

Dose-Response Modelling and Optimization of Quantitative High- Throughput Screening Assays

By

Shintaro Hagiwara

M.Sc (Probability and Statistics), Carleton University, Ottawa, ON

B.Math (Honours Statistics), Carleton University, Ottawa, ON

A thesis submitted to the Faculty of Graduate and Post Doctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Mathematics and Statistics

Carleton University

Ottawa, Ontario

©2020

Shintaro Hagiwara

Abstract

Human health risk assessment is a process designed to characterize potential health risks associated with exposure to environmental agents. Classically, it involves four main steps: (1) hazard identification, (2) dose-response assessment, (3) exposure assessment, and (4) risk characterization. Traditionally, toxicological testing has relied heavily on experimental animals to predict potential human health risk. Motivated in part by the 2007 U.S. National Research Council report, *Toxicity Testing in the 21st Century: A Vision and a Strategy*, there has been a shift towards new approach methodologies using *in vitro*, *in silico*, and *in chemico* techniques. Quantitative high-throughput screening (qHTS) is one such methodology that can rapidly produce *in vitro* assays for thousands of chemicals that can be analyzed for hazard identification and dose-response assessment purposes.

This thesis provides a statistical foundation for effectively designing qHTS assays and extends the value of information (VOI) framework to allow for more realistic comparisons in practice. Using over 8,000 qHTS assays from the Tox21 program, we develop optimal designs that maximize the efficiency of BMD estimates. Further, we extend the framework for a VOI analysis to compare the benefits realized by collecting information to aid in health risk decision-making, including variables such as timeliness, cost, and reduction in uncertainty.

Initially, the statistical properties of maximum likelihood estimators (MLEs) of the BMDs and underlying Hill model parameters are studied. A simulation study is performed to investigate the relationship between the properties of the MLEs and the observed concentration response curve (CRC).

The development of optimal designs based on BMD estimates for qHTS assays and the comparison of efficiencies for six fixed designs (FDs) is considered next. The relatively low median relative efficiencies for the FDs suggests that without some knowledge of the model parameters, no fixed design can consistently deliver high efficiency across all shapes of CRCs.

Finally, the VOI framework is extended by calculating the total social cost (TSC) over a specified time horizon instead of the annualized social cost (ASC) traditionally used in practice. The TSC uses the timeliness of information collected, thereby allowing for a more complete comparison of alternative toxicological testing methodologies.

Acknowledgements

It would not have been possible to complete this thesis without the help and support from the people around me. First and foremost, I would like to express my sincere gratitude to my supervisors Dr. Patrick Farrell and Dr. Daniel Krewski for their excellent guidance and support throughout my time at Carleton University.

I would also like to thank Dr. Salomon Sand, for his permission to use the dataset that is used in this thesis.

I would also like to thank the members of my thesis examination committee, Dr. Matthew Wheeler of National Institutes of Health, Dr. Paul Villeneuve of Carleton University, Dr. Jason Nielsen of Carleton University, Dr. Cindy Feng of University of Ottawa for taking the time to read and provide their insightful inputs on my thesis research.

I owe my greatest appreciation to my family for their support and love they provided me throughout my entire life. I would like to further thank my girlfriend, Hui Zhang, for providing daily support and warm encouragement to help me achieve this milestone.

Contents

Abstract.....	ii
Acknowledgements.....	iv
List of Tables	viii
List of Figures	ix
List of Abbreviations	x
1. Introduction	1
1.1. Motivation and Statement of Problem.....	1
1.2. Outline	5
2. Statistical Properties of the Benchmark Dose Based on the Hill Model.....	7
2.1. Abstract.....	7
2.2. Quantitative high-throughput screening.....	8
2.3. Hill model for qHTS assays and corresponding BMDs	9
2.4. Maximum Likelihood Estimators	11
2.5. Concentration-Response Modelling	12
2.6. Simulation Scenarios.....	12
2.7. Performance Evaluation Criteria for MLEs.....	13
2.8. Analysis and Results.....	15
2.8.1. Simulation under Constant Variance	15
2.8.2. Effect of the BMR.....	21
2.8.3. Effect of Error Standard Deviation.....	21
2.8.4. Effect of Number of Replicates	23
2.8.5. Effect of λ	24
2.9. Discussion and Conclusions	24
3. Optimal Design for BMD under the Hill Model for qHTS assays.....	27
3.1. Abstract.....	27
3.2. Quantitative high-throughput screening.....	28
3.3. Hill model for qHTS assays and corresponding BMDs	29
3.4. Maximum Likelihood Estimators	30
3.4.1. MLEs for the Hill Model Parameters.....	30
3.4.2. MLEs for BMR and BMD.....	31
3.5. Optimal Design for Estimating the BMD.....	32
3.6. Properties of the Optimal Designs.....	34

3.7. Relative Efficiency of Fixed Designs in Practice	36
3.7.1. Discussion of Fixed Designs	36
3.7.2. Comparing Relative efficiencies.....	36
3.8. Discussion and Conclusions	38
4. Use of Value of Information Analysis in Evaluating the Effects of Timeliness, Uncertainty Reduction, and Throughput of Information Collection Procedures	40
4.1. Abstract.....	40
4.2. Introduction	40
4.3. Methodology.....	44
4.3.1. Risk.....	44
4.3.2. Information Collection and Bayesian Updating	45
4.3.3. Objective Functions	46
4.3.4. Decision Rules	48
4.3.5. Incorporating Uncertainty	49
4.3.6. Expected Net Benefit of Sampling (ENBS)	52
4.3.7. Return on Investment (ROI).....	52
4.4. Illustrative Applications	52
4.4.1. Parameters:.....	54
4.4.2. Decision Criteria.....	54
4.4.3. Scenario 1.....	55
4.4.4. Scenario 2.....	57
4.4.5. Scenario 3.....	58
4.4.6. Cross-Scenario Findings	59
4.4.7. Testing Multiple Chemicals.....	59
4.4.8. Trade-offs between Reduction in Uncertainty and Timeliness of Testing Information	60
4.4.9. Effect of Uncertainty in Multiple Parameters.....	61
4.4.10. Effect of Control Cost.....	61
4.5. Discussion and Conclusions	62
5. Summary, Conclusion, and Future Work	65
Bibliography	71
Appendix A: Figures and Tables	75
A.1. Tables and Figures for Chapter 2.....	75
A.2. Tables and Figures for Chapter 3.....	94

A.3. Tables and Figures for Chapter 4.....	98
Appendix B: Selected Proofs and Verifications.....	108
B.1. Regularity Conditions and Identifiability	108
B.1.1. Identifiability	108
B.1.2. Regularity Conditions.....	109
B.2. Number of Concentration Levels Required for the Optimal Design for BMR	113

List of Tables

Table 2.1: Hill parameters for 8 scenarios	75
Table 2.2: Concentration levels used for 8 simulations in log-scale.....	75
Table 2.3: Performance Evaluation Statistics for Hill Model Parameters.....	76
Table 2.4: Performance Evaluation Statistics for BMD	77
Table 2.5: Mean absolute deviation (MAD) for 8 simulations.....	77
Table 2.6: Performance Evaluation Statistics for BMDs with various BMRs (BMR=10%, 5%, and 1%)	78
Table 2.7: Performance Evaluation Statistics for Hill Model Parameters with $\sigma_0 = 0.025$, $\lambda = 0$, and no replicates.....	79
Table 2.8: Performance Evaluation Statistics for Hill Model Parameters with $\sigma_0 = 0.075$, $\lambda = 0$, and no replicates.....	80
Table 2.9: Performance Evaluation Statistics for BMDs with various σ ($\sigma = 0.025, 0.05$ and 0.075).....	81
Table 2.10: Mean absolute deviation (MAD) for 8 simulations with varying replications r ($r = 1, \dots, 10$).....	82
Table 2.11: Performance Evaluation Statistics for BMDs with various λ ($\lambda = 0, 0.5$ and 1).....	83
Table 2.12: Performance Evaluation Statistics for Hill Model Parameters with $\lambda = 0.5$	84
Table 2.13: Performance Evaluation Statistics for Hill Model Parameters with $\lambda = 1$	85
Table 3.1: Six fixed designs for BMD_{ER} analysis.....	94
Table 3.2: Relative Efficiencies for six fixed designs for BMD_{ER} based on 8,456 data.	94
Table 3.3: Six fixed designs for BMD_{AR} analysis.....	95
Table 3.4: Relative Efficiencies for six fixed designs for BMD_{AR} based on 8,456 data.	95
Table 4.1: Parameters used to characterize illustrative VOI calculation scenarios	98
Table 4.2: Parameters governing toxicity and exposure of a chemical	99
Table 4.3: Information regarding Test A and Test B	99
Table 4.4: VOI analysis under target-risk decision making	99
Table 4.5: VOI analysis under benefit-risk decision making	100
Table 4.6: Parameters governing toxicity and exposure of a chemical	100
Table 4.7: Various VOI outcomes based on TRDM	101
Table 4.8: VOI analysis under benefit-risk decision making	101
Table 4.9: Parameters governing toxicity and exposure of a chemical	102
Table 4.10: Various VOI outcomes based on TRDM	102
Table 4.11: VOI analysis under benefit-risk decision making	103
Table 4.12: EVSI for various values of uncertainty reduction and timeliness of testing over 20-year time horizon under three scenarios (in \$M).....	103
Table 4.13: EVOI values for various values of the ratio CC_{max}/V^*	104

List of Figures

Figure 2.1: Eight CRCs for simulation study and estimated CRCs from 8,456 qHTS experiments .	86
Figure 2.2: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.05$, $\lambda = 0$, and no replicates.....	87
Figure 2.3: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.05$, $\lambda = 0$, and no replicates.....	88
Figure 2.4: RRMSE for BMD against various values of η for each of the 8 Scenarios	89
Figure 2.5: Performance evaluation statistics for BMD vs. number of replicates.....	89
Figure 2.6: Performance evaluation statistics for Hill model parameter estimates vs. number of replicates.....	90
Figure 2.7: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.025$, $\lambda = 0$, and no replicates.....	91
Figure 2.8: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.075$, $\lambda = 0$, and no replicates.....	92
Figure 2.9: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.05$, $\lambda = 0$, and $r = 4$ replicates	93
Figure 3.1: Relative efficiency of the COD for BMD_{ER} as a function of ER_{MCT} (left) and Dose allocations used in COD for BMD_{ER} as a function of ER_{MCT} (right)	96
Figure 3.2: Relative efficiencies for BMD_{ER} using 8,456 data	96
Figure 3.3: Differences in relative efficiencies for BMD_{ER} compared to FD_1 using 8,456 data	97
Figure 3.4: Relative efficiencies for BMD_{AR} using 8,456 data.....	97
Figure 4.1: Uncertainty distributions for Scenario 1.....	105
Figure 4.2: Uncertainty distributions for Scenario 2.....	106
Figure 4.3: Uncertainty distributions for Scenario 3.....	107

List of Abbreviations

Given the large number of acronyms used in this thesis, it is recommended that the reader ensure ease of access to the list presented in the next two pages.

ACOD: Average constrained optimal design

AC₅₀: Half maximal activation concentration

AF: Adjustment factor

AR: Additional response

ASC: Annualized social cost

BMD: Benchmark dose

BMR: Benchmark response

BRDM: Benefit-risk decision-maker

CC: Control cost

COD: Constrained optimal design

CRC: Concentration-response curve

CV: Coefficient of variation

DRC: Dose-response curve

ENBS: Expected net benefit of sampling

EPA: United States environmental protection agency

ER: Extra response

ETSC: Expected total social cost

EVIS: Expected value of information system

EVOI: Expected value of information

EVPI: Expected value of perfect information

EVPPPI: Expected value of perfect partial information

EVSI: Expected value of sample information

FD: Fixed design

HC: Health cost

ICC: Information collection cost
IC₅₀: Half maximal inhibitory concentration
IVIVE: In vitro to in vivo extrapolation
mCT: Minimum (non-zero) concentration tested
MCT: Maximum concentration tested
MLE: Maximum likelihood estimator
NRC: United States national research council
OD: Optimal design
OF: Objective function
ORE: Optimal reduction in exposure
PoD: Point of departure
QALY: Quality adjusted life year
qHTS: Quantitative high-throughput screening
RB: Relative bias
RE: Relative efficiency
ROI: Return on investment
RRMSE: Relative root mean-squared-error
TRDM: Target-risk decision-maker
TRL: Target risk level
TRV: Toxicity reference value
TSC: Total social cost
VOI: Value of information
VSL: Value of a statistical life
WTP: Willingness to pay

1. Introduction

1.1. Motivation and Statement of Problem

Human health risk assessment (HHRA) is a process designed to characterize potential health risks associated with exposure to environmental agents. Classically, this process involves four main steps, including: (1) hazard identification, (2) dose-response assessment, (3) exposure assessment, and (4) risk characterization (U.S. NRC, 1983). In step (1), hazard identification aims to determine whether exposure to the agent of interest has the potential to cause specific adverse health effects. Should a hazard be identified, dose-response assessment is undertaken to model how risk changes with the level of exposure in step (2). The response may be given as categorical (e.g., presence or absence of an adverse effect) or continuous (e.g., weight loss) variable, depending on the nature of the available data. Exposure assessment, step (3), involves measuring or estimating the human exposure. Exposure can be given in terms of intensity, frequency, and duration to the agent of interest. In step (4), risk calculation integrates the information gathered in the previous steps to arrive at the quantitative (or possibly semi-quantitative) statement about the level of risk associated with actual human exposure conditions.

Traditionally, toxicological testing has relied heavily on experimental animals to predict potential human health risk. Motivated in part by the 2007 U.S. National Research Council (U.S. NRC) report, *Toxicity Testing in the 21st Century: A Vision and a Strategy*, there has been a shift towards use of alternative test methods, including high-throughput *in vitro* screenings (U.S. NRC, 2007). This shift is emphasized in the U.S. Environmental Protection Agency (U.S. EPA) framework for the next generation of the risk science (Krewski et al. 2014), which emphasizes alternative test methods as the key to broader coverage of the large number of environmental agents, which people may be exposed. A recent review of the progress made over the last decade implementing 2007 NRC vision document

substantial uptake of the new approach methodologies (NAM) in toxicological risk assessment (Krewski et al. 2020).

High-throughput screening (HTS) technique gained popularity in early 1990s as a means of rapid hit identification, where each chemical was tested using a single concentration (Schenecke and Boström, 2006). Although HTS technique allowed testing of a large number of chemicals much more rapidly, it suffered from a large number of false positives as well (Malo et al. 2006). To overcome the shortcomings of HTS, Inglese et al. (2006) introduced quantitative-HTS (qHTS), which extends the concept of the traditional HTS by testing response at various concentration levels. Inglese et al. (2006) also report traditional HTS resulted in a false-positive rate of 2% or less, and a false-negative rate of 40% or more, when compared to the results given by qHTS. Since qHTS employs a number of experimental concentrations, it allows the analysis of concentration-response relationship, in addition to hazard identification.

The concept of the benchmark dose (BMD) as a key indicator of toxicity was introduced by Crump (1984, 1994) to overcome some of the limitations of previous indicators such as the no-observed-adverse-effect-level (NOAEL) dating back to the origin of the modern toxicological assessment (Lehman and Fitzhugh 1955). By using all of the available dose-response data to estimate the BMD – defined as the dose associated with a specified increase in risk, or change in response, called benchmark risk or benchmark response (BMR) such as 5% – the BMD better reflected impact of the sample size on precision of the indicator, with larger sample sizes reducing more precise estimates of the BMD. Presently, the BMD is widely used as a point of departure (PoD) for establishing human reference dose (RfD), obtained by dividing the BMD by the a series of appropriate adjustment factors (AFs) accounting for inter-species differences and intra-species variations in susceptibility to the agent of interest (Davis et al. 2011).

The European Food Safety Agency (EFSA) recommends the use of a default BMR of 10% extra risk with quantal response data, and a BMR of 5% change in response with respect to the mean background response (EFSA 2009). EPA suggests the use of a BMR of 1 to

10% for quantal data, and for continuous data, a BMR that corresponds to a minimal biologically significant response (e.g., 10% decrease in body weight), when such information is available (Davis et al. 2011; EPA 2012).

Estimation of the BMD requires fitting a suitable dose-response modelling to available data. Assuming that the dose-response relationship is monotonic and nonlinear, a number of models can be used for continuous response data, including: the Hill model, exponential model, power model, and polynomial model (Hill 1910; Slob 2008; EPA 2012). The Hill model is widely used in pharmacology, biology, and pharmacokinetic-pharmacodynamic modelling, in part due to its flexibility and ease of interpretability of its parameters (Goutelle et al., 2008; Shockley, 2015).

Although the BMD is widely used as the PoD for purposes of toxicological risk assessment, ultimately leading to derivation of the RfD as a human exposure guideline, other PoDs have also been discussed in literature. Sand et al. (2011) introduced the concept of the signal-to-noise crossover dose (SNCD). The SNCD is defined as the dose where the ratio of the change in response (*i.e.*, signal) and the width of the two-sided 90% confidence interval on the absolute change (*i.e.*, noise) corresponds to some threshold value. The SNCD was first applied to quantal response data.

Sand et al. (2017) extended and applied the concept of SNCD to continuous response data using qHTS assays. Their analysis included 8,456 CRCs generated on assays collected as part of the U.S. Tox21 program Phase I (Tice et al. 2013). Here, approximately 2,800 chemicals were tested with each assay using 14- or 15-concentration levels. Prior to analysis, the concentration-response data was normalized as described in Inglese et al. (2006). Data from Sand et al. (2017) include three groups of assays: cytotoxicity assays, nuclear receptor assays, and stress response assays, with many of them available from PubChem BioAssay database (Wang et al. 2012). Of the 11,240 qHTS assays collected, Sand et al. (2017) used a curve classification system (Inglese et al. 2006; Huang et al. 2011; Xia et al. 2011) to determine which assays to retain for analysis. This investigation resulted in 8,456 being kept. Class 1 (complete curve) refers to observing two asymptotes for

sigmoidal CRCs such as the Hill model, while Class 2 (incomplete curve) captures one asymptote with an inflection point. Class 3 (single point activity) shows significant response only at the maximum concentration tested (MCT), and Class 4 (inactive) denotes data where titration shows efficacy of less than 30% or no response. In their analysis, Sand et al. (2017) made use of the 8,456 assays that fell into first two classes due to difficulty in fitting an appropriate CRC to those in the two latter classes. Although this thesis focuses on the BMD rather than the SNCD, we will use the 8,456 datasets to illustrate the application of our approaches to estimating the BMDs using real world data.

In its simplest form, risk is given as a product of toxicity and exposure. Using the information collected in steps (2) and (3) of HHRA, decision-makers are tasked with determining whether a chemical should be regulated, and if so, by how much. Since the available information contains some level of uncertainty, the decision must be made by weighing the likelihood of making an incorrect (or sub-optimal) decision, and the expected loss incurred in reaching such a decision. For instance, incorrectly deciding not to regulate a hazardous chemical may lead to unacceptably high health risk, which leads to a large health cost. On the other hand, regulating a non-hazardous chemical may produce unnecessary burdens in regulation cost, as well as social cost.

Value-of-Information (VOI) analysis is a systematic process to help decision-makers to answer the following questions: does currently available information provide sufficient evidence to allow a decision to be made with confidence? If not, does the benefit of collecting additional information outweigh the cost of collecting such information? And, when there is more than one information collection methodology, which method is best? In this thesis, we will develop a VOI methodology for comparing alternative toxicity testing strategies. This methodology could be used to compare traditional and new approach toxicity testing methodologies, or to compare alternative NAMs to determine the most cost-effective approach to obtaining comprehensive toxicity testing data on the large number of substances present in the human environment that require further toxicological evaluation.

1.2. Outline

The qHTS technique is considered one of the NAMs that will modernize toxicological testing. Given the anticipated prominence of qHTS in future toxicity testing strategies, it is important that the statistical foundation for the analysis of qHTS data being well established. Of particular importance are dose-response models that can be used to discriminate the CRCs for qHTS data, and the subsequent estimation of the BMD based on the model fit. In addition to discussing optimal approaches for dose-response modeling for qHTS data, we also consider the application of VOI methods to inform future toxicity testing strategies in support of toxicological risk assessment. Given the large number of chemicals that currently require toxicological evaluation, the objective function used in VOI analysis should be extended to incorporate timeliness of information collection. The overarching objective of this thesis is to develop an optimal design for qHTS assays using BMD for the PoD and determine its effectiveness as a NAM, and to provide an extended VOI framework that incorporates timeliness and cost of information collection.

Chapter 2 investigates the statistical properties of the MLEs of the Hill model parameters and corresponding BMD under an experimental design that is commonly used in practice. We investigate via simulation the effect of the degree to which the entire CRC is captured over the experimental range on the maximum likelihood estimators of the BMD and Hill model parameters. Sensitivity analyses are also conducted to study the effect of the value of the benchmark response (BMR), the error standard deviation, the number of replicates, and the assumption of constant error variance.

Chapter 3 develops an optimal design that minimizes the variance of an MLE for the BMD under the Hill model. As the optimal design require knowledge of Hill model parameters, six fixed designs that do not require knowledge of the underlying CRC are considered, and their relative efficiencies are compared with respect to the optimal design.

Chapter 4 extends the VOI framework to incorporate timeliness and cost of information collection, when comparing two (or more) testing methodologies. The analytical framework for VOI analysis is presented first, and an illustrative prototypical application of evaluation of the VOI associated with particular testing methodologies is presented.

Finally, conclusions and potential future research directions are presented in chapter 5.

2. Statistical Properties of the Benchmark Dose Based on the Hill Model

2.1. Abstract

Over the last decade, there has been increasing use of new approach methodologies (NAMs) such as high-throughput *in vitro* assays in toxicity testing as alternatives to traditional animal bioassays. Quantitative high-throughput screening (qHTS) is an *in vitro* testing methodology that examines cellular responses to various concentrations of chemicals. This chapter investigates the statistical properties of the benchmark dose (BMD) based on the four-parameter Hill model under an experimental design that is commonly used in practice. In particular, the 8,456 datasets in Sand et al. (2017) are used to create eight representative concentration response curves (CRCs) over the realm of curves resulting from these datasets. We investigate via simulation the effect of the degree to which the entire CRC is captured over the experimental range on the maximum likelihood estimators of the BMD and Hill model parameters. Sensitivity analysis is also conducted to study the effect of the value of the benchmark response (BMR), the error standard deviation, the number of replicates, and the assumption of constant error variance. The simulation study underlines the importance of setting dose levels at the lower and upper asymptotes of the CRC, and at concentrations where the response changes rapidly. While the sensitivity analysis demonstrates that changes to the parameters listed has some effect on the estimation of the Hill model parameters and the BMDs, it is predominantly the differences captured by the eight representative CRCs that influence the quality of the estimators. The results associated with changing the number of replicates suggest that, while no replication may be necessary for screening purposes, replication is recommended if the primary objective is the estimation of the BMD.

2.2. Quantitative high-throughput screening

In 2019, US Environmental Protection Agency's (EPA) announced its decision to eliminate all mammalian toxicity testing by 2035. This aligns with the U.S. National Research Council's (NRC, 2007) view provided in its report, *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Quantitative high-throughput screening (qHTS), introduced by Ingles et al. (2006) is an *in vitro* testing methodology that examines cellular responses to various concentrations of chemicals. This approach offers considerable advantages over traditional toxicity testing in terms of timeliness and cost. Unlike mammalian toxicity testing, which may require a few years and millions of dollars to run an experiment on a single chemical (Lave and Omenn, 1986), qHTS can produce results for multiple chemicals in a matter of days.

An important aspect of the analysis of qHTS data is characterization of the concentration-response curve (CRC) for the agent of interest. The Hill model, described below, is widely used for concentration-response modelling of qHTS data (Shockley, 2015). Establishing human health guidelines for potentially hazardous chemicals requires the derivation of a value called the point of departure (PoD). In this chapter, we focus on the benchmark dose (BMD), which is used by many health organizations internationally (Davis et al. 2011) as the PoD, for deriving human exposure guidelines. In practice, the toxicity reference value (TRV) is determined by dividing the PoD by a suitable adjustment factor (AF), with $TRV = BMD/AF$. This AF is usually comprised of a product of multiple factors representing inter-species differences, inter-individual variation, and other relevant considerations involved in interpretation of the toxicological data. With qHTS data, there will also be a need for *in vitro* to *in vivo* extrapolation (IVIVE) to calibrate concentrations in the cellular test system used with corresponding doses in human tissues using pharmacokinetic models (Anderson and Krewski 2009; Krewski et al. 2009, 2011; NRC 2007).

2.3. Hill model for qHTS assays and corresponding BMDs

Let $y(x)$ represent the normalized cell activity level observed in the qHTS assay when exposed to an agent with concentration level x measured in micromolar (μM). Given that the response is a continuous variable, a 4-parameter Hill model may be employed to characterize the dose-response curve (DRC), more appropriately termed concentration-response curve (CRC) in *in vitro* systems. The Hill model is given by

$$y = \alpha + \frac{\theta - \alpha}{1 + (\kappa/x)^\eta}, \quad (2.1)$$

where α corresponds to cell activity level at zero concentration and θ represents maximum activity level at infinite concentration. The concentration corresponding to half maximal response (AC_{50} , or IC_{50} for an inhibitor) is denoted by κ , and the shape (or steepness) of the CRC is determined by η . If the chemical concentration levels are expressed in log-scale ($z = \log_{10}(x)$), the above equation becomes

$$y = \alpha + \frac{\theta - \alpha}{1 + 10^{(\kappa' - z)\eta}}, \quad (2.2)$$

where $\kappa' = \log(\kappa)$.

Consider an experiment with t concentrations, x_1, \dots, x_t , with n_i observations at concentration x_i , with a total of $n = \sum_{i=1}^t n_i$ observations. Then the observed cellular response level at concentration x_i is

$$y_{ij} = \left[\alpha + \frac{\theta - \alpha}{1 + 10^{(\kappa' - z_i)\eta}} \right] + \varepsilon_{ij} \quad (i = 1, \dots, t; j = 1, \dots, n_i), \quad (2.3)$$

with ε_{ij} representing the error term for the j^{th} replicate for the i^{th} concentration. In what follows, we assume that the error terms follow a normal distribution with mean zero and variance σ_i^2 . Following Khinkis et al. (2003), we further assume that the error variance to be described by the power model:

$$\sigma_i^2 = \sigma_0^2 y_i^{2\lambda} \quad (\lambda \geq 0). \quad (2.4)$$

Here, $\lambda = 0$ corresponds to constant variance model, while $\lambda = 1$ corresponds to increasing variance with the coefficient of variation (CV) held constant.

To assess the toxicity of the agent of interest, we will employ BMD as our PoD. The BMD is defined as the concentration level that corresponds to a specified response called the benchmark response (BMR). The BMR is commonly expressed in terms of either additional response (AR) or extra response (ER), denoted by BMR_{AR} and BMR_{ER} , respectively. Without loss of generality, we assume that the agent of interest is an activator, with an increasing CRC. (A decreasing CRC associated with an inhibitor can be handled in a complimentary manner.) For example, a BMR_{AR} of 5% means that the response is increased by 5% compared to the background activity level α , while a BMR_{ER} of 5% corresponds to an increase of 5% within the range of activity ($\theta - \alpha$). Mathematically, corresponding BMDs in log-scale are expressed as

$$\log_{10}(BMD_{AR}) = \kappa' - \frac{\log\left(\frac{\theta - \alpha}{BMR_{AR}} - 1\right)}{\eta}, \quad (2.5)$$

and

$$\log_{10}(BMD_{ER}) = \kappa' - \frac{\log\left(\frac{1}{BMR_{ER}} - 1\right)}{\eta}. \quad (2.6)$$

From (2.6), we can calculate the concentration range corresponding to a range of 1 to 99% in the extra activity level as

$$BMD_{ER,99\%} - BMD_{ER,1\%} = \frac{\log\left(\frac{1}{0.01} - 1\right) - \log\left(\frac{1}{0.99} - 1\right)}{\eta} = \frac{3.99}{\eta}. \quad (2.7)$$

This implies that, if $\eta = 1$, then 98% of the change in the CRC falls within a 4-fold change in log-concentration. Similarly, a change from 5% to 95% spans a 2.6-fold change in log-concentration, and 10% to 90% corresponds to a change in concentration of approximately 2-fold on the log scale. This relationship also implies that, as we double the value of η , the corresponding range in log-concentration would be halved.

2.4. Maximum Likelihood Estimators

The method of maximum likelihood can be used to estimate the Hill model parameters and the corresponding BMDs. The maximum likelihood estimators (MLEs) are obtained by maximizing the likelihood function L given the observed responses under the assumed distribution function for the error terms. In particular, MLEs for the Hill parameters are such that

$$\hat{\boldsymbol{\phi}}' = (\hat{\alpha}, \hat{\kappa}, \hat{\eta}, \hat{\theta}) = \underset{(\alpha, \kappa, \eta, \theta)}{\operatorname{argmax}} L(\alpha, \kappa, \eta, \theta | \mathbf{y}, \mathbf{x}), \quad (2.8)$$

where $\mathbf{y}' = (y_1, \dots, y_n)$ and $\mathbf{x}' = (x_1, \dots, x_n)$.

Suppose that the expected response μ_i at concentration x_i follows a 4-parameter Hill model ($\mu_i = f(x_i | \boldsymbol{\phi})$). Furthermore, suppose that the error terms $\epsilon_{ij} = y_{ij} - f(x_i | \boldsymbol{\phi})$ follow a normal distribution with mean zero and standard deviation σ_i , where σ_i follows the power model $\sigma_i = f(x_i | \boldsymbol{\phi})^\lambda$. The likelihood function is then given by

$$L(\boldsymbol{\phi}, \sigma | \mathbf{y}) = \prod_{i=1}^t \prod_{j=1}^{n_i} (2\pi\sigma_i^2)^{-1/2} \exp\left(-\frac{(y_{ij} - \mu_i)^2}{2\sigma_i^2}\right), \quad (2.9)$$

with log-likelihood

$$l(\boldsymbol{\phi}, \sigma | \mathbf{y}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^t \ln(\sigma_i^2) - \sum_{i=1}^t \sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_i)^2}{2\sigma_i^2}. \quad (2.10)$$

The MLEs that maximize the above equations are denoted by $\hat{\boldsymbol{\phi}}'_{MLE} = (\hat{\alpha}_{MLE}, \hat{\kappa}_{MLE}, \hat{\eta}_{MLE}, \hat{\theta}_{MLE})$. Since it is impossible to solve for $\hat{\boldsymbol{\phi}}_{MLE}$ analytically, numerical methods such as the Newton-Raphson method are required to obtain the MLEs.

Once the MLEs of the Hill model parameters have been obtained, a corresponding estimate of the desired BMD can be calculated using the invariance property of the MLEs. For example, the BMDs based on additional response and extra response can be respectively estimated as

$$\widehat{\text{BMD}}_{\text{AR}} = \frac{\left(\frac{\hat{\theta} - \hat{\alpha}}{\widehat{\text{BMR}}_{\text{AR}}} - 1\right)^{1/\hat{\eta}}}{\hat{\kappa}}, \quad (2.11)$$

and

$$\widehat{\text{BMD}}_{\text{ER}} = \frac{\left(\frac{1}{\widehat{\text{BMR}}_{\text{ER}}} - 1\right)^{1/\hat{\eta}}}{\hat{\kappa}}. \quad (2.12)$$

2.5. Concentration-Response Modelling

The primary objective of this chapter is to investigate the statistical properties of the parameters of the Hill model used to describe the CRC for the qHTS data. This will be done using computer simulation to evaluate the properties of the MLEs of the Hill model parameters, as well as the corresponding BMD. Specifically, we will examine the relative bias (RB), the coefficient of variation (CV), and relative root mean-squared-errors (RRMSE) of the MLEs of the both the Hill model parameters and the BMDs derived from these parameters.

Our simulation study is based on eight prototypical CRCs, that spans the range of shapes described by the Hill model. To ensure that the simulation results are relevant to experimental conditions encountered in practice, the eight scenarios chosen are based on analysis of over 8,000 actual qHTS CRCs previously considered by Sand et al. (2017).

2.6. Simulation Scenarios

Sand et al. (2017) collected a total of 11,240 qHTS assays from Phase I of the Tox21 program. These were broken down into three types: nuclear receptor assays, cytotoxicity assays, and stress response assays. For each assay, the associated CRC is obtained using data consisting of 14- or 15-concentration levels.

Of the 11,240 qHTS assays collected, Sand et al. (2017) used a curve classification system (Inglese et al. 2006; Huang et al. 2011; Xia et al. 2011) to determine which assays to retain

for analysis. Class 1 (complete curve) refers to observing two asymptotes for sigmoidal CRCs such as the Hill model, while Class 2 (incomplete curve) captures one asymptote with an inflection point. Class 3 (single point activity) shows significant response only at the maximum concentration tested (MCT), and Class 4 (inactive) denotes data where titration shows efficacy of less than 30% or no response. In their analysis, Sand et al. (2017) made use of the 8,456 assays that fell into the first two classes due to difficulty in fitting an appropriate CRC to those in the two latter classes. In our simulation, the eight curve scenarios are selected to cover the range of CRCs observed in these 8,456 assays.

Figure 2.1 presents the 8,456 CRCs (in grey), as well as eight CRCs (in various colours) to be used in the simulation study. Here, the minimum (non-zero) and maximum concentrations tested are labeled as mCT and MCT, respectively. Table 2.1 summarizes the values for the Hill model parameters used in the eight scenarios. We build experiments with 14 concentration levels (Table 2.2).

As Sand et al. (2017) noted that most experiments that they considered had no replicate observations at each concentration, we consider first a 14-concentration experiment with one observation at each concentration. Furthermore, we let $\lambda = 0$ and $\sigma_0 = 0.05$, reflecting constant variance across the experimental range. Initially, we set the target BMR to be 5%, but will consider BMRs of 1 and 10% in subsequent sensitivity analyses.

The MLEs are computed using *mle2* function from *bbmle* package in *R*. Since this optimization technique involves a numerical integration, convergence was not achieved in some cases. Although only convergent results are included in the analysis, we report on the number of cases where the convergence was not achieved.

2.7. Performance Evaluation Criteria for MLEs

Let $\phi' = (\alpha, \kappa, \eta, \theta)$ denote the parameters in the Hill model, which, once estimated, can be used to estimate the BMD. For the eight scenarios considered, $B = 500$ simulations were conducted. Let B_i^* denote the number that converged in the i^{th} scenario. The statistical properties of MLEs were evaluated using four criteria: relative bias (RB),

coefficient of variation (CV), relative root mean-squared-error (RRMSE), and confidence interval coverage.

Let $\hat{\phi}_b$ be the b^{th} MLE for the parameter of interest ϕ . The RB is then defined as the difference between the expected value of the MLE and the parameter, divided by the parameter of interest:

$$RB(\hat{\phi}) = E\left(\frac{\hat{\phi} - \phi}{\phi}\right). \quad (2.13)$$

Given the MLE for each of the simulated dataset, $\hat{\phi}$, we estimate RB by:

$$\widehat{RB}(\hat{\phi}) = \frac{1}{B_i^*} \sum_{b=1}^{B_i^*} \left(\frac{\hat{\phi}_b - \phi}{\phi}\right). \quad (2.14)$$

Similarly, CV can be defined as

$$CV(\hat{\phi}) = \frac{\sqrt{Var(\hat{\phi})}}{\phi}, \quad (2.15)$$

where $Var(\hat{\phi}) = E[\hat{\phi} - E(\hat{\phi})]^2$ and estimated by

$$\widehat{CV}(\hat{\phi}) = \frac{1}{\phi} [\widehat{Var}(\hat{\phi})]^{1/2} = \frac{1}{\phi} \left[\frac{\sum_{b=1}^{B_i^*} (\hat{\phi}_b - \bar{\hat{\phi}})^2}{B_i^*} \right]^{1/2}. \quad (2.16)$$

Using the relationship $MSE(\hat{\phi}) = Var(\hat{\phi}) + Bias(\hat{\phi})^2$, the $RRMSE(\hat{\phi})$, is given by

$$RRMSE(\hat{\phi}) = \sqrt{[CV(\hat{\phi})]^2 + [RB(\hat{\phi})]^2}, \quad (2.17)$$

And estimated by

$$\widehat{RRMSE}(\hat{\phi}) = \sqrt{[\widehat{CV}(\hat{\phi})]^2 + [\widehat{RB}(\hat{\phi})]^2}. \quad (2.18)$$

Finally, confidence interval coverage is calculated as the proportion of the B_i^* MLEs that fall within the 90% large sample confidence interval based on standard normal distribution. That is,

$$\text{Coverage}(\hat{\phi}) = \frac{1}{B_i^*} \sum_{b=1}^{B_i^*} I_{[L,U],b}(\phi), \quad (2.19)$$

where

$$I_{[L,U],b}(\phi) = \begin{cases} 1 & \text{if } \phi \in \left(\hat{\phi}_b \pm z_{0.95} \sqrt{\widehat{\text{Var}}(\hat{\phi}_b)} \right) \\ 0 & \text{otherwise.} \end{cases} \quad (2.20)$$

Here, $\widehat{\text{Var}}(\hat{\phi}_b)$ is given by inverting the information matrix.

To assess the goodness-of-fit of the simulated CRCs, we present the mean absolute distance (MAD) of simulated curvatures at below, between, and above the experimental range. The calculation of MAD within the experimental range is as follows.

1. Divide the experimental range into 100 log-equidistance intervals.
2. For each of the estimated CRCs, calculate the b^{th} MAD

$$\text{MAD}_{i,b,\text{between}} = \frac{\sum_{l=1}^{100} |\hat{y}_{i,b,l} - y_l|}{100}. \quad (2.21)$$

3. The MAD for the i^{th} simulation is then calculated as

$$\text{MAD}_{i,\text{between}} = \sum_{b=1}^{B_i^*} \frac{\text{MAD}_{i,b}}{B_i^*}. \quad (2.22)$$

The $\text{MAD}_{i,\text{below}}$ and $\text{MAD}_{i,\text{above}}$ are obtained in a similar manner.

2.8. Analysis and Results

2.8.1. Simulation under Constant Variance

For each of the eight scenarios, $B = 500$ simulated experiments were generated. Of the 500 simulations conducted in each scenario, convergence was obtained in 112 (22.4%),

500 (100%), 500 (100%), 499 (99.8%), 500 (100%), 229 (45.8%), 441 (88.2%), and 123 (24.6%) cases for Scenarios 1 through 8, respectively. The blue curve in Figure 2.2 represents the true CRC, with the green curves corresponding Hill models fit to each simulated experiment. Black dots indicate simulated observations. The purple vertical line indicates the value of κ , and the red square denotes the BMD, with $BMR_{AR} = BMR_{ER} = 5\%$.

The results of the simulation study are summarized in Table 2.3, which shows the RB, CV, RRMSE, and coverage for each of the four Hill model parameters. Results for α , θ , κ , and η are discussed separately below.

2.8.1.1 Estimation of α

Among the eight simulation scenarios considered, the estimation of α was exceptional for Scenarios 3, 4, 5, 6, and 8. The bias in all scenarios was negative: the maximum absolute bias was 3.0% in Scenario 1, with the bias being less than 0.6% in remaining scenarios. The root MSEs for the exceptional scenarios range between 1.4% to 5.5%. As seen in Figure 2.2, experiments in these scenarios have at least 2 concentration levels at the lower asymptote, making it easy to estimate α without incurring undue bias or variability. With the exception of Scenarios 1 and 7, the confidence coverage is within 3% from the nominal 90% level, with the lowest coverage of 82.5% observed in Scenario 7, and highest coverage of 99.1% seen in Scenario 1.

Scenarios 3 and 4 differ with respect to only the value of η , with Scenario 4 having smaller value ($\eta = 0.6$) than in Scenario 3 ($\eta = 1$). As a consequence, Scenario 4 is unable to capture the lower asymptote as well as Scenario 3, resulting in somewhat weaker performance with respect to the estimation of α .

Scenarios 1, 2, and 7 have root MSE values between 7.5% to 12.0%. In Scenario 7 with $\eta = 0.3$, the lower asymptote falls outside the experimental range, leading to greater bias, standard error, and root MSE values compared to Scenarios 3, 4, 5, 6, and 8. Scenarios 1 and 2 have comparatively small values of κ , resulting in insufficient number of test

concentrations at the lower asymptote to estimate α well. It should also be noted that even though these scenarios showed weaker performance compared to scenarios 3, 4, 5, 6, and 8, the corresponding root MSE is relatively small when compared to RRMSEs of $\hat{\theta}$, $\hat{\eta}$, and $\hat{\kappa}$.

2.8.1.2 Estimation of θ

In contrast to estimation of α , where experimental data at the lower asymptote is needed, concentrations at the upper asymptote are needed to estimate θ well. Scenarios 1, 2, 3, and 5 all have at least two concentrations at the upper asymptote, making it easy to estimate θ without incurring undue bias or variability. This is demonstrated in Table 2.3, where the RBs for these scenarios are all less than 0.64%, with the RRMSEs all less than 3.4%. Confidence coverage is within 2% the nominal 90% level for Scenarios 1 through 5, while the lowest coverage of 71.0% observed in Scenario 7.

Scenarios 3 and 4 differ only with respect to the value of η , with Scenario 4 having a smaller value ($\eta = 0.6$) than Scenario 3 ($\eta = 1$). Consequently, Scenario 4 is unable to capture the upper asymptote as well as Scenario 3, resulting in somewhat weaker performance (RB of 0.9% and RRMSE of 6.2%) with respect to the estimation of θ .

As the upper asymptote did not fall within the experimental range in Scenarios 6, 7, and 8, the performance of $\hat{\theta}$ was much worse in these cases: the RB was as large as 45.7%, and the RRMSE ranged between 69.0% to 168.5%. In addition, these scenarios had at most one concentration that was greater than κ . For Scenarios 6 and 8, this translated to only having two to three concentration levels that were not at the lower asymptote. For Scenario 7, there is a lack of data on the shape of the CRC above the inflection point. As a consequence, it is not surprising that Scenarios 6 through 8 show poor performance with respect to the estimation of the upper asymptote of the CRC. The weaker performance in these scenarios is also reflected in the deviations in confidence coverage from the nominal 90% level.

2.8.1.3 Estimation of κ

Comparing the performance of κ across the eight scenarios, Scenarios 2, 3, and 5 show the best results, with RBs ranging between 0.7% to 2.2% with CVs less than 23.1%. In contrast, $\hat{\kappa}$ is highly unstable in Scenario 7, with excessive RB and CV (greater than $10^6\%$ and $10^7\%$, respectively). Performance in Scenarios 1, 4, 6, and 8 lies between these extreme cases.

Given the comparatively large standard deviation of the error term ($\sigma = 0.05$), relative to the limited curvature of the CRC, the poor performance in Scenario 7 is understandable. With $\eta = 0.3$, the rate of change in response within the experimental range is small, with most of the experimental observations are either below the inflection point, making it difficult to infer the shape of the CRC beyond the highest concentration tested. Since κ is given in the arithmetic scale, whereas the symmetry of the Hill model is apparent on the log scale, there is a tendency for $\hat{\kappa}$ to be positively biased. (Exceptionally $\hat{\kappa}$ and $\hat{\theta}$ demonstrate negative RBs in Scenario 8.) Coverage is below 80% in Scenario 7 and reaches 100% in Scenario 1.

2.8.1.4 Estimation of η

The performance of $\hat{\eta}$ as described in Table 2.3 is similar to that of $\hat{\theta}$. Scenarios 1 through 5 have relatively small RB (ranging between 0.7% to 2.6%) and CV (between 15.2% to 21.9%), whereas Scenarios 6, 7 and 8 show RBs ranging from 4.4% to 38.2% in absolute magnitude with CVs over 36.7%. Estimation of η was particularly difficult in Scenario 7, with an RB of 38.2% and CV of 133.4%. Difficulty in estimating η in Scenarios 6, 7, and 8 may have been attributed to upper asymptote falling largely outside of the experimental range. Coverage is close to the nominal 90% level in Scenarios 2, 3, 4, and 5, while the coverage for remaining scenarios are over 97.5%.

Considering the results discussed above regarding the estimation of the four Hill model parameters, several observations can be made. Overall, Scenario 5 provided the best results in estimating the four Hill model parameters. The experiment included

concentrations at the lower and upper asymptotes, as well as on both side of the inflection point of the concentration curve. The poorest performance was seen in Scenario 7, where the experimental range was too narrow to capture data at either the lower or upper asymptote.

Scenarios 1, 6, and 8 had large values of η with κ located near the boundaries of the experimental concentration range. With a few concentrations that are not at the lower or upper asymptote, it is difficult to determine the shape of the CRC outside the experimental range. The rates of convergence of the numerical algorithm used to obtain the MLEs in these scenarios are also low. Collectively, these results highlight the need for observations at or near both the lower and upper asymptote of the CRC, along with data above and below the inflection point, in order to achieve good estimates of the four Hill model parameters. These conditions are realized in Scenario 5, in which the best estimates of these parameters are obtained.

2.8.1.5 Estimation of BMD

Having evaluated the performance of the MLEs of the four Hill model parameters, we now investigate the properties of the corresponding BMD estimates derived from these four parameters. Table 2.4 shows the statistical properties of the MLE of the BMD in the eight scenarios considered previously, with $BMR_{AR} = BMR_{ER} = 0.05$.

The performance of \widehat{BMD} varies substantially across the different scenarios. By definition, BMD_{AR} corresponds to the concentration at which the response is increased (or decreased) by 5%, compared to α . Because the BMD_{AR} is calculated relative to α , it is important that this parameter be well estimated. As seen in Table 2.4, Scenarios 5, 6, and 8 produced the most reliable estimates of the BMD_{AR} , with modest negative RBs ranging from -14.8% to -1.3% . Figure 2.4 illustrates the relationship between η and the variability in \widehat{BMD} . When the value of η range between 3 to 5, the CRCs demonstrate a high degree of curvature. For Scenarios 5, 6, and 8, this results in correspondingly low

variability in \widehat{BMD} reflected in RRMSEs that range between 15.2% to 24.3%. Coverage in Scenario 5 is 86.2% and exceeds 97% for Scenarios 6 and 8.

Scenarios 1, 2, 3, and 4 exhibit less satisfactory results, with RBs ranging from 14.0% to 57.9%, and RRMSEs between 59.3% to 164.0%. This can be attributed to greater imprecision in the estimation of α . Confidence coverage is greater than 85% in these scenarios.

Scenario 7 demonstrates poor performance, with the RB exceeding 1,000% and RRMSE exceeding 10,000%. The challenge in estimating the BMD_{AR} in this scenario is due to the limited amount of information about the shape of the CRC, with only one experimental concentration above the inflection point, and none in the upper asymptote of the curve.

Similar findings are observed with the BMD_{ER} . Scenarios 5, 6, and 8 again show the best results, with Scenario 7 providing least reliable estimates. When estimating the BMD_{ER} , it is critical to attain good estimates of the range $(\theta - \alpha)$. In Scenarios 6 through 8, where θ is poorly estimated, the performance of \widehat{BMD}_{ER} is worse than that of \widehat{BMD}_{AR} . In contrast, Scenarios 2, 3, and 4 reflect superior performance of \widehat{BMD}_{ER} relative to \widehat{BMD}_{AR} . The coverage probabilities for \widehat{BMD}_{AR} and \widehat{BMD}_{ER} are similar in all eight scenarios.

2.8.1.6 Estimation of CRC

Table 2.5 summarizes the goodness-of-fit of estimated CRCs in three parts. Within the experimental range, the goodness-of-fit is characterized by MAD_{between} ; goodness-of-fit below and above the experimental range are presented as the ratios $MAD_{\text{above}}/MAD_{\text{between}}$ and $MAD_{\text{below}}/MAD_{\text{between}}$, respectively. For the eight scenarios considered, MAD_{between} was always approximately 0.02, except for Scenario 1, where the MAD_{between} was 0.05. This value, along with the tightness of the estimated CRCs shown in Figure 2.3, suggest that the experiment with 14 concentration levels, on average, traces the underlying CRC very well. The reliability of the extrapolation of the CRC is highly dependent on whether either asymptote is captured within the experimental range. When this is true (upper asymptote for Scenarios 1 and 2, and lower asymptote for

Scenarios 5, 6, and 8), the ratio $MAD_{above}/MAD_{between}$ and $MAD_{below}/MAD_{between}$ are smaller than 1. However, some CRCs failed to provide an adequate fit outside the experimental range: in Scenarios 6, 7, and 8, where the experiment failed to capture the upper asymptote, the MAD_{above} were more than 10 times larger than $MAD_{between}$. Similarly, in Scenarios 2, 4, and 7, MAD_{below} is less stable compared to $MAD_{between}$, though to lesser degrees. This finding underscores the importance of choosing the experimental range so as to include the BMD, as variability in \widehat{BMD} can be appreciable when it falls below the mCT.

2.8.2. Effect of the BMR

Choices of the BMR typically include 10%, 5%, or 1%. As the scenarios considered to this point have focused only on a BMR of 5%, we undertake sensitivity analyses using BMRs of both 10% and 1% (Table 2.6). The RRMSE of the BMD decreases with increasing BMR (except for BMD_{ER} in Scenarios 6 and 8, where θ is not well estimated), reflecting the greater uncertainty in estimating lower and lower quantiles of the CRC. Because the BMD is often used as a PoD for establishing a TRV, with the BMD divided by an appropriate adjustment factor(s), it is desirable to choose the BMR as low as possible to minimize the degree of extrapolation below the PoD. In practice, the choice of BMR will require a balancing between a lower RRMSE achieved with higher BMR and less extrapolation below the PoD with a lower BMR. The RB of the BMD appears similar for BMRs of 10%, 5%, and 1%. With the exception of Scenario 8, where there are few experimental concentrations above the 10% response rate. Coverage probabilities are also similar for the three BMRs considered.

2.8.3. Effect of Error Standard Deviation

To investigate the effect of the magnitude of the error variance on the statistical properties of the maximum likelihood estimators of the Hill model parameters and the

associated BMDs, we repeated the above analysis with $\sigma_0 = 0.075$ and $\sigma_0 = 0.025$, values that are 50% larger and smaller than the previous value of $\sigma_0 = 0.050$.

Table 2.9 displays the performance evaluation statistics for the MLEs of the BMD_{AR} and BMD_{ER} with $\sigma_0 = 0.025, 0.05, \text{ and } 0.075$. When σ_0 is reduced from 0.050 to 0.025, the performance of the MLEs is uniformly improved for all four performance evaluation criteria (except for the RB for BMD_{ER} in Scenario 6, which has a 0.3% increase). Because the asymptotic variance of the MLEs is proportional to σ_0 , the standard deviation of the MLEs with $\sigma_0 = 0.025$ is expected to be half of that when $\sigma_0 = 0.050$. However, because the results in Table 2.9 are subject to both simulation errors and possibly differences in the number of convergent simulations, this expected relationship does not hold exactly (aberrant convergent result may also contribute to deviation from this theoretical relationship). In Scenario 5, where the BMDs have previously been shown to be well estimated when $\sigma_0 = 0.050$, the above hypothesis holds: for both \widehat{BMD}_{AR} and \widehat{BMD}_{ER} , the RRMSEs in Scenario 5 are reduced by about 52%. Conversely, in Scenario 7, where the ratio of the error standard deviation to the rate of change within the experimental range is large, a reduction in σ_0 results in tremendous improvement (>93%) in the RRMSE. In Scenarios 1, 6, and 8, the improvement in the MLEs is less than the expected 50%, ranging between 22.5% to 38.3%. This behaviour can be attributed to poor estimation of either α (in Scenario 1) or θ (Scenarios 6 and 8).

In the remaining scenarios, the reduction in the RRMSE varies between 51.7% and 64.0%, depending largely on the degree of difficulty in estimating the Hill model parameters. This is mainly due to the shape of the CRC and the extent to which the experimental concentrations capture curvature.

Conversely, when $\sigma_0 = 0.075$, the performance of the MLEs of the Hill model parameters and associated BMDs is less favourable than when $\sigma_0 = 0.050$ (except for the RB for BMD_{ER} in Scenario 6, which has a 0.7% improvement – and in three cases the coverage is slightly worse). In Scenario 5, RRMSEs for \widehat{BMD}_{AR} and \widehat{BMD}_{ER} result in about 48.5% increase.

It is also worthwhile noting that the effect of changing the magnitude of σ_0 is not necessarily symmetrical. For instance, the RRMSE for $\widehat{\text{BMD}}_{\text{ER}}$ in Scenario 8 has only declined by 22.5% when σ_0 is reduced to 0.025, while the inflation in the corresponding RRMSE is more than 260% when σ_0 is increased to 0.075. A similar pattern is observed in Scenarios 2, 4, and 7.

2.8.4. Effect of Number of Replicates

The inclusion of replicates at each concentration in an experiment can be beneficial. For instance, with r replicates, the expected standard error of the average response at each concentration tested is reduced by \sqrt{r} . In addition, a sufficient number of observations is required for the distributions of the estimators to approach the asymptotic properties of the MLEs. Under some regularity conditions, which are discussed in Chapter 2, the MLE for each Hill model parameter, as well as the BMD, are strongly consistent. Furthermore, $\sqrt{rt}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})$ converges in distribution to a multivariate normal with mean $\mathbf{0}$ and asymptotic covariance matrix $\boldsymbol{\Sigma}$.

Figures 2.5 and 2.6 show the performance evaluation statistics for the BMDs and the Hill model parameters using $r = 1, 2, 3, 4, 5$ and 10 replicates at each concentration. As r increases, RB, CV, and RRMSE approach 0, while the coverage probability nears the nominal 90% rate. This is particularly apparent in Scenarios 2, 3, 4, and 5, where the experimental concentrations capture the overall shape of the CRC. Conversely, in Scenarios 1, 6, 7, and 8, replication shows less improvement in performance (due to the difficulty in estimating Hill model parameters with concentration levels not spanning the entire CRC).

A similar pattern is seen in goodness-of-fit. As seen in Table 2.10, in Scenarios 2, 3, 4, 5, and 7, the $\text{MAD}_{\text{between}}$ is reduced by \sqrt{r} , while Scenarios 1, 6, and 8 show less of a reduction in $\text{MAD}_{\text{between}}$.

2.8.5. Effect of λ

To this point, we have investigated the performance of MLEs under the assumption of constant variance ($\lambda = 0$ in the power model). We now consider the cases of $\lambda = 0.5$ and 1 in which the error variance increases with increasing response. In all scenarios considered, α and θ are set to 0 and 1, respectively. When $\lambda > 0$, σ_i decreases as y_i approaches 0. As seen in Table 2.11, the RRMSE values for $\widehat{\text{BMD}}_{\text{AR}}$ and $\widehat{\text{BMD}}_{\text{ER}}$ are smaller when $\lambda = 0.5$ or 1 compared with $\lambda = 0$, with the exception of Scenario 1 (which has 5% increase in RRMSE). Scenarios 6 and 8 benefit the most from the use of the power model, as there are over 10 concentrations located at the lower asymptote. In Scenarios 2 through 5, moderate improvements in RRMSEs are observed; however, coverage is reduced with the increase in λ . While a significant improvement is observed in Scenario 7, the RRMSE values for $\widehat{\text{BMD}}_{\text{AR}}$ and $\widehat{\text{BMD}}_{\text{ER}}$ still exceed 100%.

2.9. Discussion and Conclusions

In this chapter, we evaluated the statistical properties of the maximum likelihood estimates of the Hill model parameters and corresponding BMDs for qHTS data. As a screening technique, qHTS assays often employ 14- or 15-concentration designs with no replicate observations to determine whether there exists a statistically significant concentration-response relationship with the agent of interest. To evaluate the effect of the shape of the CRC on the statistical properties of the MLEs, we selected eight prototypical CRCs representing the types of CRCs might be seen in practice. Statistically, these eight CRCs were chosen so as to span the shapes demonstrated by over 8,000 actual CRCs considered previously by Sand et al. (2017). For each of the eight CRCs, the RB, CV, RRMSE, and coverage probability were evaluated by simulating 500 experimental outcomes generated using a 14-concentration experimental design with one observation at each concentration. These eight simulation scenarios provide a basis for understanding the effect of the shape of the CRCs and the choice of the experimental design on the statistical properties of the Hill model parameters. In the discussion that follows, note that in the Hill model, the parameters α and θ represent the lower and upper asymptotes

of the CRC, κ represents the half maximal concentration (which is also the inflection point of the CRC), while the steepness of the CRC is characterized by η .

Scenario 1 has one concentration at the lower asymptote, while the remaining concentrations are placed at the upper asymptote. While at least five concentrations appear to be at the upper asymptote in Scenario 2, given the smaller value of η , the remaining concentrations are located on both sides of the inflection point. Scenarios 3, 4, and 5 capture the majority of the CRC. The location κ and steepness η in Scenario 5 lead to the most well-defined CRC, relative to the experimental concentrations. With κ near the MCT, Scenarios 6 and 8 place most of the concentrations at the lower asymptote, and no observations at the upper asymptote. In Scenario 7, where $\eta = 0.3$, the experimental concentrations fail to capture both the lower and upper asymptotes.

Since α and θ represent the lower and upper asymptotes, these parameters are well estimated when there are at least two concentrations placed at the respective asymptote. As the slope of the curvature is maximized at κ , placing several experimental concentrations at both sides of the inflection point lead to smaller RRMSEs for $\hat{\kappa}$ (as in Scenarios 2, 3, 4, and 5). Our ability to estimate η well depends on the location of κ relative to the experimental concentration range and value of η relative to spacing of the experimental concentrations: this is particularly problematic when η is small, relative to the spacing of the experimental concentrations, or when η is large and κ is near the boundary of the experimental concentration range.

The MAD enables the assessment of the goodness-of-fit of the estimated CRC, as well as our ability to extrapolate outside the experimental range. Our simulation results showed that the MAD_{between} is similar in most scenarios considered, indicating a generally good fit of the Hill model within the experimental range. When there are less than two concentrations at the upper or lower asymptote, the ratios $MAD_{\text{above}}/MAD_{\text{between}}$ or $MAD_{\text{below}}/MAD_{\text{between}}$ are greater than one, indicating uncertainty above or below the experimental range, respectively.

The BMD for qHTS data can be defined in terms of additional response or extra response. Since BMR is typically chosen in the range of 1 to 10%, it is essential to have information about the lower asymptote α , as well as the lower increasing part of the CRC, including concentrations with responses beyond the BMR. Where more than eight concentrations are placed at the lower asymptote (Scenarios 5, 6, and 8), yield the smallest RRMSEs for the BMD_{AR} . The largest RRMSEs are observed when α is poorly estimated (Scenarios 1, 2, 4, and 7). Since the BMR_{ER} is defined relative to the range of the qHTS response, it requires that both α and θ be well estimated; when θ is not well estimated (Scenarios 6 and 8), the RRMSEs for the BMD_{ER} are larger than those of the BMD_{AR} .

The RRMSEs of BMD_{AR} and BMD_{ER} generally decrease with increasing BMR, decreasing error variance, and increasing number of replicates per concentration. The BMD demonstrates the lowest RRMSEs in Scenarios 5, 6, and 8, in which there are several concentrations below and above the BMD. In general, it is desirable that the experimental concentrations capture the entire CRC. Although increasing the number of replicates r will reduce the expected RRMSE by \sqrt{r} , this increase in precision will need to be balanced against the additional testing cost.

This chapter demonstrates that a strategic placement of experimental concentrations will notably improve the statistical properties of the estimators of the Hill model parameters and associated BMDs. The inference of the experimental design on the statistical properties of these estimators raises the question as to whether an optimal design that maximizes the precision and accuracy of the BMD can be constructed. This challenge is taken up in Chapter 2, in which optimal designs for qHTS assays that minimize the variance of the estimated BMD are developed.

3. Optimal Design for BMD under the Hill Model for qHTS assays

3.1. Abstract

High throughput *in vitro* tests are increasingly used in toxicity testing. In this chapter, we develop optimal experimental designs for minimizing the variance of the benchmark dose (BMD) under four-parameter Hill model used to describe continuous *in vitro* dose-response data. Under the assumption of constant error variance, the optimal design requires four experimental concentrations, including a control group at concentration zero. The placement of the optimal doses depends on the shape of the concentration-response curve (CRC), with the optimal doses chosen to provide maximum information on overall shape. Because more than the minimum number of experimental doses in the optimal design may be desirable in practice, we also consider fixed designs with between four and 15 experimental doses. Because the shape of the CRC is not known, in advance of conducting the experiment, the doses in the fixed designs are predetermined fractions of the highest dose tested. To evaluate the efficiency of the fixed designs relative to the optimal design, we analyzed over 8,000 qHTS datasets, calculating the relative efficiency (RE) of each of the fixed designs relative to the optimal design (assuming the fitted Hill model represents the true underlying CRC). The largest median REs for the fixed designs was below 22% (with average REs for these designs ranged from 5% to 24%). The results of this investigation will be helpful in guiding the development of future experimental designs for qHTS data. As prior knowledge of the shape of the CRC would be helpful in developing specific experimental designs, a two-phase empirical Bayes approach, in which the preliminary estimates of the CRC is obtained in phase 1 and used to guide the final experimental design in phase 2, would be worth exploring in the future.

3.2. Quantitative high-throughput screening

Traditionally, toxicological studies have relied heavily on animal testing to assess potential human health risk. Recently, there has been a shift towards alternative testing methodologies through *in vitro*, *in silico*, and *in chemico* analysis. Last year, the U.S. Environmental Protection Agency's (EPA) announced a decision to eliminate all mammalian toxicity testing by 2035. This is in alignment with the report *Toxicity Testing in the 21st Century: A Vision and a Strategy*, provided by the U.S. National Research Council (NRC 2007). The term new approach methodologies (NAMs) is used to describe any non-animal testing approach that can be used to provide information on chemical hazard and risk assessment (ICCVAM 2018). One particular NAM, that of quantitative high-throughput screening (qHTS), is an *in vitro* testing methodology that examines cellular responses to various concentrations of chemicals. This approach, introduced by Ingles et al. (2006), offers considerable advantages over traditional toxicity testing in terms of timeliness and cost. Unlike mammalian toxicity testing, which may require a number of years and millions of dollars to run an experiment on a single chemical (Lave and Omenn, 1986), qHTS can produce results for multiple chemicals in a matter of days.

The characterization of a concentration-response curve (CRC) for the agent of interest is an important aspect of qHTS analysis. To establish human health guidelines for potentially hazardous chemicals, the derivation and estimation of a value called the point of departure (PoD) is first required. In this chapter, we focus on the benchmark dose (BMD), which is used by many health organizations internationally (Davis et al. 2011), as the PoD for deriving human exposure guidelines. The toxicity reference value (TRV) is calculated by dividing the PoD by a suitable adjustment factor (AF); that is $TRV = PoD/AF$. This AF is usually comprised of variability caused by inter-species differences, intra-individual variations, and other relevant considerations, such as *in vitro* to *in vivo* extrapolation (IVIVE), which is used to calibrate concentrations in the cellular test system with corresponding doses in human tissues using pharmacokinetic models (Anderson and Krewski 2009; Krewski et al. 2009, 2011; NRC 2007). Although there are several models that are currently available for the characterization of a concentration-response

relationship, the Hill model is widely used for modelling qHTS data (Shockley 2015) and is used in this chapter.

3.3. Hill model for qHTS assays and corresponding BMDs

For a given concentration x (in micromolar μM), let $y(x)$ be the normalized response in the qHTS assay. Since y is a continuous variable, we will use a 4-parameter Hill model

$$y = \alpha + \frac{\theta - \alpha}{1 + (\kappa/x)^\eta}, \quad (3.1)$$

to describe the CRC. Here, α represents the cell activity level at zero concentration; θ denotes the maximum activity level at an effectively infinite concentration; κ is the concentration that corresponds to a response of $(\theta + \alpha)/2$; and η determines the shape (or steepness) of the CRC.

Suppose that an experiment with t concentrations, x_1, \dots, x_t , is conducted. The number of replicates at concentration x_i is given by n_i . Further, let $n = \sum_{i=1}^t n_i$ be the total number of trials. Then the observed response for the j^{th} replicate at concentration x_i is

$$y_{ij} = \left[\alpha + \frac{\theta - \alpha}{1 + (\kappa/x_i)^\eta} \right] + \varepsilon_{ij} \quad (i = 1, \dots, t; j = 1, \dots, n_i), \quad (3.2)$$

where ε_{ij} represents the error term associated with this observation. For each x_i , the error terms are assumed to follow a normal distribution with mean zero and variance σ_i^2 . Following Khinkis et al. (2003), the error variance is assumed to be described by the power model:

$$\sigma_i^2 = \sigma_0^2 y_i^{2\lambda} \quad (\lambda \geq 0). \quad (3.3)$$

When $\lambda = 0$, the model assumes constant variance across the concentrations, while the error terms would possess a constant coefficient of variation (CV) if $\lambda = 1$.

3.4. Maximum Likelihood Estimators

3.4.1. MLEs for the Hill Model Parameters

In this chapter, we use the method of maximum likelihood to estimate the Hill model parameters, along with the corresponding BMDs. The maximum likelihood estimators (MLEs) are determined by maximizing the likelihood function L given the observed data under the assumed distribution function for the error terms. Mathematically, the MLEs for the Hill parameters are such that

$$\hat{\boldsymbol{\phi}}' = (\hat{\alpha}, \hat{\kappa}, \hat{\eta}, \hat{\theta}) = \underset{(\alpha, \kappa, \eta, \theta)}{\operatorname{argmax}} L(\alpha, \kappa, \eta, \theta | \mathbf{y}, \mathbf{x}), \quad (3.4)$$

where $\mathbf{y}' = (y_1, \dots, y_n)$ and $\mathbf{x}' = (x_1, \dots, x_n)$.

Suppose that at each concentration x_i , the expected response μ_i follows a 4-parameter Hill model ($\mu_i = f(x_i | \boldsymbol{\phi})$). Further assume that the error terms $\epsilon_{ij} = y_{ij} - f(x_i | \boldsymbol{\phi})$ follow a normal distribution with mean zero and standard deviation σ_i , where σ_i follows the power model $\sigma_i = f(x_i | \boldsymbol{\phi})^\lambda$. The likelihood function is then given by

$$L(\boldsymbol{\phi}, \sigma | \mathbf{y}) = \prod_{i=1}^t \prod_{j=1}^{n_i} (2\pi\sigma_i^2)^{-1/2} \exp\left(-\frac{(y_{ij} - \mu_i)^2}{2\sigma_i^2}\right), \quad (3.5)$$

with log-likelihood function

$$l(\boldsymbol{\phi}, \sigma | \mathbf{y}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^t \ln(\sigma_i^2) - \sum_{i=1}^t \sum_{j=1}^{n_i} \frac{(y_{ij} - \mu_i)^2}{2\sigma_i^2}. \quad (3.6)$$

The asymptotic covariance matrix $\boldsymbol{\Sigma}_\phi$ of the Hill model parameter estimates $\hat{\boldsymbol{\phi}}$ is given by the inverse of the information matrix

$$\mathbf{X}_\phi = -E \left[\frac{\partial^2 l(\boldsymbol{\phi})}{\partial \phi_r \partial \phi_s} \right] = \mathbf{ADA}', \quad (3.7)$$

where

$$\mathbf{A} = \begin{bmatrix} f_\alpha(x_1) & f_\alpha(x_2) & \cdots & f_\alpha(x_t) \\ f_\kappa(x_1) & f_\kappa(x_2) & \cdots & f_\kappa(x_t) \\ f_\eta(x_1) & f_\eta(x_2) & \cdots & f_\eta(x_t) \\ f_\theta(x_1) & f_\theta(x_2) & \cdots & f_\theta(x_t) \end{bmatrix}, \quad (3.8)$$

and

$$\mathbf{D} = \text{diag} \left\{ \frac{n_i}{\sigma_i^2} \right\}. \quad (3.9)$$

Thus, an estimate of the covariance matrix of $\hat{\boldsymbol{\phi}}$ is given by $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\phi}}} = \mathbf{X}_{\hat{\boldsymbol{\phi}}}^{-1} = (\mathbf{A}\mathbf{D}\mathbf{A}')^{-1}|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}$.

Once the MLEs of the Hill model parameters have been obtained, corresponding estimates, \hat{x}_0 and $\hat{\pi}_0$, of the desired BMD and BMR respectively, can be calculated by substituting $\hat{\boldsymbol{\phi}}$ in place of $\boldsymbol{\phi}$.

3.4.2. MLEs for BMR and BMD

Krewski & Ryzin (1981) showed that, under mild regularity conditions (see Appendix B.1), the asymptotic variance of the estimated BMR can be approximated by

$$V(\hat{\pi}_0) = \boldsymbol{\Pi}' \mathbf{X}_{\hat{\boldsymbol{\phi}}}^{-1} \boldsymbol{\Pi} |_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}, \quad (3.10)$$

where $\boldsymbol{\Pi}' = (\Pi_1, \dots, \Pi_t)$ with $\Pi_r(x) = \frac{\partial}{\partial \phi_r} \left[\frac{\theta - \alpha}{1 + (\kappa/x)^\eta} \right]$ for BMR_{AR} and $\Pi_r(x) = \frac{\partial}{\partial \phi_r} \left[\frac{1}{1 + (\kappa/x)^\eta} \right]$ for BMR_{ER} .

Note that for a pre-specified $\text{BMR} > 0$, we have

$$x_{0,\text{AR}} = \text{BMD}_{\text{AR}} = \frac{\left(\frac{\theta - \alpha}{\text{BMR}_{\text{AR}}} - 1 \right)^{1/\eta}}{\kappa}, \quad (3.11)$$

and

$$x_{0,\text{ER}} = \text{BMD}_{\text{ER}} = \frac{\left(\frac{1}{\text{BMR}_{\text{ER}}} - 1 \right)^{1/\eta}}{\kappa}. \quad (3.12)$$

Using the delta-method, the asymptotic variance of the corresponding BMD estimate is given by

$$V(\hat{x}_0) = \left[\frac{\partial f(x)}{\partial x} \right]_{x=\hat{x}_0}^{-2} V(\hat{\pi}_0), \quad (3.13)$$

where

$$\frac{\partial f(x)}{\partial x} = \frac{\eta(\theta - \alpha)(\kappa/x)^\eta}{x(1 + (\kappa/x)^\eta)^2}. \quad (3.14)$$

Thus, we have

$$\text{Var}(\hat{x}_0) = \left[\frac{\eta(\theta - \alpha)(\kappa/\hat{x}_0)^\eta}{\hat{x}_0(1 + (\kappa/\hat{x}_0)^\eta)^2} \right]^{-2} \mathbf{\Pi}' \mathbf{X}_{\hat{\phi}}^{-1} \mathbf{\Pi}. \quad (3.15)$$

Since

$$\hat{x}_0 = \pi^{-1}(\pi_0 | \hat{\phi}), \quad (3.16)$$

asymptotically, we have

$$\text{BMD} \sim E[\pi^{-1}(\pi_0 | \hat{\phi})] \sim \pi^{-1}(\pi_0 | \phi). \quad (3.17)$$

Hence, (3.17) is constant under a given set of ϕ . This implies that minimizing the variance of $\hat{\pi}_0$ is equivalent to minimizing the variance of corresponding \hat{x}_0 , so that an optimal design for BMR is also an optimal design for BMD.

3.5. Optimal Design for Estimating the BMD

Chernoff's theorem (Chernoff 1952) states that if the information matrix $\mathbf{X}_{\hat{\phi}}$ is closed and bounded at any $x \in [0, X]$ under any given set of $\phi \in \Phi$, then there exists a set of concentration levels x , which minimizes the variance of the estimator $\hat{\pi}_0$. Chernoff's theorem also tells us that such an optimal design would require four experimental concentrations corresponding to the number of Hill model parameters being estimated (see Appendix B.2).

For experiments with four concentrations, we have

$$\mathbf{A}^{-1} = \{a^{ir}\} \quad (3.18)$$

($i = 1, \dots, 4$; $r = 1, \dots, 4$) with

$$\mathbf{D}^{-1} = \text{diag}\{\sigma_i^2/n_i\}. \quad (3.19)$$

Hence,

$$\Sigma_{\hat{\phi}} = \begin{bmatrix} \sum_{i=1}^4 (a^{i1})^2 \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 a^{i2} a^{i1} \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 a^{i3} a^{i1} \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 a^{i4} a^{i1} \frac{\sigma_i^2}{n_i} \\ \sum_{i=1}^4 a^{i1} a^{i2} \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 (a^{i2})^2 \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 a^{i3} a^{i2} \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 a^{i4} a^{i2} \frac{\sigma_i^2}{n_i} \\ \sum_{i=1}^4 a^{i1} a^{i3} \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 a^{i2} a^{i3} \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 (a^{i3})^2 \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 a^{i4} a^{i3} \frac{\sigma_i^2}{n_i} \\ \sum_{i=1}^4 a^{i1} a^{i4} \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 a^{i2} a^{i4} \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 a^{i3} a^{i4} \frac{\sigma_i^2}{n_i} & \sum_{i=1}^4 (a^{i4})^2 \frac{\sigma_i^2}{n_i} \end{bmatrix}, \quad (3.20)$$

which leads to

$$V(\hat{\pi}_0) = \mathbf{\Pi}' \Sigma_{\hat{\phi}} \mathbf{\Pi} = \sum_{i=1}^4 \frac{\sigma_i^2}{n_i} \left(\sum_{j=1}^4 \Pi_j a^{ij} \right)^2. \quad (3.21)$$

To find the values of n_1, \dots, n_4 that minimize the variance of $\hat{\pi}_0$ holding $\mathbf{x}' = (x_1, x_2, x_3, x_4)$ fixed, let

$$Q = V(\hat{\pi}_0) = V(\hat{\pi}_0) + \psi \left(n - \sum_{i=1}^4 n_i \right) = \sum_{i=1}^4 \frac{\sigma_i^2}{n_i} \left(\sum_{j=1}^4 \Pi_j a^{ij} \right)^2 + \psi \left(n - \sum_{i=1}^4 n_i \right). \quad (3.22)$$

Taking the partial derivatives of Q with respect to n_i and ψ yields

$$\frac{\partial Q}{\partial n_i} = \frac{\sigma_i^2}{n_i^2} \left(\sum_{j=1}^4 \Pi_j a^{ij} \right)^2 + \psi, \quad (3.23)$$

and

$$\frac{\partial Q}{\partial \psi} = \left(n - \sum_{i=1}^4 n_i \right). \quad (3.24)$$

Setting (3.23) to zero and solving for n_i gives

$$n_i = \sqrt{\frac{\sigma_{j=1}^2 (\sum_{j=1}^4 \Pi_j a^{ij})^2}{\psi}}. \quad (3.25)$$

Since $n = \sum_{i=1}^4 n_i$, the asymptotically optimal choices of the sample proportions $c_i = n_i/n$ are given by

$$c_i^* = \frac{\sigma_i |\sum_{j=1}^4 \Pi_j a^{ij}|}{\sum_{h=1}^4 (\sigma_h |\sum_{j=1}^4 \Pi_j a^{hj}|)}. \quad (3.26)$$

Given the optimal sample proportions $\mathbf{c}^* = (c_1^*, c_2^*, c_3^*, c_4^*)'$, a numerical search can be used to find the optimal experimental concentrations $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*)'$.

Substituting (3.26) into (3.21), the variance of $\hat{\pi}_0$ is

$$V(\hat{\pi}_0) = \sum_{i=1}^4 \left(\sigma_i \left| \sum_{j=1}^4 \Pi_j (\hat{x}_0) a^{ij} \right| \right). \quad (3.27)$$

3.6. Properties of the Optimal Designs

The optimal experimental concentration vector $\mathbf{x}^* = (x_1^*, x_2^*, x_3^*, x_4^*)'$ is determined using a numerical grid search. The grid is given in 0.1% increment in extra response between 0.1% to 99.9%. For each of the lower and upper asymptotes below 0.1% and above 99.9%, respectively, a grid with five increments between 10^{-8} and 10^{-4} , spread equally in the log-scale, is used. To validate that an OD follows a four-concentration experiment, (1) no pair of tested concentrations must be adjacent to each other, and (2) each of the sample proportions, c_i must be at least 1%.

The theoretical optimal designs (ODs) for the BMD_{ER} set at 5% for the eight prototypical CRCs considered in chapter 1 are almost invariant with respect to the optimal response rates under the assumption of constant error variance, regardless of the values of the Hill model parameters for these eight CRCs. Specifically, the optimal design includes one concentration of $x_1^* \cong 0$ corresponding to the control group with $ER_{x_1^*} = 0$, and another concentration x_4^* within the upper asymptote with $ER_{x_4^*}$ approaching one. The remaining two concentrations x_2^* and x_3^* correspond to ERs of $ER_{x_2^*} = 21.8\%$ and $ER_{x_3^*} = 78.2\%$ (symmetric about κ on the log-scale, with $\log(\kappa) = \log(x_2^*x_3^*)/2$), respectively. The corresponding optimal sample proportions are $\mathbf{c}^* = (31.9\%, 45.8\%, 18.1\%, 4.2\%)$, regardless of the underlying Hill model parameters.

These optimal designs provide a useful benchmark against which sub-optimal designs can be compared in terms of relative efficiency (RE). Specifically, the efficiency of the design D_2 relative to design D_1 with respect to estimating the BMD is defined as $RE_{1,2} = Var(\widehat{BMD})|_{D_2} / Var(\widehat{BMD})|_{D_1}$. We will use RE as a basis for evaluation of sub-optimal fixed designs that may be considered for use in practice in the absence of perfect knowledge of the underlying CRC required to determine the optimal design.

In practice, the maximum concentration tested (MCT) may not be equivalent to the maximum concentration in the theoretical OD. (In order to avoid excessively high concentrations that would not provide biologically meaningful results, the MCT is often less than x_4^* .) Consequently, we will consider the constrained optimal design (COD) in which $x_4 \leq MCT$. The relative efficiencies (REs) of the CODs is summarized in Figure 3.1, for decreasing extra response ER_{MCT} corresponding to decreasing upper limits on the experimental concentrations given by the MCT. These results show that the efficiency of the COD decreases notably as the MCT decreases. When the MCT is less than 93% of the ER, the OD places almost all of the weight on three of the four experimental concentrations in the COD, with x_3^* and x_4^* being very close when ER_{MCT} is in the range of 70-90%, x_2^* and x_3^* often being very close when ER_{MCT} ranges from 52-69%. Note also

that the lowest concentration x_1^* in the COD is somewhat greater than zero when ER_{MCT} is 53% or less, thereby eliminating a control group from the COD.

3.7. Relative Efficiency of Fixed Designs in Practice

3.7.1. Discussion of Fixed Designs

Derivation of the optimal design requires knowledge of the Hill model parameters. Since this information is not typically available in practice, we will consider a series of fixed designs (FDs) with prespecified experimental concentrations and sample allocations. Specifically, we investigate the REs of the FDs relative to the CODs given for six FDs, each of which includes a control group at concentration of zero and the highest concentration at (or close to) the MCT. FD_1 is a 15-concentration experimental design that is commonly used in practice. As seen in Table 3.1, this design places more concentrations in the higher experimental range. FD_2 also employs 15 experimental concentrations, equally spaced in log-scale. FD_3 and FD_4 use 10 and 5 concentrations, respectively, equally spaced in log-scale. This allows us to investigate whether reducing the number of concentrations might increase efficiency (as shown previously, COD includes exactly four experimental concentrations). The fifth and sixth fixed designs are average constrained optimal designs (ACODs), derived from 8,456 datasets considered in Sand et al. (2017). $ACOD_{ave}$ (FD_5) is determined by taking the average experimental concentrations from the 8,456 CODs, while $ACOD_{med}$ (FD_6) is constructed using the median experimental concentrations.

3.7.2. Comparing Relative efficiencies

Table 3.2 and Figure 3.2 summarize the REs for the six FDs in estimating BMD_{ER} across the 8,456 datasets. Due to the wide variation in overall shape, it is difficult for a single fixed design to do well across the entire range of CRCs considered by Sand et al. (2017). Because the distribution of REs shown in the boxplot are right skewed for all six FDs, comparisons among the six designs will be largely based on the median RE in what follows. FD_1 performed the best with $RE_{med,FD_1} = 22.2\%$, followed by FD_2 with $RE_{med,FD_2} = 18.3\%$. Both FD_1 and FD_2 use 15 experimental concentrations, with FD_1 placing more

concentrations in the higher experimental range. Given that the majority of the observed CRCs do not depart from the lower asymptote at smaller doses, placement of the experimental concentrations is most effective in FD_1 .

The remaining designs have fewer experimental concentrations, with more than one observation at some of the design points. The $ACOD_{ave}$ and $ACOD_{med}$ designs perform better than FD_3 and FD_4 because they tend to better capture curvature around the BMD more often. While the fixed designs FD_1 and FD_2 show better performance on average, the maximum RE achieved of 50% (with FD_1) is lower than the REs realized with $ACOD_{med}$ in almost a quarter of the cases.

Figure 3.3 displays boxplots of the differences in the REs of $ACOD_{ave}$ and $ACOD_{med}$ relative to FD_1 . This paired comparison allows us to study the frequency and the magnitude of the differences in REs across the 8,456 datasets. While the median difference in the REs between FD_1 and $ACOD_{med}$ is 4.4% the boxplot reveals that in 25% of the experiments, the RE from $ACOD_{med}$ is greater than that of FD_1 by 22.3% to 88.0%. FD_1 is better than $ACOD_{med}$ by 16.8% to 50.0% in 25% of the experiments. The results for $ACOD_{ave}$ are similar to those for $ACOD_{med}$, although $ACOD_{med}$ performed slightly better than $ACOD_{ave}$. Similar findings are observed when the BMD is defined with respect to AR (Table 3.4 and Figure 3.4).

Because of singularities, the differences in the average REs are necessarily based on experiments for which the convergent optimal design was obtained. Although there were no singularities for FD_1 and FD_2 , less than approximately 2% singularities for FD_3 , $ACOD_{ave}$, and $ACOD_{med}$, FD_4 resulted in almost 60% non-convergent designs, so that the differences in average REs for FD_4 are based on only 40% of the 8,456 experiments. The large number of singularities for FD_4 may be attributed to the dose spacing in this design, which is inadequate to effectively capture the curvature in the majority of these datasets.

This suggests that strategic placement of experimental concentrations could be beneficial, and a two-phase Bayesian approach could be considered.

3.8. Discussion and Conclusions

In this chapter, we developed theoretical optimal designs that minimize the variance of the maximum likelihood estimates of the BMD for qHTS data using a 4-parameter Hill model to describe the CRC.

For the BMD_{ER} corresponding to an extra benchmark response BMR_{ER} of 5% as the PoD, the theoretical OD includes a control group at $x_1^* = 0$, with the maximum experimental concentration x_4^* in the upper asymptote of the CRC (with $ER_{x_4^*}$ approaching 1). The two intermediate concentrations x_2^* and x_3^* are such that $ER_{x_2^*} = 21.8\%$ and $ER_{x_3^*} = 78.2\%$. Under the assumption of the constant error variance, these optimal design points are invariant with respect to the underlying Hill model parameters.

As practical constraints may preclude using an MCT as high as x_4^* , a COD subject to $x_4 \leq MCT$ that minimizes the variance of \widehat{BMD}_{ER} may be determined. When ER_{MCT} is less than 93%, the COD may include adjacent doses that are very close to each other. The efficiency of the COD was seen to decrease with MCT, for example, $RE < 10\%$ when $ER_{MCT} = 40\%$, because of the reduced coverage of the overall shape of the CRC.

Derivation of both the theoretical OD and COD requires knowledge of the Hill model parameter values. In the absence of the prior knowledge of the Hill model parameters, practitioners have used fixed designs with up to 15 experimental concentrations. To explore the efficiency of alternative fixed designs, we calculated the RE of six FDs relative to the COD for over 8,000 actual CRCs previously considered by Sand et al. (2017). These fixed designs included a commonly used design with 15 experimental concentrations, and a variant of this design with 15 concentrations equally spaced in log scale. Two additional designs with fewer (10 and 5) experimental concentrations are also considered. The final two designs were determined by taking the average (or median) experimental concentrations in the CODs for the 8,456 CRCs.

Our results show that the ACODs considered in this chapter have the potential to do better than the current 15-dose design that is widely used in practice. However, even the ACODs do not perform uniformly well because of the wide variation in the shapes of CRCs

seen in practice. If prior information on the shape of the CRC was available, a COD could be developed based on this information. Absent such prior information, a two-phase approach in which a small pilot experiment is conducted to obtain preliminary information on the CRC could be considered.

In such a two-phase approach, the total sample size of n is divided into two groups of sizes n_1 and n_2 , where samples in the first group are used in phase 1 and the remaining samples held for phase 2. The design employed in phase 1 should include a large number of experimental concentrations with a minimal number of replicates at each concentration (similar to FD_1), in order to gain preliminary information about the overall shape of the CRC. The estimates of the Hill model parameters, $\hat{\phi}_1$, can then be used to develop the phase 2 design using the optimality criteria presented in this chapter. Upon completion of phase 2, the updated Hill model parameter estimates $\hat{\phi}_2$ can then be used to estimate the BMD. Although the full treatment of the this two phase approach is outside of the scope of this chapter, the correlation between the experimental data obtained in phase 1 and phase 2 will need to be considered in a formal derivation of the two phase optimal design.

4. Use of Value of Information Analysis in Evaluating the Effects of Timeliness, Uncertainty Reduction, and Throughput of Information Collection Procedures

4.1. Abstract

A number of investigators have explored the use of value of information (VOI) analysis to evaluate alternative information collection procedures in different decision-making contexts. This chapter presents an analytic framework for valuing information in support of risk decisions involving toxic substances, providing a set of methods for optimizing efforts to gather, process, and apply information to help decision-makers achieve their goals. The framework can be used to understand the trade-offs between toxicological tests that can be used to accelerate the pace of chemical risk assessment in comparison with more resource intensive testing strategies. The use of our framework is demonstrated by three illustrative applications which, although based on simplified assumptions, show the types of insights that can be obtained through the use of our methods in VOI analysis. Specifically, our results suggest timeliness of information collection can improve the VOI associated with testing a single chemical, even when the reduction in uncertainty is less. Our analysis also suggests that high-throughput testing may be beneficial in terms of public health benefits by increasing the number of substances that can be evaluated within a given budget.

4.2. Introduction

In addressing public and population health risk issues, it is important that decision-makers be provided with the best possible scientific evidence to support risk decision. In the

presence of uncertainty, decision-makers must weigh the likelihood of making an incorrect (or sub-optimal) decision and the magnitude of the expected loss in choosing such a course of action. VOI analysis provides an approach for systematically assessing the trade-offs between the benefits of acquiring additional information (such as reduction in uncertainty, which may lead to reduction in probability of making incorrect decision) and the cost of collecting such information. Specifically, VOI analysis aims to answer the following questions: Does currently available information provide sufficient evidence to allow a decision to be made with confidence (e.g., whether a chemical should be regulated, and if so, by how much)? If not, does the benefit of collecting additional information outweigh the cost of collecting such information? And, when there are multiple information collection methodologies, which method is best?

There are a number of ways to define the expected VOI (EVOI) to determine whether the collection of additional information is beneficial. If the additional information eliminates the uncertainty inherent in the decision context, then such information allows decision-makers to make an optimal decision with certainty. This notion of EVOI is referred to as the expected value of perfect information (EVPI). While the EVPI is unattainable in practice, it can be used as an *upper limit* on the value of collecting additional information.

In many cases, the risk models underlying risk decisions will involve a number of parameters, each of which is subject to uncertainty. The overall structure of the uncertainty associated with a given risk context can lead to joint uncertainty distributions that are complex. Since additional information is often relevant to only a subset of uncertainty parameters, the expected value of partial perfect information (EVPPI) may be used to determine which parameter(s) should be targeted in attempts to reducing uncertainty. As with the EVPI, the EVPPI is also unattainable in practice, but provides a benchmark defining the maximum benefit that could be achieved by reducing uncertainty in key parameters.

While VOI metrics focusing on perfect information provide theoretical upper limits on which might be achieved through the collection of additional information, other

measures of the value of additional information based on what is practically achievable are available for use in practice. The expected value of sample information (EVSI) provides a realistic estimate against which to compare the information collection cost (ICC), and to determine whether the additional information obtained is worthwhile. The expected net benefit of sampling (ENBS), defined as the difference between the EVSI and the ICC, as well as the return on investment (ROI), calculated as the ratio of ENBS and ICC, also provide useful metrics for evaluating VOI in practice.

Historically, most VOI applications in the health sciences prior to 1995 were based on uncertainty distributions that were either discrete or discretized (Yokota and Thompson, 2004). A classic example is provided by work of Lave et al. (1988), who evaluated the cost-effectiveness of lifetime rodent bioassays against no testing using EVSI calculated based on a 2x2 contingency table (positive/negative human carcinogenicity cross-classified with positive/negative lifetime rodent bioassays).

As VOI methodologies advanced, computational trends shifted towards simulation-based analysis. For example, Thompson and Evans (1997) investigated the value of collecting additional exposure, toxicological, and pharmacokinetic data for perchloroethylene, a popular dry-cleaning solvent. The decision context was more complex than the discrete VOI analyses mentioned above. The affected population was partitioned into four groups based on potential exposure levels (workers, families of workers, customers, and general public): and various information collection schemes were evaluated using EVPI and EVPPI, comparing factors such as optimization by dry cleaning facility or machine type, with varying ICCs.

While VOI methodologies such as the EVPPI and EVSI have been shown to provide more realistic and informative estimates compared to the EVPI, their implementation has been limited due to conceptual and computational complexity. Keisler et al. (2014) found that studies using EVSI are associated with fewer uncertainties and alternative courses of action, compared to studies with measures such as EVPI. Tuffaha et al. (2016) compared non-parametric regression approaches to calculating EVPPI and EVSI to the traditional

Monte Carlo approach and showed that non-parametric regression could dramatically reduce the computational time, oftentimes leading to VOI estimates with smaller standard errors.

While a highly useful approach to evaluating the benefits of collecting additional information to support risk decisions, VOI analysis is by no means a simple undertaking. Andronis (2015) notes that valuation of consequences such as willingness-to-pay (WTP) for quality-adjusted-life-years (QALYs) gained, and value of a statistical life (VSL) must be chosen with care, as these choices can have a strong influence on the outcome of the analysis. Koffijberg et al. (2018a) discusses additional factors that may affect both the outcome and the validity of the analysis, including: correlations amongst input parameters; uncertainty in model structure; and choice of time horizon and discount rate. Koffijberg et al. (2018b) further notes that external constraints such as the maximum allowable control cost and health equity considerations may alter the outcome of the analysis as well.

Although VOI represents a comparatively new analytic approach for determining the value in collecting additional information in support of more informed risk decisions, it is becoming increasingly applied in practice. A systematic review by Keisler et al. (2014) documented applications of VOI in a wide range of disciplines, including agriculture, anthropology, chemistry, defense, ecology, economics, education, energy, environmental science, geology, information science, infrastructure, medicine, and transportation. Based on the collective experience accumulated to date with this methodology, VOI would appear to be a valuable tool for evaluating the benefits gained by collecting additional information to strengthen risk decision making in the presence of uncertainty.

A key aspect of public health decision-making with respect to potentially toxic chemicals present in the human environment involves the choice between making an immediate decision with currently available information versus delaying a decision until additional data is collected and analyzed. This choice is often informed by the urgency of the public health need and the costs, in terms of both time and resources, of acquiring additional

relevant information that may lead to better decisions. Given that the toxicological properties of only a small fraction of some 140,000 chemicals currently available in commerce have been investigated in detail, it is important to evaluate the trade-offs between timeliness, cost, and the degree of uncertainty reduction that various testing strategies offer.

The goal of this chapter is to present an algorithm for VOI analysis that can be used to compare alternative information collection techniques. Our analytic framework for VOI analysis is presented in section 4.2. Although an in-depth evaluation of the VOI associated with particular testing methodologies is outside the scope of this chapter, an illustrative prototypical application of how this can be done is provided in section 4.3. A summary and discussion of our progress to date is given in section 4.4, along with suggestions for further work in this area.

4.3. Methodology

4.3.1. Risk

Let x denotes the exposure level to a chemical (mg/kg/day). Furthermore, let $f_{\text{exp}}(x|\boldsymbol{\theta})$ be the probability density function of the exposure and $g(x|\boldsymbol{\theta})$ be the toxicity of a chemical.

In this chapter, the population average risk of a chemical R is given as a function of its toxicity and population exposure. Let $\boldsymbol{\theta}$ denote the set of parameters that determine the risk level. The toxicity of a chemical $g(x|\boldsymbol{\theta})$ is defined as the probability of an adverse effect to be present at a given exposure level y . Mathematically, this means that

$$g(x|\boldsymbol{\theta}) = P(\text{Adverse effect present} | X \leq x, \boldsymbol{\theta}). \quad (4.1)$$

Let $f_{\text{exp}}(\cdot)$ denote the probability density of the population exposure to a chemical. Thus, the population average risk R can be derived by

$$R = R(\boldsymbol{\theta}) = E[g(x|\boldsymbol{\theta})] = \int_0^{\infty} g(x|\boldsymbol{\theta}) f_{\text{exp}}(x|\boldsymbol{\theta}) dy. \quad (4.2)$$

To calculate the risk, let us suppose that θ can be partitioned into $\theta = [\theta_{\text{tox}}, \theta_{\text{exp}}]$, where θ_{tox} represents a set of parameters that govern the toxicity, and θ_{exp} controls the exposure. Suppose further that θ_{tox} and θ_{exp} are statistically independent. This assumption seems reasonable, since the prevalence rate of an untested chemical should not be related to its toxicity, and vice versa. Then, the risk can be expressed as

$$R = \int_0^{\infty} g(x|\theta_{\text{tox}})f_{\text{exp}}(x|\theta_{\text{exp}})dy. \quad (4.3)$$

Following Chiu et al. (2018), we assume that lognormal distribution can be used to approximate both the toxicity and exposure distributions. Since the lognormal distributions can be completely described by its mean and standard deviation, this assumption implies that $g(x|\theta_{\text{tox}}) \sim LN(\mu_{\text{tox}}, \sigma_{\text{tox}})$ and $f_{\text{exp}}(x|\theta_{\text{exp}}) \sim LN(\mu_{\text{exp}}, \sigma_{\text{exp}})$, and thus $\theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\mu_{\text{tox}}, \sigma_{\text{tox}}, \mu_{\text{exp}}, \sigma_{\text{exp}})$. This also implies that the risk can be further simplified to

$$R = \Phi\left(\frac{\mu_{\text{exp}} - \mu_{\text{tox}}}{\sqrt{\sigma_{\text{exp}}^2 + \sigma_{\text{tox}}^2}}\right), \quad (4.4)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

4.3.2. Information Collection and Bayesian Updating

The lack of perfect information leads to uncertainty distribution about the risk parameters. Suppose that $\pi(\varphi)$ denotes the prior uncertainty distribution based on currently available information about θ . Suppose further that additional information $s \in \mathcal{S}$ is collected in hopes to reduce uncertainty about θ . Then, the uncertainty distribution may be updated using the above information. Following Bayes' rule, the updated (i.e., posterior) uncertainty distribution $\kappa(\theta|s)$ can be given by

$$\kappa(\boldsymbol{\theta}|S = s) = \frac{L(s|\boldsymbol{\theta})\pi(\boldsymbol{\varphi})}{h(s)}, \quad (4.5)$$

where $L(s|\boldsymbol{\theta})$ is the likelihood function of $S = s$ given $\boldsymbol{\theta}$, and $h(s)$ is the unconditional probability density function of S at s .

Assume further that prior uncertainty distributions for each θ_i follows a lognormal distribution with parameters $(\mu_{\theta_i}^0, \sigma_{\theta_i}^0)$ for $i = 1, 2, 3, 4$. If the additional information $S|\boldsymbol{\theta}$ also follows a lognormal distribution with parameters (μ_S, σ_S) , then

$$\kappa^j(\theta_i|S^j) \sim LN(\mu_{\theta_i}^j, \sigma_{\theta_i}^j), \quad (4.6)$$

where $j = 1, \dots, J$ denotes the different information collection methodology, and

$$\mu_{\theta_i}^j = (\sigma_{\theta_i}^j)^2 \left(\frac{\mu_{\theta_i}^0}{(\sigma_{\theta_i}^0)^2} + \frac{S^j}{\sigma_{S^j}^2} \right), \quad (4.7)$$

and

$$(\sigma_{\theta_i}^j)^2 = \left(\frac{\sigma_{S^j}^2 (\sigma_{\theta_i}^0)^2}{\sigma_{S^j}^2 + (\sigma_{\theta_i}^0)^2} \right). \quad (4.8)$$

4.3.3. Objective Functions

In its most general form, VOI analysis seeks to minimize the objective function (OF), and it is based on the decision context and criteria. In VOI literature, the annualized social cost (ASC) is often used as the objective function (such as Taylor et al. 1993). It should also be noted that, however, decision-makers may be required to determine appropriate action solely based on controlling the risk. In what follows, we will discuss potential objective functions that may be used by benefits-risks decision-maker (BRDM) and target-risk decision-maker (TRDM).

4.3.3.1. Annualized Social Cost (ASC)

Suppose that there are $K + 1$ potential regulatory actions to take based on the decision (where $k = 0$ implies no regulatory action is taken). Then, the social cost can be partitioned into two parts: control cost (CC) and health cost (HC). The control cost incurs as a result of decision that risk must be mitigated, and the health cost results from loss due to adverse effect. Let CC_k and HC_k denote the amount of annual control cost (per capita) and the annual health cost (per capita) associated with reducing the exposure by k^{th} regulatory action, respectively. Then, at any given time passed y , the annualized social cost (ASC) can be expressed as:

$$ASC_k = CC_k + HC_k = \frac{N_y \times C_k}{(1+r)^y} + \frac{N_y \times V(R - R_k)}{(1+r)^y} = \frac{N_y [C_k + V(R - R_k)]}{(1+r)^y}, \quad (4.9)$$

where

- N_y is the number of exposed populations at time y
- C_k is the annual cost of control (per capita) due to reduced exposure via k^{th} regulatory action
- V represents the value of a statistical life
- R is the risk of the adverse effect (in particular, death) due to exposure without control strategy
- R_k is the risk of the adverse effect (in particular, death) due to reduced exposure via k^{th} regulatory action
- r is the discount rate

4.3.3.2. Total Social Cost (TSC)

Calculation of the total social cost (TSC) requires specification of a number of inputs for accurate estimation of the VOI. Suppose that there are $J + 1$ information collection methodologies (where $j = 0$ implies no additional information is collected). For each method $j = 0, \dots, J$ and regulatory action $k = 0, \dots, K$, we need the following information: the time required to collect information via the j^{th} method ($t_{IC,j}$); the time

required to assess the information collected and make a decision via the j^{th} method ($t_{DM,j}$); the time required to implement the k^{th} regulatory action once the decision is made ($t_{imp,k}$); and the time required to implement this decision ($t_{eff,k}$).

Then, over a predetermined time horizon y_{TH} , the control cost incurs once the regulatory action is implemented (i.e., $y_{imp,j,k} = t_{IC,j} + t_{DM,j} + t_{imp,k}$) and the reduction in health cost is observed after $y_{eff,j,k} = y_{imp,j,k} + t_{eff,k}$. The ICC must be added to account for the cost of testing.

$$TSC_k^j = CC_{j,k} + HC_{j,k}$$

$$= \sum_{y=y_{imp,j,k}}^{y_{TH}} \frac{N_y \times C_k}{(1+r)^y} + \left[\sum_{y=1}^{y_{eff,j,k}^{-1}} \frac{N_y \times VR}{(1+r)^y} + \sum_{y_{eff,j,k}}^{y_{TH}} \frac{N_y \times VR_k}{(1+r)^y} \right]. \quad (4.10)$$

Here,

- $y_{imp,j}$ is the time it takes for implement the decision based on j^{th} information collection methodology
- $y_{eff,j,k}$ is the time required for the regulation effect to take place from k^{th} regulatory action based on information collected via j^{th} methodology
- y_{TH} is the time horizon (e.g., 10 years, 20 years, or 50 years.)

The TSC introduces a new dimension to the objective function compared to the ASC, time component. In practice, it is important to consider the time required to achieve a decision and implement actions based on it, rather than simply comparing the amount of uncertainty reduction.

4.3.4. Decision Rules

The TRDM's risk management goal is to control exposure to the substance of interest so that the average population risk R does not exceed a predetermined risk threshold called

target risk level (TRL). Thus, if we have perfect information about R , the decision rule $D(R)$ can be specified as:

$$D_{\text{TRL}}(R) = \begin{cases} 1 & \text{if } R > \text{TRL} \\ 0 & \text{if } R \leq \text{TRL} \end{cases}, \quad (4.11)$$

with corresponding action $A(R)$

$$A_{\text{TRL}}(R) = \begin{cases} R \rightarrow \text{TRL} & \text{if } D_{\text{TRL}}(R) = 1 \\ R \rightarrow R & \text{if } D_{\text{TRL}}(R) = 0 \end{cases}, \quad (4.12)$$

Under this decision paradigm, the TRDM would reduce exposure so that the TRL is achieved if the current risk exceeds the target risk.

In contrast, the BRDM would engage in exposure mitigation if the economic value of the public health benefits of exposure mitigation is greater than the control cost. This decision rule is specified by:

$$D(R) = \begin{cases} 1 & \text{if } \exists R' < R \ni OF(R') < OF(R) \\ 0 & \text{otherwise} \end{cases}, \quad (4.13)$$

and

$$A(R) = \begin{cases} R \rightarrow R^* \ni R^* = \text{argmin } OF(R) & \text{if } D(R) = 1 \\ R \rightarrow R & \text{if } D(R) = 0 \end{cases}. \quad (4.14)$$

4.3.5. Incorporating Uncertainty

4.3.5.1. Uncertainty Distribution with Reduced Exposure

Let ε_k denote the reduction in mean exposure associated with regulatory action k . The resulting risk, R_k , is then calculated as:

$$R_k = \int_0^{\infty} g(y|\boldsymbol{\theta}_{\text{tox}})[f_{\text{exp}}(y|\boldsymbol{\theta}_{\text{exp}}) - \varepsilon_k] dy. \quad (4.15)$$

4.3.5.2. Expected Total Social Cost (ETSC)

In the presence of uncertainty, the true risk is unknown. Thus, given an uncertainty distribution on R , based on prior information, $f^0(R)$, the ETSC is given by:

$$\text{ETSC}_k^0(R) = E \left(\text{TSC}_k^0(R) \right) = \int \text{TSC}_k^0(R) f^0(R) dR. \quad (4.16)$$

Similarly, the posterior expected total social cost using the j^{th} information collection methodology is given by:

$$\text{ETSC}_k^j(R) = E \left(\text{TSC}_k^j(R) \right) = \int \text{TSC}_k^j(R) f^j(R) dR = \int \left[\int \text{TSC}_k^j(R|s) f^j(R|s) dR \right] f^j(s) ds. \quad (4.17)$$

This implies that, if a decision-criterion is to minimize the TSC, the optimal prior (or j^{th} posterior) decision k^j is:

$$k^j = \underset{k}{\text{argmin}} \text{ETSC}_k^j(R). \quad (4.18)$$

4.3.5.3. Expected Value of Perfect Information (EVPI)

In previous section, we discussed how to determine optimal prior and posterior actions under the assumption that the decision-maker is interested in minimising the TSC. Let $k^{j*} = k^{j*}|R = \underset{k}{\text{argmin}} \text{TSC}_k^j(R)$, that is, k^{j*} is the locally optimal action within the j^{th} information collection framework for a given risk R . The EVPI is then defined as the difference between the optimal prior (or posterior) TSC and the locally optimal TSC, integrated over the uncertainty distribution of R :

$$\text{EVPI}^j = \int \left[\text{TSC}_{k^0}^0(R) - \text{TSC}_{k^{j*}}^j(R) \right] f^j(R) dR. \quad (4.19)$$

In practice, it is impossible to remove uncertainty completely, and hence the EVPI is only of theoretical interest. It can, however, be used as an upper limit on the cost of collecting of further information, over and above the cost of information obtained via the j^{th} method.

4.3.5.4. Expected Value of Sample Information (EVSI)

Although it is impossible to remove uncertainty altogether, it is possible to reduce uncertainty with additional information. Since sampling variability will affect decisions resulting from the use of updated information, the EVSI can be used to obtain a realistic estimate of the value of information provided by additional information through toxicity test data. If we are interested in comparing the EVSI of j^{th} information collection methodology against prior uncertainty distribution, then

$$\text{EVSI}^j = \int \left[\text{TSC}_{k^0}^0(R) - \text{TSC}_{k^j}^j(R) \right] f^j(R) dR. \quad (4.20)$$

Here, the EVSI is defined as the expected difference between the TSCs calculated using the prior optimal action k^0 and the posterior counterpart k^j . Note that the former assumes no delay is incurred due to information collection, while the latter incorporates such a delay.

It is also possible to calculate the differential EVSI between two risk management options. Suppose that the decision maker is considering two alternative risk reduction strategies: one is less burdensome in terms of both cost and time, although the reduction in uncertainty is also minor, whereas the other is expected to provide a much greater reduction in uncertainty reduction but at much higher cost. The decision maker may compute the differential EVSI to determine whether employing the j^{th} information collection strategy rather than the l^{th} information collection methodology is beneficial. The differential EVSI is calculated as:

$$\text{EVSI}^{jl} = \int \left[\text{TSC}_{k^l}^l(R) - \text{TSC}_{k^j}^j(R) \right] f^j(R) dR. \quad (4.21)$$

The differential EVSI is useful in comparing the benefits brought by timeliness of information collection and the reduction in uncertainty of two testing methodologies, as it does not take into account testing costs. Positive differential EVSI values imply that the greater reduction in uncertainty affords appreciable benefits, even after accounting for the fact that the more favourable test is of longer duration. On the other hand, non-

positive values suggest that the additional the testing time does not justify the greater reduction in uncertainty.

4.3.6. Expected Net Benefit of Sampling (ENBS)

The EVSI provides the potential benefit in collecting additional information; however, it does not consider the cost of testing in its derivation. Thus, a decision-maker may calculate the ENBS as:

$$\text{ENBS}^j = \text{EVSI}^j - \text{ICC}^j. \quad (4.22)$$

A positive ENBS implies that the j^{th} information collection methodology is favourable, whereas negative ENBS values mean that the additional information collected will not offset the data collection cost.

4.3.7. Return on Investment (ROI)

Another way to measure the expected benefit of collecting additional information is by calculating the ROI, defined as:

$$\text{ROI}^j = \frac{\text{ENBS}^j}{\text{ICC}^j}. \quad (4.23)$$

Similar to ENBS, a positive ROI value indicates that additional data is beneficial, while a negative ROI implies that the further testing is not worthwhile. It may be noted that, unlike ENBS, ROI provides the greatest value per dollar spent. The best testing procedure j^* based on ROI is determined as:

$$j^* = \underset{j}{\text{argmax}} \text{ROI}^j. \quad (4.24)$$

4.4. Illustrative Applications

In order to demonstrate how VOI is calculated, we present a series of illustrative applications under different assumptions designed to represent situations that could be

of general interest in practice. Each of the situations considered is characterized by a set of parameters given in Table 4.1.

The parameter N in Table 4.1 represents the size of the target population involved in the scenarios of interest. (This could be the entire population, or a population subgroup depending on the context and scope of the analysis.) TH denotes the time horizon over which VOI calculation is done. The time required to implement a regulation designed to reduce population health risk is denoted by t_{imp} , and the time for the regulation to achieve its intended effect in terms of risk reduction is denoted by t_{eff} . In the scenarios considered, we will compare the VOI associated with two testing strategies – denoted Test A and Test B – that will provide the information on the potential population health risk, each with its own sensitivity and specificity. The times required to conduct each of these testing strategies and evaluate the results are denoted by t_A and t_B , respectively.

A discount rate of r will be used to adjust future costs and benefits to net present value. The maximum annualized control cost associated with complete elimination of exposure is denoted CC_{max} . The annualized overhead control cost reflecting the differential cost of control as the regulatory intervention proceeds is denoted by CC_{OH} . Note that for simplicity, the overhead control cost is set to 0.

As the scenarios considered will focus on reduction in mortality, we will use VSL as the monetary benefit associated with risk reduction. Specifically, an overall VSL of \$9.1M annualized over a 70-year lifetime will be used in this analysis. The annualized testing budget B can be used to determine the number of tests A and B that can be done every year, enabling the calculation of the throughput of these tests.

The target risk level, denoted by TRL, serves as a threshold for an acceptable residual risk after regulation, as it is often impossible to completely eliminate exposure to environmental agents.

4.4.1. Parameters:

The control cost can then be computed by

$$CC_{OH} + (CC_{max} - CC_{OH}) \frac{10^{h\epsilon_k} - 1}{10^h - 1}. \quad (4.25)$$

Where $h > 0$ controls the steepness of the cost function. The default value of h is set to 3.

4.4.2. Decision Criteria

In addition to describing the costs and benefits of alternative testing strategies as discussed above, it is necessary to consider how risk decisions are made. Here, we consider two types of decision makers, one seeking to limit the population health risk to a specific target level, and the other seeking to achieve a balance between the costs of regulation and population health risks. Whereas the TRDM focuses exclusively on reducing risk to TRL regardless of the cost, the BRDM is motivated to minimize the TSC, thereby achieving a balance between costs and risks. Both decision making approaches may be appropriate, depending on the context in which the risk decision is to be made. These and other principles of the risk decision making are discussed in detail by Krewski et al. (2020).

In order to be confident that the population risk is below the TRL, the TRDM would like the 95th percentile of the uncertainty distribution (q_{95}) of the average population risk to be below the TRL: in this case, no regulatory action would be taken by the TRDM. Conversely, if the 5th percentile of this distribution (q_{05}) was above the TRL, the TRDM would take action to reduce risk. If the uncertainty interval between the 5th and 95th percentiles of the uncertainty distribution of the average population risk includes TRL, the TRDM would want additional data to reduce its uncertainty in order to decide whether risk reduction action is required.

Unlike the TRDM, the BRDM is interested in balancing the costs against benefits. This can be done by minimizing the TSC, which effectively balances the cost of obtaining additional

information to support decision making against the health benefits measured in economic terms. In this decision-making context, VOI may be evaluated in terms of ROI of the toxicological testing strategy used to collect the additional information. The BRDM would first calculate the optimal reduction in exposure (ORE) that would minimize the ETSC based on prior (k^0) and posterior (k^A and k^B , corresponding to Tests A and B, respectively) uncertainty distributions. To calculate the EVSI for Test A, the BRDM would compute the difference between the TSCs using k^0 and k^A , integrated over the posterior uncertainty distribution $f^A(R)$. Finally, the ROI is calculated as the ratio ENBS/ICC for Tests A and B. The BRDM will then choose the test with largest (positive) ROI. If the ROIs for both Tests A and B are negative, the BRDM would conclude that no additional information is required for decision-making purposes.

4.4.3. Scenario 1

4.4.3.1. Description

The toxicity and exposure parameters for a hypothetical chemical of interest is given in Table 4.2, along with the prior uncertainty distributions for these parameters. Note that although these parameters are subject to uncertainty, it is assumed that the uncertainty distributions are centered around the true parameter values, reflecting our assumption of unbiasedness with respect to prior uncertainty. In this initial scenario, we consider uncertainty only in μ_{tox} , the expected value of the uncertainty distribution for toxicity.

Attributes of Test A and Test B in this scenario are given in Table 4.3. Whereas the time to collect and analyze information via Test A is only 2 years, Test B requires 10 years for its information to be collected and analyzed. Reproducibility of Test A, denoted by $\sigma_{\text{test,A}}$, is assumed to be half of the prior uncertainty about μ_{tox} , whereas the reproducibility of Test B assumed to be one tenth of the prior uncertainty. Under these assumptions, Test B is assumed to be 80% less variable than Test A, and the uncertainty surrounding the posterior mean for Test A is calculated to be $\sqrt{101}/\sqrt{5} = 4.5$ times greater than that of Test B. The information collection cost, ICC, is 10 times higher for Test B (\$5M) than Test

A (\$0.5M). These characteristics will remain constant for all three scenarios discussed in this section.

The prior uncertainty distribution for population risk is shown in Figure 4.1, along with posterior distributions of risk following implementation of Tests A and B. Note that the fifth percentile, q_{05} , of the prior uncertainty distribution is below the TRL, whereas q_{95} is above the TRL. Thus, the TRDM cannot conclude whether a regulation should be imposed. Posterior uncertainty distribution from Test A is narrower; however, the TRL still falls in the interval (q_{05}, q_{95}) , implying that no decision would be made in most cases. On the other hand, as q_{95} based on Test B is expected to be below the TRL. Thus, we anticipate that Test B would lead to a correct decision in the majority of cases.

4.4.3.2. VOI Analysis under Target-Risk Decision Making:

Figure 4.1 shows that, as TRL is greater than the true risk, the correct decision is not to mitigate the chemical. As shown in Table 4.4, the likelihood that Test A can correctly conclude that the no regulation is required is low at 7.9%, while Test B leads to a correct conclusion in 57.6% of the time. Furthermore, with its testing variability, Test A is found to produce incorrect decision (i.e., regulation should be imposed) at 1% of the time. However, due to the difference in the testing costs, the less expensive Test A yields a greater number of correct decisions per year for chemicals with similar characteristics.

4.4.3.3. VOI Analysis under Benefit-Risk Decision Making:

In this scenario, a small reduction in exposure (12.5%) would minimize the true TSC to \$6.1M over the 20-year time horizon. As shown in Table 4.5, the prior ETSC is large (\$1,906.5M), due to high uncertainty about μ_{tox} . Both Tests A and B greatly reduce uncertainty, leading to large EVSI values (over \$300M in both cases). Even though the differential EVSI is \$3.4M, suggesting that the timeliness of Test A does not outweigh the benefits brought by the larger uncertainty reduction of Test B, the ROI for Test A (\$628 per \$1 spent in testing) is almost 10-fold higher than that of Test B (\$66 per \$1 spent in

testing), implying that the Test A would yield greater benefit due to the much lower cost of testing.

4.4.4. Scenario 2

4.4.4.1. Description

Table 4.6 presents the toxicity and exposure parameters used in Scenario 2. Compared to Scenario 1, the increase in the mean toxicity μ_{tox} implies that the true risk is increased, and the smaller prior uncertainty reflects a narrower prior uncertainty distribution.

4.4.4.2. VOI Analysis under Target-Risk Decision Making:

As shown in Figure 4.2, the true average population risk is larger than the TRL, suggesting that the regulatory action is required to ensure the safety of the population. Again, the prior uncertainty distribution precludes the TRDM from confidently concluding that risk mitigation is required. While the reduction in uncertainty based on Test B allows the TRDM to correctly conclude that risk mitigation is necessary, Test A would lead to the same conclusion in less than 40% of the time required for Test B (Table 4.7).

4.4.4.3. VOI Analysis under Benefit-Risk Decision Making

As shown in Table 4.8, the substance of interest poses an appreciable risk at the current level of exposure, so that risk mitigation is required. In this circumstance, the TSC would be minimized if the average population exposure is reduced by 40.5%. Since the prior uncertainty is narrower than that of Scenario 1, the EVPPI and EVSI are also smaller. Both Tests A and B successfully reduce the ORE, resulting in sufficiently large EVOI values that render these tests worthwhile. The EVSI of Test A is greater than that of Test B, implying the timeliness of Test A outweighs the greater uncertainty reduction provided by Test B (the negative differential EVSI value also supports this determination). The benefits of Test A are further supported by the difference in the testing costs: the ROI for Test A

(\$38.8 per \$1 spent on testing) is more than 17-fold greater than that of Test B (\$2.2 per \$1 spent on testing).

4.4.5. Scenario 3

4.4.5.1. Description

Table 4.9 describes the toxicity and exposure parameters Scenario 3. The chemical of interest poses least hazard among the three scenarios considered. As can be seen in Figure 4.3, the true risk is smaller than the TRL in this scenario.

4.4.5.2. VOI Analysis under Target-Risk Decision Making

Scenario 3 showcases the chemical with the least risk, where the TRL is slightly below q_{95} based on the prior uncertainty distribution. As a result, both Test A and Test B provide sufficient evidence to conclude that the chemical of interest is safe 94.7% and 100% of the time, respectively. The effectiveness of Test A is enhanced by the lower testing cost, allowing the TRDM to evaluate 9 times chemicals than Test B (Table 4.10).

4.4.5.3. VOI Analysis under Benefit-Risk Decision Making

As the true average population risk is relatively small, there is no need to regulate the chemical of interest, if the goal is to minimize the TSC. As shown in Table 4.11, both Tests A and B successfully reduce the ORE to 0%, implying that these tests, on average, would allow the BRDM to avoid implementing an unnecessary regulation. As the differential EVSI value of \$0 suggests, Test A achieves the same benefit as Test B. However, given the resource intensive nature of Test B, its ENBS and ROI values are negative. On the other hand, Test A provides an assurance that risk mitigation action is not required by reducing uncertainty in a cost-effective manner.

4.4.6. Cross-Scenario Findings

Three scenarios described above illustrate a variety of prototypical situations that a decision-maker may face in practice. Comparisons among these scenarios provide additional insight into the utility of value of information analysis in toxicological testing. As expected, the two decision-making styles (target-risk vs. benefit-risk) sometimes disagree as to whether risk mitigation action is required. In Scenario 1, the BRDM embraced low control costs to regulate the chemical of interest and reduce the TSC, while TRDM concluded that risk was sufficiently small such that no regulation was required once testing was completed. In scenario 2, the chemical of interest posed an appreciable risk, which compelled both the TRDM and BRDM to take action to reduce exposure. Since the TRDM does not consider the cost of regulation in their decision-making process, the TRDM is expected to employ more aggressive control measure than the BRDM, if a decision is taken to reduce exposure. Finally, if the TRL was increased slightly in Scenario 3, then the TRDM may find additional testing to be unnecessary, whereas the BRDM would appreciate the benefits both tests offer.

In all three scenarios considered, Test B attained EVSI values that are near the EVPPI; however, due to greater testing costs, the ROI for Test B was always less than (that of) Test A. A similar observation can be made for the TRDM: while correct decisions can be made more often by employing Test B, the expected testing cost per correct decision is less for Test A, favouring Test A as the more cost-effective testing strategy.

4.4.7. Testing Multiple Chemicals

In order to evaluate whether Test A produces greater benefit compared to Test B across the entire space of chemicals that require evaluation, we may consider the throughput of each testing strategy. Throughput is defined as the number of chemicals evaluated and appropriate risk mitigation action taken within a fixed budget for chemical testing. The expected value of the information system (EVIS) is then defined as the product of throughput and the average EVOI (across multiple chemicals) of Tests A and B.

To illustrate, the annual budget for toxicity testing is set to \$20M, and hence the regulatory agency could test 40 chemicals using Test A, or 4 chemicals using Test B. Then the ENBS from Test A on a hypothetical chemical in Scenario 1 would be \$16.5B, while Test B would provide the annual ENBS of \$1.6B, suggesting that Test A would achieve greater benefit.

More specifically, under Scenario 1, the TRDM would be able to make 18.2 correct decisions annually based on Test A, while Test B allows only 3.9 correct decisions per year. This implies that Test A can facilitate decision-making on multiple chemical substances of interest to a much greater extent than Test B.

4.4.8. Trade-offs between Reduction in Uncertainty and Timeliness of Testing Information
Both the timeliness of information collection and the amount of uncertainty reduced play critical roles in determining the effectiveness of the testing strategies. Table 4.12 provides the EVSIs for various testing times (ranging from 0 to 10 years), as well as various ratios of the testing variability and the prior uncertainty about μ_{tox} for each of the three scenarios considered above.

In Scenario 1, true optimal decision is to regulate the chemical, which is subject to relatively large prior uncertainty. This results in appreciable EVSI values at all levels of uncertainty reduction considered. Note that the reduction in uncertainty has a greater effect on the EVSI than timeliness of information collection.

In Scenario 2, the prior uncertainty is much less than that in Scenario 1. In this case, the control cost would be optimized with more accurate testing results. The timeliness of information is critical, as change in EVSI over time is significant. (At prolonged information collection times, the EVSI may become negative.)

In Scenario 3, the optimal action is not to regulate the chemical of interest. While the prior uncertainty suggests that the regulation may be required, even a small reduction in uncertainty would result in posterior decision not to regulate. When the decision is to not

regulate, the timeliness of information collection does not matter for decisions involving individual chemicals. However, when testing multiple chemicals, the least costly strategy would be preferred as it allows evaluation of a larger number of chemicals, some of which could require risk reduction action.

4.4.9. Effect of Uncertainty in Multiple Parameters

In all three scenarios considered here, we assumed that the uncertainty only lies in the mean toxicity μ_{tox} . In practice, it is reasonable to assume that there is also uncertainty about the other parameters used to characterize the toxicity of the chemical of interest. Since the toxicological testing can only reduce uncertainty in the toxicity of the chemical, the level of uncertainty associated with chemical exposure will remain unchanged. Assuming overall uncertainty remains unchanged, incorporating additional uncertainty in multiple parameters would result in reduced values of the EVSI, ENBS, and ROI for both Tests A and B.

4.4.10. Effect of Control Cost

The optimal reduction in exposure to minimize the TSC depends strongly on the control cost required to reduce the mean population exposure. In the three scenarios considered above, we assumed that a maximum annual control cost of \$50M, which translates to \$0.5 per capita. The ratio CC_{max}/V (per capita) is then $\$0.5/\$0.13\text{M} = 3.8 \times 10^{-6}$. Table 4.13 shows the effect of changing CC_{max} on the BRDM's decision under Scenario 2. The increased control cost would result in a lower level of exposure reduction in order to achieve the optimal TSC, with control cost increasing until risk mitigation is no longer justifiable. At a ratio of $CC_{\text{max}}/V = 10^{-3}$, the prior ORE and EVSI values for both tests are all 0, implying that no testing is necessary. Note that at all values of the ratio CC_{max}/V , the ROI for Test A is always equal to or greater than that of Test B.

4.5. Discussion and Conclusions

In this chapter, we present a methodology for evaluating the VOI provided by alternative toxicological tests that might be used to characterize the toxicity of chemical substances. In its most general form, the methodology described here allows for consideration of public health benefits derived from risk mitigation actions that may be taken on the basis of the test results. These benefits can be balanced against the cost of testing, so as to minimize the TSC associated with the general population exposure to the substance or substances of interest. More narrowly, the VOI analysis can be configured without consideration of TSC (which requires economic valuation of public health benefits derived from exposure reduction), focusing only on the achievement of target risk level within the general population level.

The analytic framework for VOI analysis presented here can help in choosing among alternative testing strategies, as well as among alternative risk mitigation strategies. This framework extends previous work on the EVOI by explicitly considering the timeliness of information collection and analysis.

When there exist great uncertainties in chemical's potency and prevalence rate, there is an appreciable likelihood that the average population risk of fatal adverse events is also high. Consequently, toxicological testing is likely to be rated as cost beneficial, as it will reduce the uncertainty about the population risk and provide a basis for cost-effective risk mitigation, while avoiding unnecessary regulation on chemicals that actually do not pose appreciable risk (or social cost) to the population.

The use of TSC, rather than ASC, allow decision-makers to obtain more realistic EVOI values. In particular, we found that the time taken both to acquire toxicological test data and to implement risk control measures based on this data is an important factor in achieving public health benefit: the earlier availability of the test data to support the earlier intervention, if necessary, will result in higher EVOI values, because they are realized earlier. Even when a substance does not pose a significant risk, cost-effective tests that sufficiently reduce prior uncertainty so as to conclude that risk mitigation is not

required would be beneficial as both the timeliness and reduced testing cost allow larger numbers of chemicals to be evaluated. (Because of the large number of environmental agents with incomplete toxicological profiles, less expensive and more rapid toxicological testing strategies will achieve greater overall public health benefit because of the broader coverage of the totality of the agents present in the human environment.)

In the context of present chapter, we considered two types of decision makers: the target-risk decision maker (TRDM) and the benefit-risk decision maker (BRDM). As expected, different decision-making criteria would lead to different costs and benefits. A TRDM who is motivated by achieving a target risk regardless of the control cost will behave differently than a BRDM who is seeking to balance the control cost against public health benefit: whereas the former decision maker may engage in more aggressive control strategies in order to ensure that all risks of concern are reduced to desired target level, the latter decision maker will seek to reduce risk in a more cost effective manner.

Since we assumed that the TRDM would like to equate q_{95} of the uncertainty distribution to the TRL to ensure the average population risk does not exceed the threshold, the presence of uncertainty will result in overestimation of the required level of exposure reduction. Similarly, the assumption of lognormality for each of the parameters implies that the uncertainty about risk would be right-skewed, leading the ORE for the BRDM to also be overestimated. In other words, the presence of uncertainty may lead to over-regulation.

As the primary purpose of this chapter is to present a comprehensive methodology for evaluating VOI of alternative toxicological tests, we have not applied this methodology to specific types of tests that could be of interest in regulatory toxicology. Future work using this methodology to compare the VOI associated with alternative toxicological testing strategies that might be considered in regulatory decision making would be useful. (Krewski et al. 2020, has recently provided a comprehensive review of current and emerging toxicological tests that could be compared using the method outlined in this article.)

Although we have not applied our VOI methodology to specific toxicology tests, we have presented three illustrative applications as to how our method can be used in practice to calculate the VOI associated with alternative toxicological tests. It should also be noted that the three scenarios considered here are based on simplified assumptions that provide insights into how VOI analysis can be informative in practice. One simple assumption made in the three scenarios considered here is that there exists uncertainty only with respect to the expected toxicity of the substance of interest. Other sources of uncertainty, such as uncertainty in model structure, can be accommodated within our framework and included in future applications. (In the three scenarios considered here, we assumed that the parameters follow lognormal uncertainty distribution and that this is the correct model.) The assumption of independence among the parameters greatly simplifies the calculation of the posterior uncertainty distribution but could be relaxed with additional computational effort.

Looking to the future, there are a number of natural extensions of the analytical framework for VOI analysis proposed here. For example, the target population may be partitioned into subgroups based on their vulnerability (e.g., highly sensitive subpopulations or highly exposed occupational populations.) Our methodology can be easily adopted to take into account multiple adverse effects of different severity associated with exposure to single chemical, which may occur at varying level of exposure. More comprehensive model is expected to increase associated EVOI values. Relaxing some of the simplifying assumptions made in the illustrative applications presented here – such as acknowledging uncertainty in model structure and uncertainty in multiple parameters – can be expected to reduce the estimated EVOI values.

5. Summary, Conclusion, and Future Work

Scientific-based analysis plays a key role in decisions to regulate environmental agents because of potential health risk (NRC 2009). Human health risk assessment (HHRA) is a well-established process comprised of four main steps: hazard identification, dose-response assessment, exposure assessment, and risk characterization (NRC 1983). In support of such decisions, decision-makers need to collect all available information and assess whether there is sufficient evidence to make a decision, or if additional information must be collected.

Since uncertainty exists in both toxicological and exposure data, there is always a possibility of making sub-optimal or even incorrect decisions. Value of information (VOI) analysis is a systematic approach that considers the risk of making an erroneous decision and its magnitude based on currently available information, and how further reduction in uncertainty may reduce such risk.

Given that only a fraction of the approximately 140,000 chemicals currently available in commerce have been evaluated in depth, there is a strong need for new approaches in toxicological and exposure assessment that provide reliable information more rapidly. Quantitative high-throughput screening (qHTS) introduced by Inglese et al. (2006) can produce *in vitro* assays for thousands of chemicals in a matter of days, which can be analyzed for hazard identification and dose-response assessment purposes.

Since qHTS is an *in vitro* toxicity testing technology, the results need to be extrapolated to humans when used for purposes of human health risk assessment. This is done using *in vitro* to *in vivo* extrapolation (IVIVE) with appropriate adjustment factors in order to establish a human reference dose (RfD). Such extrapolation may lead to greater uncertainty than traditional mammalian toxicity testing approaches. However, since the cost and time required to run qHTS assays are far less than those of traditional toxicity

testing approaches, the timeliness and greater throughput of qHTS should be included in determining the VOI of various toxicity testing strategies.

In addition to using VOI to optimize the value of toxicity test data, optimal experimental designs for qHTS or other toxicological assays may be employed. For qHTS data, the practitioner may choose the number of experimental concentrations, the concentration levels, and the number of replicates to be run at each concentration in order to results that satisfy a relevant optimality criterion. Here, we have focused on optimal experimental designs that minimize the variance of the BMD based on qHTS assays.

Prior to developing optimal designs that would minimize the asymptotic variance of the estimated BMD, we investigated the statistical properties of the maximum likelihood estimators (MLEs) of the BMDs and underlying Hill model parameters for qHTS data in chapter 2. A simulation study was conducted using eight representative CRCs chosen based on 8,456 data from qHTS assays conducted as part of Tox21 program Phase I.

Four performance evaluation statistics – relative bias (RB), coefficient of variation (CV), relative root mean squared error (RRMSE), and confidence coverage – were considered. The quality of the estimation of Hill model parameters α and θ depend largely on whether a sufficient number of concentrations are placed at the respective asymptotes. In particular, having two concentrations placed at the asymptotes lead to smaller RB, CV, and RRMSE. As κ represents the inflection point of the CRC, placing concentrations at both side of it, while not on the asymptote, lead to smaller RRMSEs. Our ability to estimate η well depended not only on the spacing of concentrations, but also on the location of κ with respect to mCT and MCT. In particular, when η is small compared to dose spacings, or η is large and κ is near the boundary of the experimental concentration range, the RB, CV, RRMSE were poor.

Overall, using a typical design with 14 concentrations, the estimation of the Hill model parameters and corresponding CRC would be superior when concentrations are placed at both lower and upper asymptotes, as well as on both sides of the inflection point but not on the asymptote.

Goodness-of-fit of the CRC using mean absolute deviation (MAD) revealed that CRCs are generally well estimated within the experimental range. However, if either of the asymptotes are not well captured, then extrapolation may be difficult.

Since the BMR_{AR} is typically set to be a small deviation from the background response, the estimator \widehat{BMD}_{AR} performed well as long the experiment captured information about the lower asymptote, as well as near the BMD well. On the other hand, the value of the BMR_{ER} is dependent on the range $(\theta - \alpha)$, performance of \widehat{BMD}_{ER} was poorer compared to \widehat{BMD}_{AR} when the experiment failed to estimate θ well.

The results from this chapter clearly indicate that the statistical properties of the Hill model parameters and associated BMDs are dependent on the shape of the underlying CRC and to extent to which the experimental concentrations capture key aspects of the CRC linked to the corresponding Hill model parameters.

Motivated by the findings in chapter 2, it is desirable to derive optimal designs for qHTS analysis, as both the strategic placement of experimental concentrations, and the increase in number of replicates would lead to better estimates. In chapter 3, we showed that, using Chernoff's theorem (Chernoff 1952), the optimal design (OD) that minimizes the asymptotic variance of the BMD estimate requires four experimental concentrations. When a BMR_{ER} of 5% is used, optimal concentrations included a control group at $x_1^* = 0$, and an optimal maximum concentration x_4^* corresponding to $ER_{x_4^*}$ approaching 1, with x_2^* and x_3^* such that $ER_{x_2^*} = 21.8\%$ and $ER_{x_3^*} = 78.2\%$. Under the assumption of constant error variance, the optimal concentration vector \mathbf{x}^* is invariant with respect to the underlying Hill model parameters.

When practical constraints prohibit setting the MCT as large as x_4^* , a constrained optimal design (COD), subject to $x_4 \leq MCT$, can be developed. We found that the relative efficiency (RE) of the COD with respect to theoretical OD decreased as ER_{MCT} was reduced.

As both the theoretical OD and COD require knowledge about the underlying Hill model parameters, fixed designs not requiring the knowledge of these parameters are used in practice. A comparison of fixed designs with between 4 and 15 concentrations across over 8,000 qHTS datasets showed a median RE was at most 22%. While FD_1 with 15 experimental concentrations typically used in practice, performed best based on median RE, the $ACOD_{med}$ with only 4 concentrations produced much better results in a number of instances. FD_4 , with five experimental concentrations, showed worst performance overall. Given a large variability in the shape of CRCs observed in the experimental range, no fixed designs can consistently produce high efficiency.

In chapter 4, we present an analytical framework for VOI analysis that takes into account the timeliness of information collection over a given time horizon when calculating the total social cost (TSC). Our analysis revealed that the time taken both to acquire toxicological test data and to implement risk control measures based on this data is an important factor in achieving public health benefit: the earlier availability of the test data to support the earlier intervention, if necessary, will result in higher EVOI values because they are realized earlier. Even when a substance does not pose a significant risk, cost-effective tests that sufficiently reduce prior uncertainty so as to conclude that risk mitigation is not required would be beneficial as both the timeliness and reduced testing cost allow larger numbers of chemicals to be evaluated.

In the context of chapter 4, we considered two types of decision makers: the target-risk decision maker (TRDM), who seek to achieve desired low level of risk, and the benefit-risk decision maker (BRDM). A TRDM who is motivated by achieving a target risk regardless of the control cost will behave differently than a BRDM who is seeking to balance the control cost against public health benefit: whereas the former decision maker may engage in more aggressive control strategies in order to ensure that all risks of concern are reduced to desired target level, the latter decision maker will seek to reduce risk in a more cost effective manner.

The overarching objective of this thesis was to develop toxicological risk assessment methods that would generate better toxicity testing data to support the development of exposure guidelines for environmental agents to which people may be exposed. This objective was accomplished through derivation of optimal experimental designs for qHTS assays and the development of VOI methods for comparing alternative toxicity testing strategies.

The work carried out in chapters 2 and 3 provides a statistical foundation to improve the experimental design of future qHTS assays. Derivation of optimal designs that minimize the asymptotic variance of the estimated BMD allows the comparison of the relative efficiencies of various fixed designs. The eight prototypical CRCs used in chapter 2, as well as the comparison of relative efficiencies of various fixed designs (including ACODs) are based on over 8,000 qHTS datasets collected from past experiments. By using real-world qHTS data in this analysis, the results presented in these chapters will be relevant to practitioners.

The framework for VOI analysis was extended by including both the timeliness and cost of information collection, which allows for a more complete comparison of the VOI associated with different toxicity testing strategies. Incorporation of timeliness is particularly important, as the faster the additional information is collected and assessed (assuming that it provides sufficient evidence to allow a decision to be made), public health benefit may be increased because risk management action may be taken earlier, if required. Even if regulation is unnecessary, the enhanced timeliness of information collection benefits both decision-makers and the public by providing assurances of safety. Since only a fraction of substances present in human environment have been evaluated, qHTS may be an effective testing strategy because the cost and time of data collection is substantially reduced, facilitating greater coverage of the universe of untested agents.

Although this thesis has produced considerable insights in the use of optimal designs for qHTS assays and VOI methods for evaluating alternative toxicity testing strategies, the results are subject to some limitations.

In chapters 2 and 3, we have assumed that the underlying dose-response curves for the qHTS assays follow a Hill model. Although the Hill model is widely used in practice, there are other models (e.g., the exponential model) that have been used in practice. Therefore, it would be beneficial to evaluate the statistical properties of other models, as was done in chapter 2 for the Hill model.

None of the fixed designs performed well across the entire range of CRCs that can be observed in practice. Further, derivation of the OD requires knowledge about the underlying Hill model parameters. When such information is available, the variance of the estimated BMD can be greatly reduced by strategically constructing the experiment. When no prior knowledge of the Hill model parameters is available, designs with a greater number of experimental concentration levels typically have an advantage. In chapter 3, the concept of two-phase empirical Bayes approach was introduced as a potential design strategy to improve the quality of the BMD estimates. It would therefore be beneficial to develop the theoretical foundation for such a design.

The VOI framework was presented in chapter 4. A comparison of the VOIs resulting from various toxicological testing methodologies was also performed in the context of three illustrative examples. Since the primary objective of these examples was to demonstrate how the inclusion of the timeliness of information and regulatory action can play a role in the determination of VOI, they were based on simplified assumptions. In future, the assumptions could be altered to follow more closely what a practical application might suggest. It may also be worthwhile to accommodate additional sources of uncertainty, such as in the choice of model structure. Finally, the framework could be extended to allow for partitioning of the target population into subgroups, such as highly vulnerable sub-populations and workers with high occupational exposures.

Bibliography

1. Andersen M.E., and Krewski D. (2009). Toxicity Testing in the 21st Century: Bringing the Vision to Life. *Toxicol. Sci.* 107, 324-30.
2. Andronis, L. (2015). Analytic Approaches for Research Priority-Setting: Issues, Challenges and the Way Forward. *Expert Review of Pharmacoeconomics and Outcomes Research.* 15(5), 745-54.
3. Chernoff, H. (1952). Locally Optimal Designs for Estimating Parameters. *Topics in Applied Statistics.* 1150, 1-72.
4. Chiu, W.A., Axelrad, D.A., Dalaijamts, C., Dockins, C., Shao, K., Shapiro, A.J., and Paoli, G. (2018). Beyond the RfD: Broad Application of a Probabilistic Approach to Improve Chemical Dose-Response Assessments for Noncancer Effects. *Environmental Health Perspectives.* 126(6), 067009.
5. Crump, K. (1984). A New Method for Determining Allowable Daily Intakes. *Fundamental and Applied Toxicology.* 4(5), 854–871.
6. Crump, K. (1994). Calculation of Benchmark Doses from Continuous Data. *Risk Analysis,* 15(1), 79–89.
7. Davis A., Gift S., and Zhao J. (2011). Introduction to Benchmark Dose Methods and U.S. EPA's Benchmark Dose Software (BMDS) Version 2.1.1. *Regul. Toxicol. Pharmacol.* 254, 181–191.
8. EPA. (2012). Benchmark Dose Technical Guidance. United States Environmental Protection Agency. https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf
9. Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Doucher, M., and Maire, P. (2008). The Hill Equation: A Review of Its Capabilities in Pharmacological Modelling. *Fundamental and Clinical Pharmacology.* 22, 633-648.
10. Hill, A.V., (1910). The Possible Effects of the Aggregation of the Molecules of Haemoglobin on its Dissociation Curves. *Journal of Physiology.* 40, 4-7.
11. Huang, R., Xia, N., Cho, M-H., Sakamuru, S., Shinn, P., Houck, K.A., Dix, D.J., Judson, R.S., Witt, K.L., Kavlock, R.j., Tice, R.R., and Austin, C.P. (2011). *Chemical Genomics*

- Profiling of Environmental Chemical Modulation of Human Nuclear Receptors. *Environmental Health Perspectives*. 119(8), 1142-1148.
12. ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods). (2018). A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States. Available: <https://ntp.niehs.nih.gov/go/iccvam-rdmp>.
 13. Inglese J., Auld D.S., Jadhav A., Johnson, R.L, Simeonov, A., Yasgar, A., Zheng, W., and Austin, C.P. (2006). Quantitative High-Throughput Screening: A Titration-Based Approach that Efficiently Identifies Biological Activities in Large Chemical Libraries. *Proc. Natl. Acad. Sci. U.S.A.* 103(31), 11473-11478.
 14. Keisler, J.M., Collier, Z.A., Chu, E., Sinatra, N., and Linkov, I. (2014). Value of Information Analysis: The State of Application. *Environmental Systems and Decisions*. 34, 3-23.
 15. Khinkis L.A., Levasseur L., Faessel H., and Greco W.R. (2003). Optimal design for estimating parameters of the 4-parameter hill model. *Nonlinearity Biol Toxicol Med*. 1(3), 363-377.
 16. Koffijberg, H., Rothery, C., Chalkidou, K., and Grutters, J. (2018). Value of Information Choices that Influence Estimates: A Systematic Review of Prevailing Considerations. *Medical Decision Making*. 38(7), 888-900.
 17. Koffijberg, H., Knies, S., and Janssen, M.P. (2018). The Impact of Decision Makers' Constraints on the Outcome of Value of Information Analysis. *Value in Health*. 21(2), 203-209.
 18. Krewski D., Andersen M.E., Mantus E., and Zeise L. (2009). Toxicity testing in the 21st century: Implications for Human Health Risk Assessment. *Risk. Anal.* 29, 474-79.
 19. Krewski, D., Westphal, M., Andersen, M.E., Paoli, G.M., Chiu, W.A., Al-Zoughool, M., Croteau, C., Burgoon, L.D., and Cote, I. (2014). A framework for the next generation of risk science. *Environmental Health Perspectives*. 122(8), 796-805.
 20. Krewski, D., Andersen, M.E., Tyshenko, M.G., Krishnan, K., Hartung, T., Boekelheide, K., Wambaugh, J.F., Jones, D., Whelan, M., Thomas, R., Yauk, C., Barton-Maclaren, T.,

- and Cote, I. (2020). Toxicity Testing in the 21st Century: Progress in the Past Decade and Future Perspectives. *Archives of Toxicology*. 94, 1-58.
21. Krewski, D., and Ryzin, J.V. (1981). Dose Response Models for Quantal Response Toxicity Data. *Statistics and Related Topics*. 201-231.
 22. Krewski, D., Saunders-Hastings, P., Westphall, M., Tyshenko, M.G., Larkin, P., Leiss, W., Dusseault, M., Jerrett, M., and Coyle, D. (2020). *Journal of Toxicology and Environmental Health, Part B: Critical Reviews*. (submitted)
 23. Krewski D., Westphal M., Al-Zoughool M., Croteau M.C., and Andersen M.E. (2011). New Directions in Toxicity Testing. *Annu. Rev. Public Health* 32, 161-178.
 24. Lave, L.B., and Omenn, G.S. (1986). Cost Effectiveness of Short-term Tests for Carcinogenicity. *Nature*. 324, 29-34.
 25. Lave, L.B., Ennever, F.K., Rosenkranz, H.S., and Omenn, G.S. (1988). Information Value of the Rodent Bioassay. *Nature*. 336, 631-633.
 26. Malo, N., Hanley, J., Cerquozzi, S., Pelletier, J., and Nadon, R. (2006). Statistical practice in high-throughput screening data analysis. *Nature Biotechnology*, 24(2), 167-175.
 27. National Research Council. (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: The National Academies Press.
 28. Paustenbach, D. (2003). Human and Ecological Risk Assessment: Theory and Practice. *Human and Ecological Risk Assessment: An International Journal*, 9(4), 1089–1090.
 29. Sand, E., Portier, C.J., and Krewski, D. (2011). A Signal-to-Noise Crossover Dose as the Point of Departure for Health Risk Assessment. *Environmental Health Perspectives*. 119, 1766-1774.
 30. Sand S., Parham F., Portier C.J., Tice R.R., and Krewski D. (2017). Comparison of Points of Departure for Health Risk Assessment Based on High-Throughput Screening Data. *Environ Health Perspect*. 125(4), 623-633.
 31. Shockley, K.R. (2015). Quantitative High-Throughput Screening Data Analysis: Challenges and Recent Advances. *Drug Discovery Today*. 20, 296-300.

32. Slob, W. (2001). Dose-Response Modeling of Continuous Endpoints. *Toxicological Sciences*. 66, 298-312.
33. Taylor, A., Evans, J., and McKone, T. (1993). The Value of Animal Test Information in Environmental Control Decisions, *Risk Analysis*. 13(4), 403–412.
34. Thompson, K.M., and Evans, J.S. (1997). The Value of Improved National Exposure Information for Perchloroethylene (Perc): A Case Study for Dry Cleaners. *Risk Analysis*. 17(2), 253-271.
35. Tice R.R., Austin C.P., Kavlock R.J., and Bucher J.R. (2013). Improving the Human Hazard Characterization of Chemicals: A Tox21 Update. *Environ. Health Perspect.* 121, 756-765.
36. Tuffaha, H.W., Strong, M., Gordon, L.G., and Scuffham, P.A. (2016). Efficient Value of Information Calculation Using a Nonparametric Regression Approach: An Applied Perspective. *Value in Health*. 19(4), 505-9.
37. Xia M., Shahane S., Huang R., Titus S.A., Shum E., Zhao Y., Southall, N., Zheng, W., Witt, K.L., Tice, R.R., and Austin, C.P. (2011). Identification of Quaternary Ammonium Compounds as Potent Inhibitors of hERG Potassium Channels. *Toxicology and Applied Pharmacology*. 252(3), 250-258.
38. Yokota, F., and Thompson, K.M. (2004). Value of Information Analysis in Environmental Health Risk Management Decisions: Past, Present, and future. *Risk Analysis*. 24(3), 635-650.

Appendix A: Figures and Tables

A.1. Tables and Figures for Chapter 2

Table 2.1: Hill parameters for 8 scenarios

Simulation	α	κ'	η	θ
1	0	-3	5	1
2	0	-2	1	1
3	0	-0.5	1	1
4	0	-0.5	0.6	1
5	0	0.8	3	1
6	0	1.7	5	1
7	0	1.7	0.3	1
8	0	2	3	1

Table 2.2: Concentration levels used for 8 simulations in log-scale

i	z_i	i	z_i
1	-3.36	8	-0.19
2	-2.41	9	-0.02
3	-1.93	10	0.69
4	-1.46	11	0.93
5	-1.24	12	1.40
6	-0.98	13	1.56
7	-0.50	14	1.88

Table 2.3: Performance Evaluation Statistics for Hill Model Parameters. In the case $\alpha = 0$, the bias, standard error, and root MSE are reported rather than RB, CV, and RRMSE.

Scenario	Parameter Value	RB	CV	RRMSE	Coverage
α^*					
1	0	-3.0%	11.6%	12.0%	99.1%
2	0	-0.4%	7.5%	7.5%	91.6%
3	0	- 0.1 %	3.3%	3.3%	90.6%
4	0	-0.3%	5.5%	5.5%	91.0%
5	0	- 0.1 %	1.6%	1.6%	91.6%
6	0	-0.3%	1.4%	1.4%	93.0%
7	0	-0.6%	8.9%	8.9%	82.5%
8	0	-0.4%	1.5%	1.5%	91.1%
θ					
1	1	0.2%	1.3%	1.4%	93.8%
2	1	0.2%	2.0%	2.0%	88.4%
3	1	0.3%	2.9%	2.9%	90.0%
4	1	0.9%	6.1%	6.2%	92.2%
5	1	0.6%	3.3%	3.4%	90.8%
6	1	45.7%	101.8%	111.5%	98.3%
7	1	57.7%	158.3%	168.5%	71.0%
8	1	-16.5%	67.0%	69.0%	71.5%
κ					
1	0.001	54.6%	62.1%	82.7%	100.0%
2	0.01	2.2%	23.0%	23.1%	91.8%
3	0.32	2.2%	16.4%	16.5%	91.0%
4	0.32	9.0%	46.7%	47.6%	90.8%
5	6.31	0.7%	6.4%	6.4%	92.0%
6	50.12	23.6%	54.1%	59.0%	97.4%
7	50.12	>10 ⁶ %	>10 ⁷ %	>10 ⁷ %	71.9%
8	100.00	-16.4%	67.5%	69.5%	82.9%
η					
1	5	2.4%	19.4%	19.5%	99.1%
2	1	2.3%	18.0%	18.2%	89.2%
3	1	2.3%	15.2%	15.4%	89.6%
4	0.6	2.6%	18.9%	19.0%	91.2%
5	3	0.7%	21.9%	21.9%	88.2%
6	5	-4.4%	36.7%	37.0%	99.6%
7	0.3	38.2%	133.4%	138.8%	97.5%
8	3	24.8%	57.3%	62.4%	100.0%

Table 2.4: Performance Evaluation Statistics for BMD

Scenario	BMR _{AR} = 0.05					BMR _{ER} = 0.05				
	BMD	RB	CV	RRMSE	Coverage	BMD	RB	CV	RRMSE	Coverage
1	5.55x10 ⁻⁴	57.9%	75.9%	95.4%	100.0%	5.55x10 ⁻⁴	57.9%	74.9%	94.7%	100.0%
2	5.26x10 ⁻⁴	22.8%	83.4%	86.5%	100.0%	5.26x10 ⁻⁴	19.4%	74.8%	77.3%	100.0%
3	1.66x10 ⁻²	14.0%	57.6%	59.3%	90.2%	1.66x10 ⁻²	12.9%	53.2%	54.7%	90.8%
4	2.34x10 ⁻³	54.4%	154.7%	164.0%	85.6%	2.34x10 ⁻³	43.9%	127.7%	135.0%	88.0%
5	2.36	-1.3%	22.3%	22.3%	86.2%	2.36	-1.1%	21.9%	21.9%	86.0%
6	2.78x10	-6.4%	13.8%	15.2%	100.0%	2.78x10	3.3%	16.1%	16.5%	100.0%
7	2.74x10 ⁻³	>10 ³ %	>10 ⁴ %	>10 ⁴ %	86.2%	2.74x10 ⁻³	>10 ⁴ %	>10 ⁵ %	>10 ⁵ %	94.3%
8	3.75x10	-14.8%	19.3%	24.3%	97.6%	3.75x10	-19.3%	30.3%	35.9%	97.6%

Table 2.5: Mean absolute deviation (MAD) for 8 simulations

Scenario	MAD _{between}	MAD _{below}	MAD _{above}
		MAD _{between}	MAD _{between}
1	5.10x10 ⁻²	1.42	0.21
2	2.02x10 ⁻²	2.75	0.76
3	2.12x10 ⁻²	1.26	1.09
4	2.08x10 ⁻²	1.95	1.94
5	1.84x10 ⁻²	0.70	1.44
6	1.63x10 ⁻²	0.72	30.21
7	1.95x10 ⁻²	2.38	14.55
8	1.55x10 ⁻²	0.81	32.13

Table 2.6: Performance Evaluation Statistics for BMDs with various BMRs (BMR=10%, 5%, and 1%)

Scenario	BMR _{AR} = 0.1				
	BMD	RB	CV	RRMSE	Coverage
1	6.44x10 ⁻⁴	56.9%	72.3%	92.0%	100.0%
2	1.11x10 ⁻³	15.3%	65.3%	67.1%	100.0%
3	3.51x10 ⁻²	9.0%	44.1%	45.0%	90.6%
4	8.12x10 ⁻³	32.9%	104.8%	109.9%	87.4%
5	3.03	-1.4%	17.5%	17.6%	86.2%
6	3.23x10	-4.4%	9.6%	10.5%	100.0%
7	3.31x10 ⁻²	459.6%	>10 ³ %	>10 ³ %	89.1%
8	4.81x10	-15.4%	16.2%	22.4%	94.3%

Scenario	BMR _{AR} = 0.05				
	BMD	RB	CV	RRMSE	Coverage
1	5.55x10 ⁻⁴	57.9%	75.9%	95.4%	100.0%
2	5.26x10 ⁻⁴	22.8%	83.4%	86.5%	100.0%
3	1.66x10 ⁻²	14.0%	57.6%	59.3%	90.2%
4	2.34x10 ⁻³	54.4%	154.7%	164.0%	85.6%
5	2.36	-1.3%	22.3%	22.3%	86.2%
6	2.78x10	-6.4%	13.8%	15.2%	100.0%
7	2.74x10 ⁻³	>10 ³ %	>10 ⁴ %	>10 ⁴ %	86.2%
8	3.75x10	-14.8%	19.3%	24.3%	97.6%

Scenario	BMR _{AR} = 0.01				
	BMD	RB	CV	RRMSE	Coverage
1	3.99x10 ⁻⁴	60.6%	85.3%	104.6%	100.0%
2	1.01x10 ⁻⁴	47.8%	142.5%	150.3%	100.0%
3	3.19x10 ⁻³	30.3%	98.1%	102.7%	88.2%
4	1.49x10 ⁻⁴	140.0%	377.3%	402.4%	82.6%
5	1.36	-0.1%	32.9%	32.9%	86.2%
6	2.00x10	-10.0%	24.1%	26.1%	100.0%
7	1.12x10 ⁻⁵	>10 ⁵ %	>10 ⁶ %	>10 ⁶ %	77.8%
8	2.16x10	-9.7%	29.9%	31.4%	100.0%

Scenario	BMR _{ER} = 0.1				
	BMD	RB	CV	RRMSE	Coverage
1	6.44x10 ⁻⁴	56.9%	71.2%	91.1%	100.0%
2	1.11x10 ⁻³	12.5%	57.2%	58.6%	100.0%
3	3.51x10 ⁻²	8.3%	39.9%	40.8%	91.6%
4	8.12x10 ⁻³	25.6%	84.4%	88.2%	90.4%
5	3.03	-1.1%	17.0%	17.1%	85.6%
6	3.23x10	7.2%	21.4%	22.6%	100.0%
7	3.31x10 ⁻²	>10 ⁴ %	>10 ⁵ %	>10 ⁵ %	93.7%
8	4.81x10	-19.9%	32.6%	38.2%	92.7%

Scenario	BMR _{ER} = 0.05				
	BMD	RB	CV	RRMSE	Coverage
1	5.55x10 ⁻⁴	57.9%	74.9%	94.7%	100.0%
2	5.26x10 ⁻⁴	19.4%	74.8%	77.3%	100.0%
3	1.66x10 ⁻²	12.9%	53.2%	54.7%	90.8%
4	2.34x10 ⁻³	43.9%	127.7%	135.0%	88.0%
5	2.36	-1.1%	21.9%	21.9%	86.0%
6	2.78x10	3.3%	16.1%	16.5%	100.0%
7	2.74x10 ⁻³	>10 ⁴ %	>10 ⁵ %	>10 ⁵ %	94.3%
8	3.75x10	-19.3%	30.3%	35.9%	97.6%

Scenario	BMR _{ER} = 0.01				
	BMD	RB	CV	RRMSE	Coverage
1	3.99x10 ⁻⁴	60.5%	84.5%	103.9%	100.0%
2	1.01x10 ⁻⁴	42.6%	130.7%	137.4%	100.0%
3	3.19x10 ⁻³	28.3%	92.4%	96.6%	88.8%
4	1.49x10 ⁻⁴	117.7%	316.8%	337.9%	83.8%
5	1.36	0.0%	32.5%	32.5%	86.0%
6	2.00x10	-3.0%	18.4%	18.6%	100.0%
7	1.12x10 ⁻⁵	>10 ⁵ %	>10 ⁶ %	>10 ⁶ %	90.2%
8	2.16x10	-15.3%	34.4%	37.7%	100.0%

Table 2.7: Performance Evaluation Statistics for Hill Model Parameters with $\sigma_0 = 0.025$, $\lambda = 0$, and no replicates. [In the case $\alpha = 0$, the bias, standard error, and root MSE are reported rather than RB, CV, and RRMSE.]

Scenario	Parameter Value	RB	CV	RRMSE	Coverage
α^*					
1	0	-0.6%	5.4%	5.4%	100.0%
2	0	<0.1%	3.5%	3.5%	89.8%
3	0	<0.1%	1.6%	1.6%	90.8%
4	0	<0.1%	2.6%	2.6%	90.4%
5	0	<0.1%	0.8%	0.8%	91.2%
6	0	-0.1%	0.7%	0.7%	91.3%
7	0	0.4%	4.0%	4.0%	85.9%
8	0	< 0.1 %	0.7%	0.7%	92.1%
θ					
1	1	0.1%	0.7%	0.7%	95.7%
2	1	0.1%	1.0%	1.0%	87.8%
3	1	0.1%	1.4%	1.4%	89.2%
4	1	0.2%	2.6%	2.6%	90.4%
5	1	0.2%	1.6%	1.6%	91.0%
6	1	26.7%	67.9%	73.0%	91.0%
7	1	7.8%	51.4%	52.0%	78.4%
8	1	-4.9%	60.7%	60.9%	74.9%
κ					
1	0.00	47.2%	44.4%	64.8%	100.0%
2	0.01	0.9%	11.1%	11.1%	91.4%
3	0.32	0.7%	7.9%	8.0%	91.0%
4	0.32	2.1%	15.0%	15.1%	89.6%
5	6.31	0.2%	3.0%	3.1%	91.8%
6	50.12	14.1%	35.9%	38.6%	98.1%
7	50.12	>10 ⁴ %	>10 ⁵ %	>10 ⁵ %	80.0%
8	100	-13.7%	34.8%	37.4%	85.7%
η					
1	5	2.1%	13.3%	13.5%	100.0%
2	1	0.6%	8.6%	8.6%	90.0%
3	1	0.7%	7.2%	7.3%	89.2%
4	0.6	0.8%	9.0%	9.1%	91.0%
5	3	< 0.1 %	10.5%	10.5%	89.8%
6	5	1.1%	35.8%	35.8%	97.0%
7	0.3	13.1%	32.2%	34.8%	95.3%
8	3	35.7%	55.4%	65.9%	100.0%

Table 2.8: Performance Evaluation Statistics for Hill Model Parameters with $\sigma_0 = 0.075$, $\lambda = 0$, and no replicates. [In the case $\alpha = 0$, the bias, standard error, and root MSE are reported rather than RB, CV, and RRMSE.]

Scenario	Parameter Value	RB	CV	RRMSE	Coverage
α^*					
1	0	-5.6%	22.3%	22.9%	97.2%
2	0	-1.5%	13.3%	13.4%	92.4%
3	0	-0.2%	5.0%	5.0%	90.2%
4	0	-1.1%	9.5%	9.5%	90.8%
5	0	-0.1%	2.5%	2.5%	91.2%
6	0	-0.5%	2.2%	2.2%	91.5%
7	0	-6.2%	52.9%	53.3%	80.4%
8	0	-1.2%	4.0%	4.2%	89.8%
θ					
1	1	0.4%	2.0%	2.1%	91.7%
2	1	0.4%	3.0%	3.0%	89.4%
3	1	0.6%	4.6%	4.6%	90.6%
4	1	3.1%	20.1%	20.3%	91.8%
5	1	1.4%	5.6%	5.8%	91.6%
6	1	48.5%	94.8%	106.4%	96.4%
7	1	100.9%	251.4%	270.9%	64.2%
8	1	-28.7%	68.8%	74.6%	65.7%
κ					
1	0.00	81.9%	82.7%	116.4%	100.0%
2	0.01	4.0%	36.4%	36.6%	91.8%
3	0.32	4.8%	27.1%	27.6%	89.8%
4	0.32	>10 ³ %	>10 ⁵ %	>10 ⁵ %	92.6%
5	6.31	1.8%	11.4%	11.5%	92.4%
6	50.12	24.7%	57.3%	62.4%	98.2%
7	50.12	>10 ⁷ %	>10 ⁷ %	>10 ⁷ %	64.6%
8	100	>10 ⁶ %	>10 ⁷ %	>10 ⁷ %	78.7%
η					
1	5	8.7%	37.9%	38.9%	98.2%
2	1	6.1%	31.3%	31.9%	92.0%
3	1	5.4%	24.5%	25.1%	92.2%
4	0.6	5.9%	31.0%	31.6%	92.4%
5	3	3.1%	34.6%	34.7%	87.4%
6	5	-5.0%	37.9%	38.2%	99.4%
7	0.3	89.2%	224.6%	241.7%	100.0%
8	3	17.4%	65.2%	67.5%	98.1%

Table 2.9: Performance Evaluation Statistics for BMDs with various σ ($\sigma = 0.025, 0.05$ and 0.075)

Scenario	BMR _{AR} = 0.05, $\sigma = 0.025$					BMR _{AR} = 0.05, $\sigma = 0.05$					BMR _{AR} = 0.05, $\sigma = 0.075$				
	BMD	RB	CV	RRMSE	Coverage	BMD	RB	CV	RRMSE	Coverage	BMD	RB	CV	RRMSE	Coverage
1	5.55x10 ⁻⁴	49.2%	50.8%	70.8%	100.0%	5.55x10 ⁻⁴	57.9%	75.9%	95.4%	100.0%	5.55x10 ⁻⁴	92.9%	115.1%	147.9%	98.8%
2	5.26x10 ⁻⁴	6.0%	37.0%	37.5%	90.0%	5.26x10 ⁻⁴	22.8%	83.4%	86.5%	88.2%	5.26x10 ⁻⁴	55.1%	159.1%	168.4%	85.4%
3	1.66x10 ⁻²	3.9%	25.9%	26.2%	90.2%	1.66x10 ⁻²	14.0%	57.6%	59.3%	88.6%	1.66x10 ⁻²	31.8%	100.2%	105.1%	87.6%
4	2.34x10 ⁻³	13.8%	57.5%	59.1%	89.2%	2.34x10 ⁻³	54.4%	154.7%	164.0%	85.4%	2.34x10 ⁻³	135.6%	370.2%	394.2%	83.2%
5	2.36	-0.5%	10.7%	10.7%	90.2%	2.36	-1.3%	22.3%	22.3%	87.2%	2.36	-1.5%	33.1%	33.2%	85.2%
6	2.78x10	-2.4%	9.5%	9.8%	99.5%	2.78x10	-6.4%	13.8%	15.2%	99.6%	2.78x10	-7.8%	18.6%	20.2%	99.4%
7	2.74x10 ⁻³	346%	735%	813%	91.1%	2.74x10 ⁻³	>10 ³ %	>10 ⁴ %	>10 ⁴ %	94.2%	2.74x10 ⁻³	>10 ³ %	>10 ⁴ %	>10 ⁴ %	90.6%
8	3.75x10	-1.2%	15.0%	15.0%	94.6%	3.75x10	-14.8%	19.3%	24.3%	96.7%	3.75x10	-28.7%	24.3%	37.6%	91.5%

Scenario	BMR _{ER} = 0.05, $\sigma = 0.025$					BMR _{ER} = 0.05, $\sigma = 0.05$					BMR _{ER} = 0.05, $\sigma = 0.075$				
	BMD	RB	CV	RRMSE	Coverage	BMD	RB	CV	RRMSE	Coverage	BMD	RB	CV	RRMSE	Coverage
1	5.55x10 ⁻⁴	49.04%	50.23%	70.20%	100.00%	5.55x10 ⁻⁴	57.88%	74.92%	94.67%	100.00%	5.55x10 ⁻⁴	92.6%	113.8%	146.8%	100.0%
2	5.26x10 ⁻⁴	5.11%	33.48%	33.87%	89.60%	5.26x10 ⁻⁴	19.44%	74.83%	77.32%	88.60%	5.26x10 ⁻⁴	47.4%	141.8%	149.6%	85.8%
3	1.66x10 ⁻²	3.60%	23.95%	24.22%	90.40%	1.66x10 ⁻²	12.87%	53.19%	54.72%	89.60%	1.66x10 ⁻²	29.3%	92.5%	97.0%	88.4%
4	2.34x10 ⁻³	11.24%	49.05%	50.32%	90.40%	2.34x10 ⁻³	43.86%	127.71%	135.03%	86.17%	2.34x10 ⁻³	109.8%	305.4%	324.5%	83.8%
5	2.36	-0.40%	10.57%	10.57%	89.80%	2.36	-1.09%	21.87%	21.90%	86.60%	2.36	-1.1%	32.5%	32.5%	84.6%
6	2.78x10	3.62%	10.98%	11.56%	98.37%	2.78x10	3.29%	16.13%	16.46%	99.13%	2.78x10	2.6%	20.9%	21.1%	98.8%
7	2.74x10 ⁻³	235%	666%	706%	93.33%	2.74x10 ⁻³	>10 ⁴ %	>10 ⁵ %	>10 ⁵ %	94.49%	2.74x10 ⁻³	>10 ³ %	>10 ⁴ %	>10 ⁴ %	92.8%
8	3.75x10	-2.57%	27.70%	27.82%	83.74%	3.75x10	-19.29%	30.29%	35.91%	82.11%	3.75x10	-25.4%	126.5%	129.1%	77.4%

Table 2.10: Mean absolute deviation (MAD) for 8 simulations with varying replications r ($r = 1, \dots, 10$)

r=1			
Scenario	MAD _{between}	$\frac{MAD_{below}}{MAD_{between}}$	$\frac{MAD_{above}}{MAD_{between}}$
1	5.10×10^{-2}	1.42	0.21
2	2.02×10^{-2}	2.75	0.76
3	2.12×10^{-2}	1.26	1.09
4	2.08×10^{-2}	1.95	1.94
5	1.84×10^{-2}	0.70	1.44
6	1.63×10^{-2}	0.72	30.21
7	1.95×10^{-2}	2.38	14.55
8	1.55×10^{-2}	0.81	32.13

r=2			
Scenario	MAD _{between}	$\frac{MAD_{below}}{MAD_{between}}$	$\frac{MAD_{above}}{MAD_{between}}$
1	4.87×10^{-2}	1.28	0.18
2	1.48×10^{-2}	2.62	0.79
3	1.50×10^{-2}	1.24	1.10
4	1.49×10^{-2}	1.88	1.86
5	1.31×10^{-2}	0.71	1.37
6	1.26×10^{-2}	0.67	37.34
7	1.41×10^{-2}	2.48	18.08
8	1.02×10^{-2}	0.77	48.46

r=3			
Scenario	MAD _{between}	$\frac{MAD_{below}}{MAD_{between}}$	$\frac{MAD_{above}}{MAD_{between}}$
1	4.59×10^{-2}	1.12	0.14
2	1.21×10^{-2}	2.55	0.77
3	1.23×10^{-2}	1.23	1.09
4	1.23×10^{-2}	1.82	1.79
5	1.06×10^{-2}	0.72	1.38
6	1.10×10^{-2}	0.62	34.51
7	1.15×10^{-2}	2.56	21.28
8	9.26×10^{-3}	0.75	47.61

r=4			
Scenario	MAD _{between}	$\frac{MAD_{below}}{MAD_{between}}$	$\frac{MAD_{above}}{MAD_{between}}$
1	4.48×10^{-2}	1.17	0.13
2	1.06×10^{-2}	2.60	0.76
3	1.07×10^{-2}	1.25	1.08
4	1.06×10^{-2}	1.89	1.79
5	9.16×10^{-3}	0.72	1.39
6	9.98×10^{-3}	0.59	38.34
7	9.95×10^{-3}	2.63	23.07
8	8.39×10^{-3}	0.75	49.53

r=5			
Scenario	MAD _{between}	$\frac{MAD_{below}}{MAD_{between}}$	$\frac{MAD_{above}}{MAD_{between}}$
1	4.43×10^{-2}	1.05	0.13
2	9.41×10^{-3}	2.60	0.77
3	9.47×10^{-3}	1.21	1.09
4	9.50×10^{-3}	1.89	1.85
5	8.22×10^{-3}	0.72	1.40
6	9.15×10^{-3}	0.58	39.75
7	8.94×10^{-3}	2.65	25.35
8	7.52×10^{-3}	0.74	54.28

r=10			
Scenario	MAD _{between}	$\frac{MAD_{below}}{MAD_{between}}$	$\frac{MAD_{above}}{MAD_{between}}$
1	4.04×10^{-2}	0.44	0.09
2	6.66×10^{-3}	2.60	0.75
3	6.74×10^{-3}	1.29	1.06
4	6.68×10^{-3}	1.96	1.79
5	5.88×10^{-3}	0.72	1.36
6	7.34×10^{-3}	0.52	32.43
7	6.26×10^{-3}	2.60	18.52
8	5.21×10^{-3}	0.69	61.77

Table 2.11: Performance Evaluation Statistics for BMDs with various λ ($\lambda = 0, 0.5$ and 1)

Scenario	$BMR_{AR} = 0.05, \lambda = 0$					$BMR_{AR} = 0.05, \lambda = 0.5$					$BMR_{AR} = 0.05, \lambda = 1$				
	BMD	RB	CV	RRMSE	Coverage	BMD	RB	CV	RRMSE	Coverage	BMD	RB	CV	RRMSE	Coverage
1	5.55×10^{-4}	84.4%	63.5%	105.7%	78.6%	5.55×10^{-4}	85.9%	58.3%	103.8%	98.9%	5.55×10^{-4}	86.5%	60.8%	105.7%	98.9%
2	5.26×10^{-4}	22.8%	83.4%	86.5%	88.2%	5.26×10^{-4}	7.4%	49.7%	50.3%	82.2%	5.26×10^{-4}	3.8%	41.0%	41.2%	72.4%
3	1.66×10^{-2}	14.0%	57.6%	59.3%	88.6%	1.66×10^{-2}	3.5%	26.2%	26.4%	67.3%	1.66×10^{-2}	1.6%	19.7%	19.7%	22.6%
4	2.34×10^{-3}	54.4%	154.7%	164.0%	85.4%	2.34×10^{-3}	14.0%	62.8%	64.4%	76.0%	2.34×10^{-3}	6.9%	46.3%	46.8%	40.4%
5	2.36	-1.3%	22.3%	22.3%	87.2%	2.36	-1.0%	14.3%	14.3%	70.6%	2.36	-1.0%	10.7%	10.7%	20.0%
6	2.78x10	-6.4%	13.8%	15.2%	99.6%	2.78x10	-0.3%	4.3%	4.3%	89.1%	2.78x10	<0.1%	1.0%	1.0%	82.2%
7	2.74×10^{-3}	> $10^3\%$	> $10^4\%$	> $10^6\%$	73.7%	2.74×10^{-3}	246.2%	490.5%	548.8%	91.5%	2.74×10^{-3}	80.7%	162.9%	181.8%	64.5%
8	3.75x10	-14.8%	19.3%	24.3%	96.7%	3.75x10	-0.1%	5.8%	5.8%	95.0%	3.75x10	-0.2%	1.5%	1.5%	81.5%

Scenario	$BMR_{ER} = 0.05, \lambda = 0$					$BMR_{ER} = 0.05, \lambda = 0.5$					$BMR_{ER} = 0.05, \lambda = 1$				
	BMD	RB	CV	RRMSE	Coverage	BMD	RB	CV	RRMSE	Coverage	BMD	RB	CV	RRMSE	Coverage
1	5.55×10^{-4}	83.9%	63.1%	105.0%	78.6%	5.55×10^{-4}	85.6%	58.0%	103.4%	100.0%	5.55×10^{-4}	86.2%	60.3%	105.2%	100.0%
2	5.26×10^{-4}	19.4%	74.8%	77.3%	88.6%	5.26×10^{-4}	6.6%	46.1%	46.5%	83.0%	5.26×10^{-4}	3.4%	37.9%	38.0%	73.2%
3	1.66×10^{-2}	12.9%	53.2%	54.7%	89.6%	1.66×10^{-2}	3.5%	24.4%	24.7%	67.2%	1.66×10^{-2}	1.6%	18.0%	18.0%	28.8%
4	2.34×10^{-3}	43.9%	127.7%	135.0%	86.2%	2.34×10^{-3}	11.8%	54.1%	55.4%	77.0%	2.34×10^{-3}	5.7%	39.2%	39.6%	39.4%
5	2.36	-1.1%	21.9%	21.9%	86.6%	2.36	-0.8%	14.0%	14.0%	71.5%	2.36	-0.8%	10.4%	10.5%	24.6%
6	2.78x10	3.3%	16.1%	16.5%	99.1%	2.78x10	1.6%	5.4%	5.6%	92.9%	2.78x10	0.2%	1.6%	1.6%	91.1%
7	2.74×10^{-3}	> $10^3\%$	> $10^4\%$	> $10^6\%$	73.9%	2.74×10^{-3}	148.6%	402.7%	429.2%	92.0%	2.74×10^{-3}	42.6%	94.0%	103.2%	83.0%
8	3.75x10	-19.3%	30.3%	35.9%	82.1%	3.75x10	1.7%	18.2%	18.3%	86.7%	3.75x10	-0.2%	3.7%	3.7%	99.0%

Table 2.12: Performance Evaluation Statistics for Hill Model Parameters with $\lambda = 0.5$. [In the case $\alpha = 0$, the bias, standard error, and root MSE are reported rather than RB, CV, and RRMSE.]

Scenario	Parameter Value	RB	CV	RRMSE	Coverage
α^*					
1	0	<0.1%	10.2%	10.2%	95.7%
2	0	-0.2%	2.9%	2.9%	83.8%
3	0	<0.1%	0.9%	0.9%	36.0%
4	0	-0.1%	2.0%	2.0%	65.8%
5	0	< 0.1 %	0.1%	0.1%	46.2%
6	0	<0.1%	0.0%	0.0%	55.4%
7	0	0.8%	2.2%	2.4%	78.8%
8	0	<0.1%	0.0%	0.0%	60.1%
θ					
1	1	0.4%	1.3%	1.4%	92.4%
2	1	0.2%	1.9%	1.9%	89.2%
3	1	0.3%	2.8%	2.8%	85.8%
4	1	0.8%	5.0%	5.1%	88.2%
5	1	0.5%	3.2%	3.2%	89.0%
6	1	7.9%	24.7%	25.9%	90.6%
7	1	-1.6%	38.3%	38.3%	75.9%
8	1	7.8%	53.5%	54.1%	85.6%
κ					
1	0.00	82.6%	41.0%	92.2%	48.9%
2	0.01	0.7%	12.2%	12.3%	2.8%
3	0.32	1.6%	11.2%	11.4%	74.8%
4	0.32	6.0%	28.9%	29.6%	81.4%
5	6.31	0.6%	4.9%	4.9%	99.6%
6	50.12	4.5%	15.5%	16.1%	98.2%
7	50.12	779.2%	>10 ³ %	>10 ³ %	99.2%
8	100	-1.1%	24.6%	24.6%	98.3%
η					
1	5	2.0%	20.5%	20.6%	98.9%
2	1	0.7%	13.6%	13.7%	84.8%
3	1	0.7%	9.2%	9.2%	69.0%
4	0.6	0.7%	12.1%	12.1%	81.2%
5	3	0.1%	15.7%	15.7%	72.8%
6	5	0.1%	19.1%	19.1%	83.5%
7	0.3	15.6%	28.5%	32.5%	88.9%
8	3	8.4%	23.8%	25.3%	96.3%

Table 2.13: Performance Evaluation Statistics for Hill Model Parameters with $\lambda = 1$. [In the case $\alpha = 0$, the bias, standard error, and root MSE are reported rather than RB, CV, and RRMSE.]

Scenario	Parameter Value	RB	CV	RRMSE	Coverage
α^*					
1	0	-1.3%	14.9%	15.0%	95.6%
2	0	-0.3%	2.3%	2.4%	67.4%
3	0	< 0.1 %	0.7%	0.7%	2.2%
4	0	-0.1%	1.6%	1.6%	22.0%
5	0	< 0.1 %	0.1%	0.1%	0.0%
6	0	<0.1%	0.0%	0.0%	0.0%
7	0	0.4%	1.3%	1.4%	60.3%
8	0	< 0.1 %	0.0%	0.0%	0.0%
θ					
1	1	0.3%	1.3%	1.3%	95.6%
2	1	0.2%	1.9%	1.9%	89.2%
3	1	0.3%	2.7%	2.7%	84.0%
4	1	0.7%	4.6%	4.7%	82.8%
5	1	0.4%	3.1%	3.2%	87.8%
6	1	0.9%	5.8%	5.9%	89.1%
7	1	-0.3%	27.6%	27.6%	75.1%
8	1	0.5%	8.3%	8.3%	99.0%
κ					
1	0.00	84.5%	41.5%	94.1%	43.3%
2	0.01	0.3%	8.9%	9.0%	1.8%
3	0.32	1.3%	9.4%	9.5%	61.0%
4	0.32	4.8%	22.9%	23.4%	75.4%
5	6.31	0.6%	4.1%	4.2%	100.0%
6	50.12	0.4%	3.0%	3.0%	100.0%
7	50.12	230.6%	888.2%	917.6%	99.8%
8	100	-0.2%	3.1%	3.1%	100.0%
η					
1	5	0.5%	22.2%	22.2%	98.9%
2	1	0.2%	12.1%	12.1%	78.4%
3	1	0.3%	7.6%	7.6%	35.6%
4	0.6	0.1%	10.1%	10.1%	53.6%
5	3	-0.3%	12.7%	12.7%	11.6%
6	5	-0.1%	3.2%	3.2%	72.8%
7	0.3	7.0%	17.7%	19.1%	75.9%
8	3	0.1%	2.6%	2.6%	55.8%

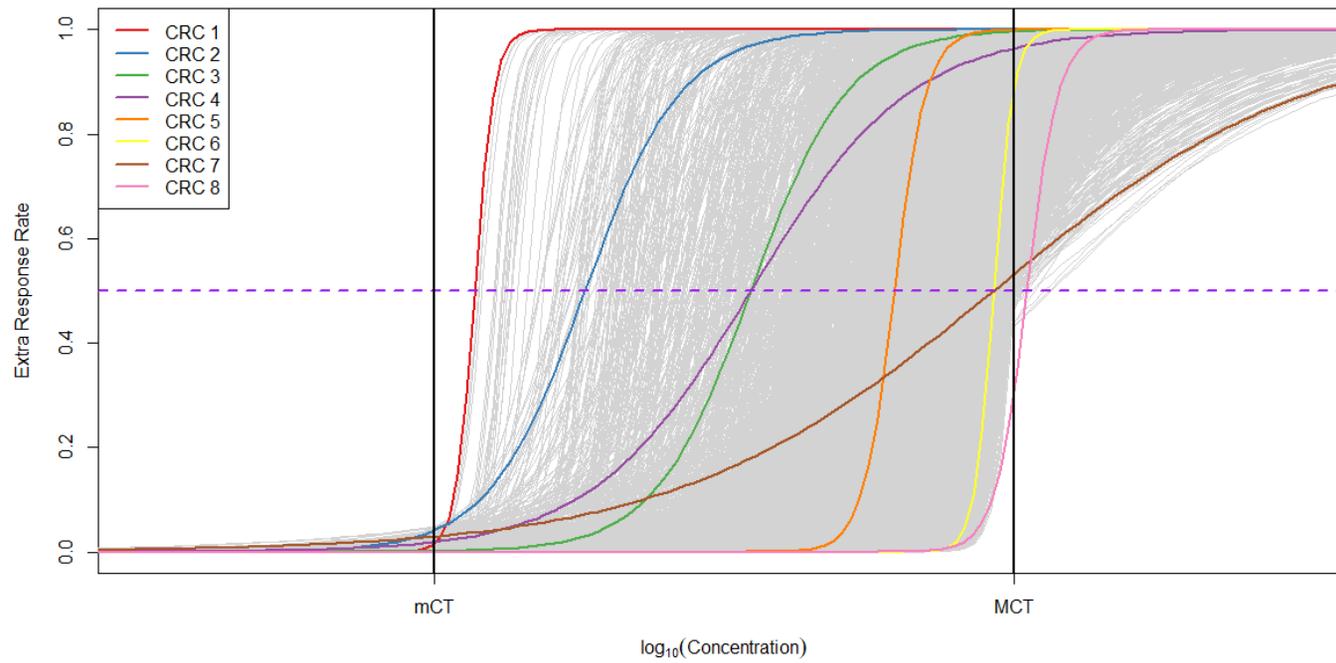


Figure 2.1: Eight CRCs for simulation study and estimated CRCs from 8,456 qHTS experiments

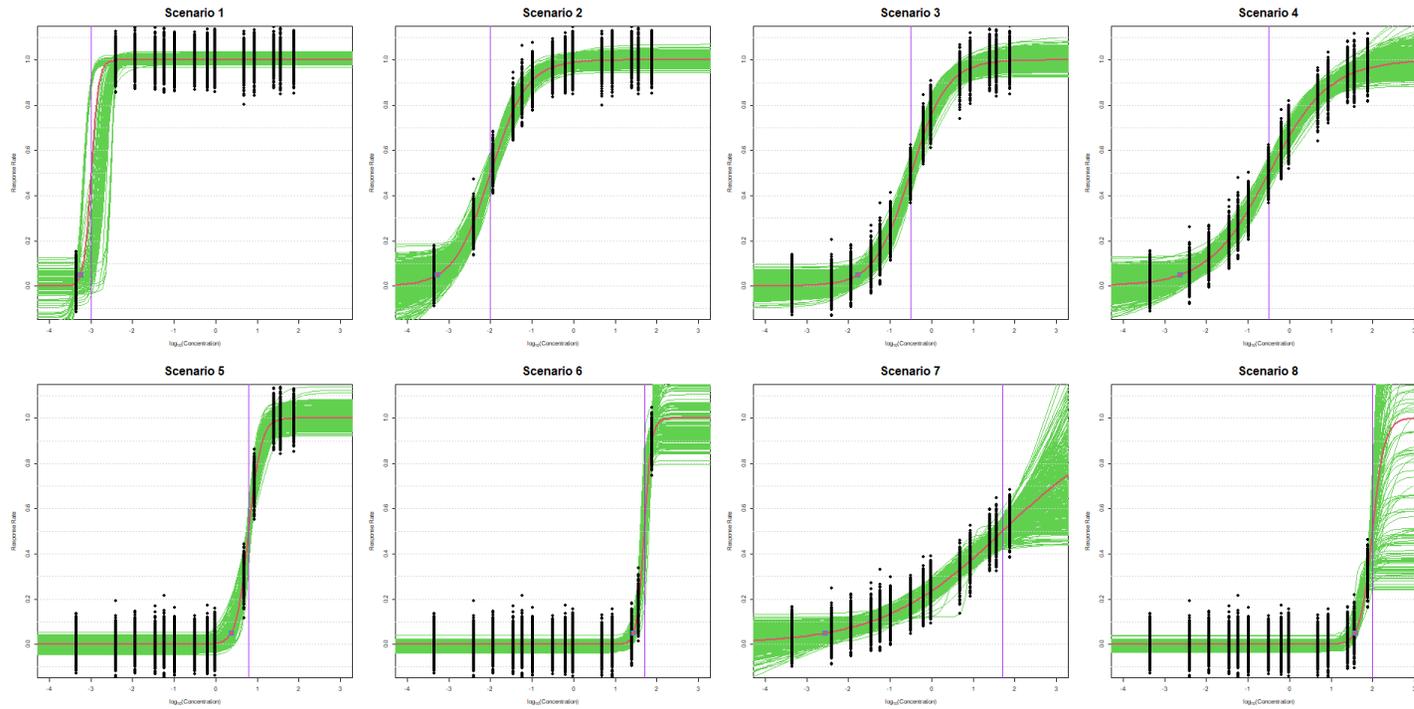


Figure 2.2: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.05$, $\lambda = 0$, and no replicates

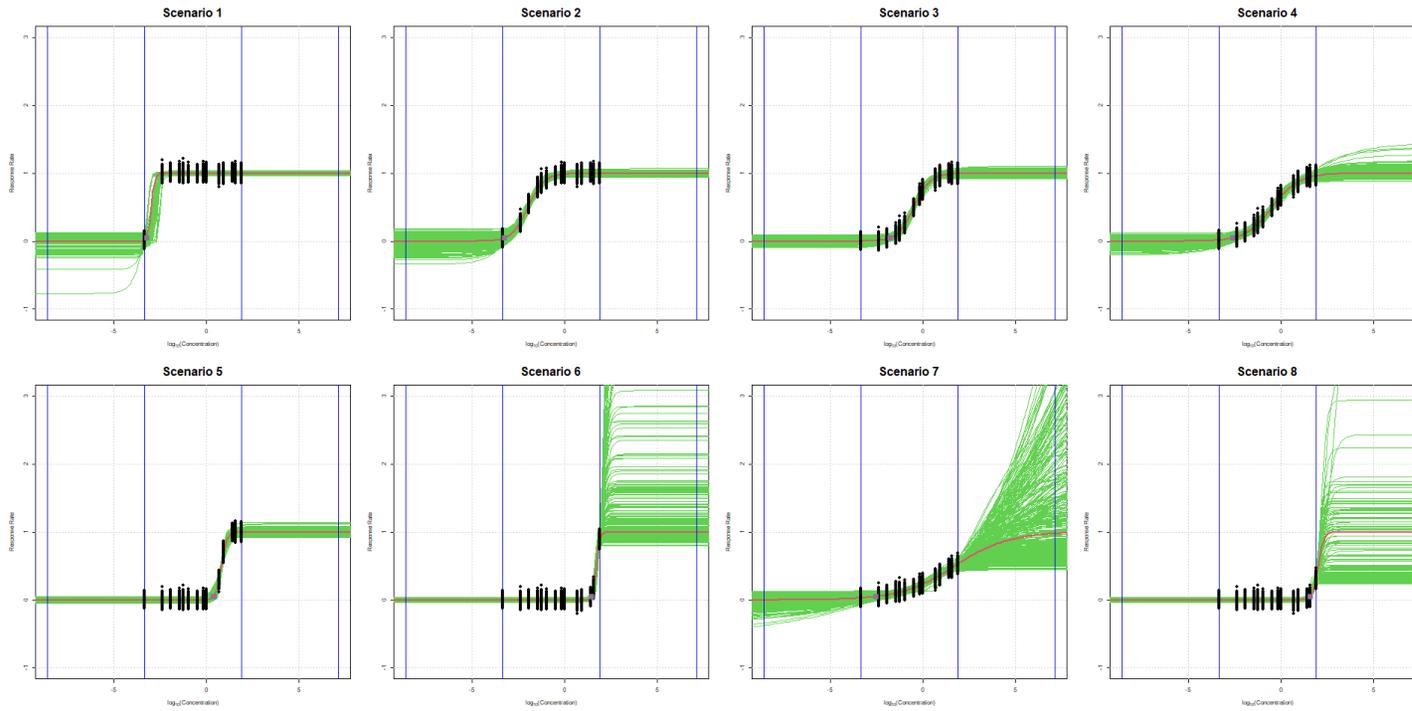


Figure 2.3: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.05$, $\lambda = 0$, and no replicates

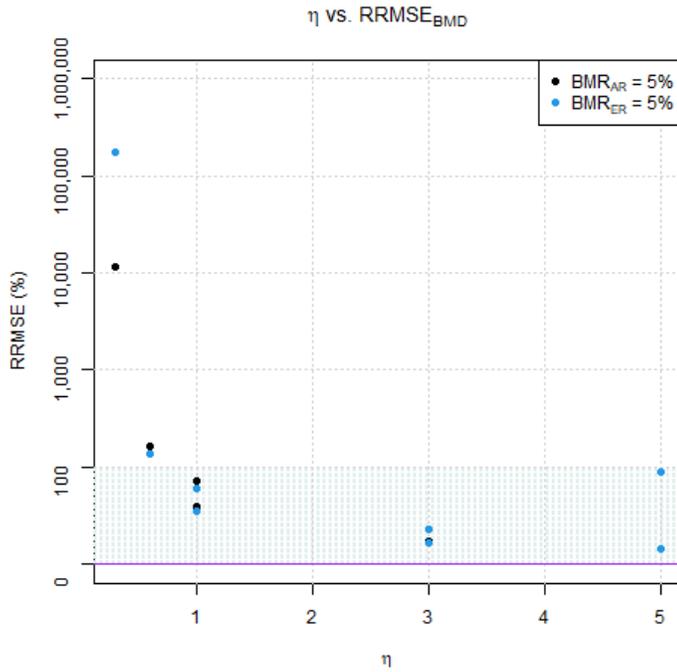


Figure 2.4: RRMSE for \widehat{BMD} against various values of η for each of the 8 Scenarios. (Note that some values of η repeat.) Note that the graph is given in arithmetic scale between 0 to 100, while the log-10 scale is used between 100 to 1,000,000 for RRMSE.

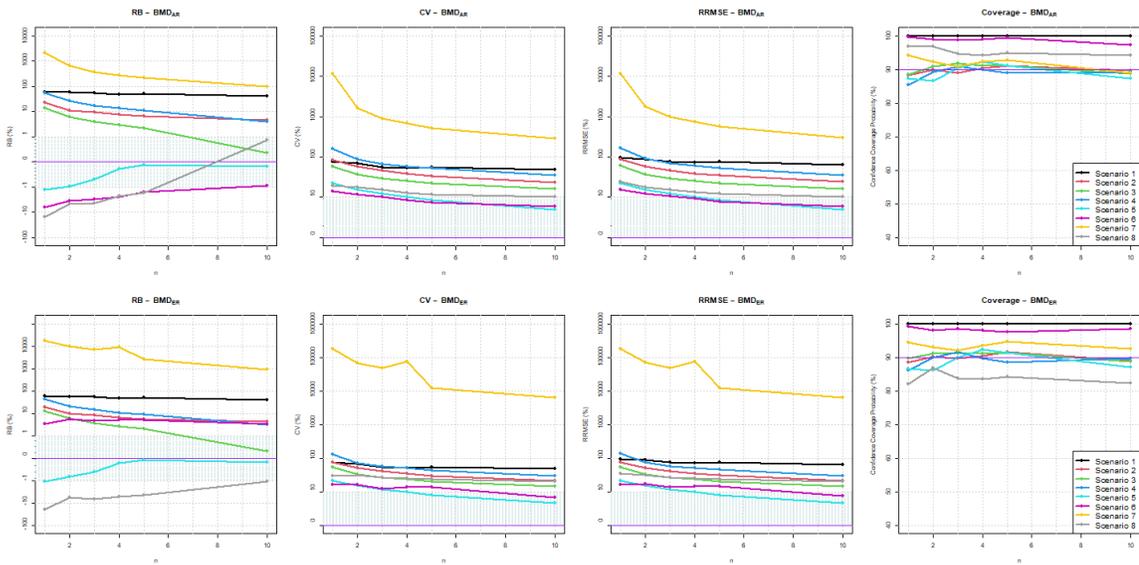


Figure 2.5: Performance evaluation statistics for \widehat{BMD} vs. number of replicates. Note that for RB, CV, and RRMSE, the graphs are given in log-10 scale, except between 0 to 1 (for RB) and 0 to 10 (for CV and RRMSE), where they are expressed in arithmetic scale.

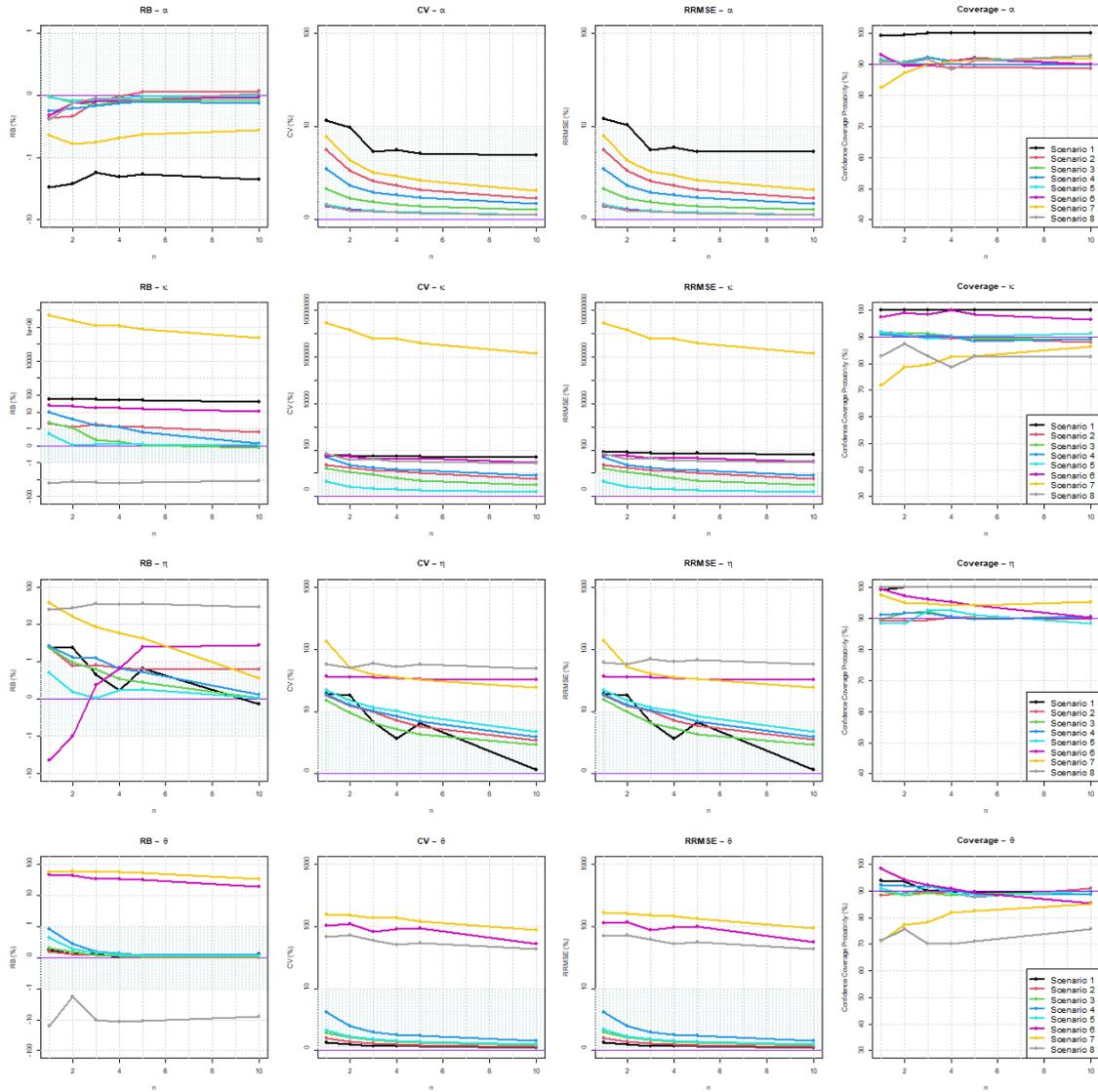


Figure 2.6: Performance evaluation statistics for Hill model parameter estimates vs. number of replicates. Note that for RB, CV, and RRMSE, the graphs are given in log-10 scale, except between 0 to 1 (for RB) and 0 to 10 (for CV and RRMSE), where they are expressed in arithmetic scale.

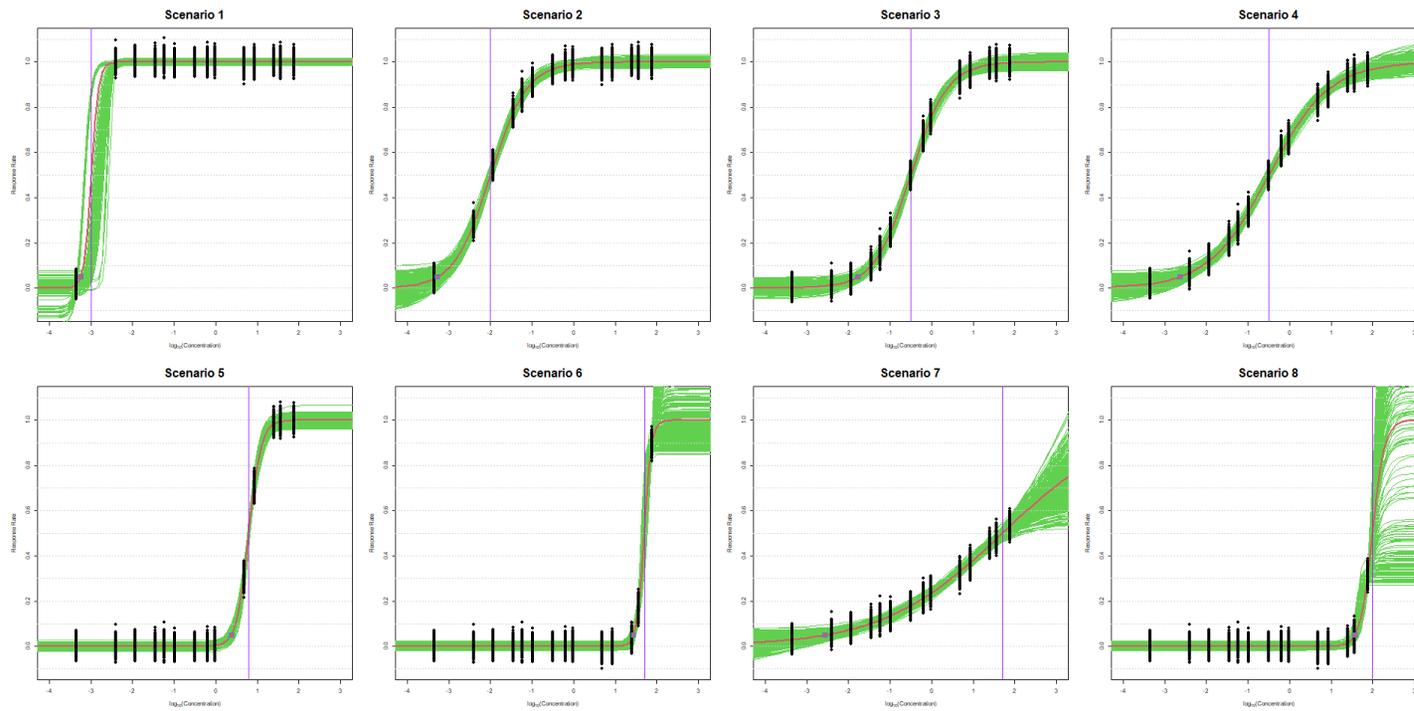


Figure 2.7: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.025$, $\lambda = 0$, and no replicates

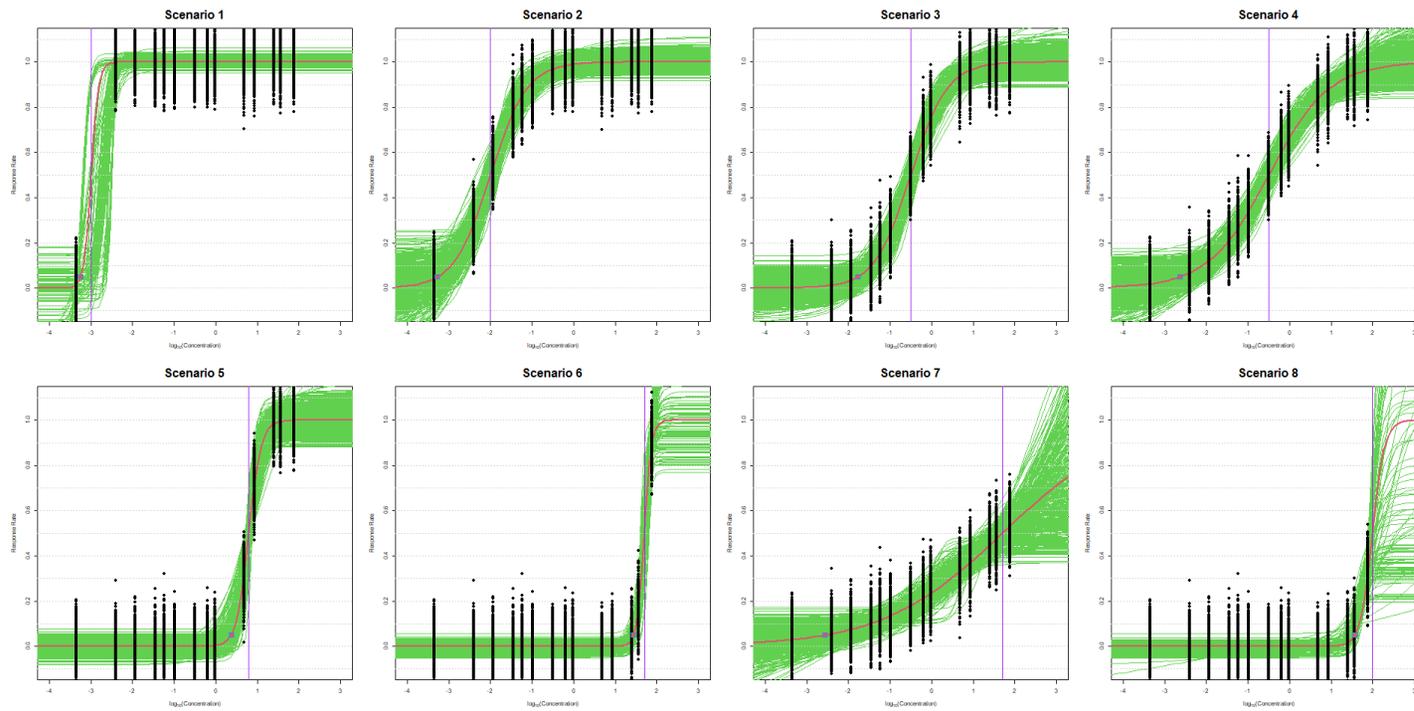


Figure 2.8: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.075$, $\lambda = 0$, and no replicates

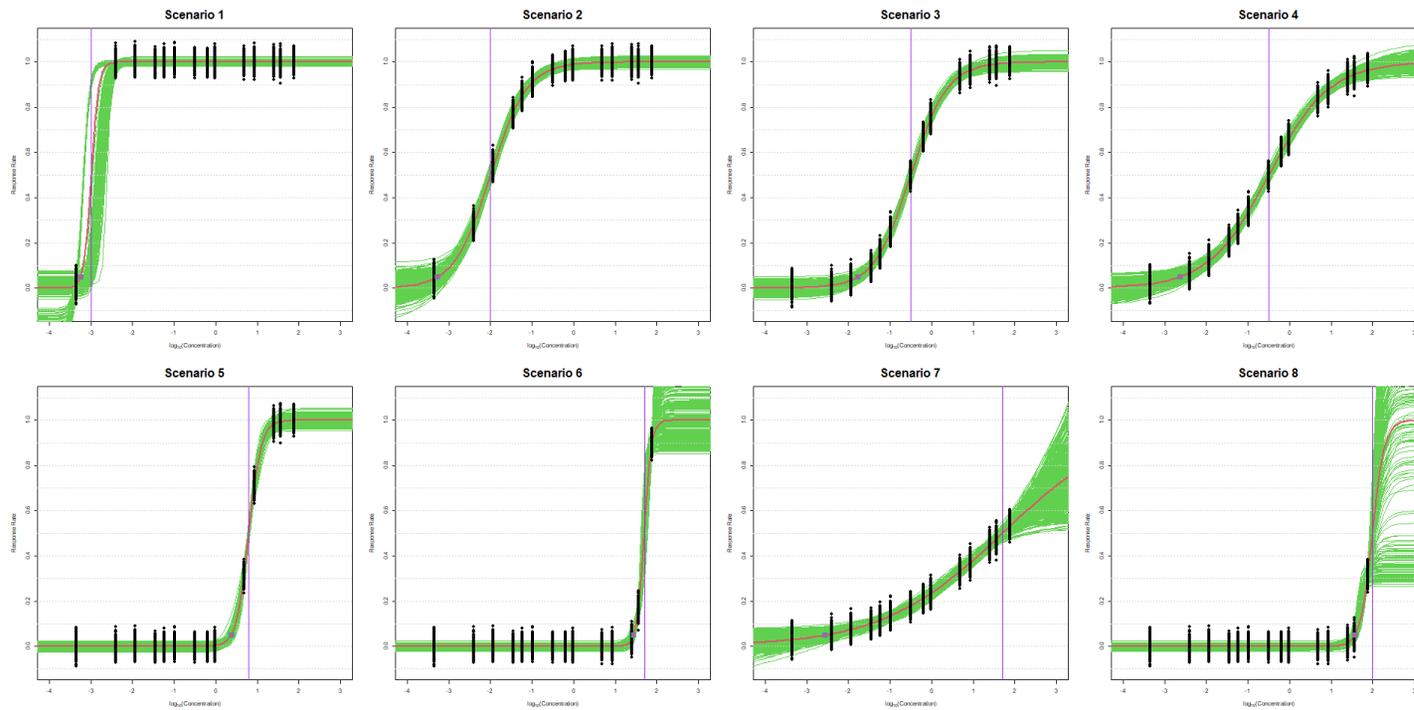


Figure 2.9: Estimated CRCs of simulated experiments for eight scenarios with $\sigma_0 = 0.05$, $\lambda = 0$, and $r = 4$ replicat

A.2. Tables and Figures for Chapter 3

Table 3.1: Six fixed designs for BMD_{ER} analysis.

Design	Dose Spacing	Number of Doses (control + treatment)	Design Parameters*	Experimental Concentrations														
				x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
FD ₁	A typical design	1+14	Dose levels (%)	0.000	0.001	0.005	0.015	0.045	0.074	0.14	0.4	0.8	1.3	6.3	11.1	32.8	47.3	100.0
			Dose allocations (%)	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7
FD ₂	Equispread (log scale)	1+14	Dose levels (%)	0.000	0.001	0.003	0.007	0.017	0.041	0.10	0.2	0.6	1.3	3.1	7.4	17.7	42.0	100.0
			Dose allocations (%)	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7
FD ₃	Equispread (log scale)	1+9	Dose levels (%)	0.000	0.001	0.005	0.021	0.088	0.4	1.5	6.0	24.5	100.0					
			Dose allocations (%)	6.7	6.7	6.7	6.7	6.7	12.9	12.9	12.9	12.9						
FD ₄	Equispread (log scale)	1+4	Dose levels (%)	0.000	0.001	0.055	2.3	100.0										
			Dose allocations (%)	20	20	20	20	20										
FD ₅ =ACOD _{ave}	Average optimal (log scale)	1+3	Dose levels (%)	0.00	11.6	28.9	93.3											
			Dose allocations (%)	33.3	46.7	13.3	6.7											
FD ₆ =ACOD _{med}	Median optimal (log scale)	1+3	Dose levels (%)	0.00	20.4	44.9	100.0											
			Dose allocations (%)	33.3	46.7	13.3	6.7											

*Dose levels expressed as a percentage of the MCT.

Table 3.2: Relative Efficiencies for six fixed designs for BMD_{ER} based on 8,456 data.

Design	avg RE	med RE	min RE	max RE	Q1	Q3	Proportion of singularities	Proportion of outliers
FD ₁	22.2%	20.0%	0.0%	50.0%	12.3%	34.5%	0.0%	0.0%
FD ₂	18.6%	18.3%	0.0%	47.8%	2.9%	31.3%	0.0%	0.0%
FD ₃	15.1%	2.3%	0.0%	55.5%	0.0%	32.8%	0.3%	0.0%
FD ₄	4.9%	0.0%	0.0%	62.9%	0.0%	1.0%	59.2%	8.3%
FD ₅ =ACOD _{ave}	20.3%	4.3%	0.0%	97.1%	0.2%	37.0%	1.6%	0.4%
FD ₆ =ACOV _{med}	25.5%	12.4%	0.0%	97.2%	1.3%	46.9%	2.1%	0.0%

Table 3.3: Six fixed designs for BMD_{AR} analysis.

Design	Dose Spacing	Number of Doses (control + treatment)	Design Parameters*	Experimental Concentrations														
				x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
FD ₁	A typical design	1+14	Dose levels (%)	0.000	0.001	0.005	0.015	0.045	0.074	0.14	0.4	0.8	1.3	6.3	11.1	32.8	47.3	100.0
			Dose allocations (%)	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7
FD ₂	Equispread (log scale)	1+14	Dose levels (%)	0.000	0.001	0.003	0.007	0.017	0.041	0.10	0.2	0.6	1.3	3.1	7.4	17.7	42.0	100.0
			Dose allocations (%)	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7	6.7
FD ₃	Equispread (log scale)	1+9	Dose levels (%)	0.000	0.001	0.005	0.021	0.088	0.4	1.5	6.0	24.5	100.0					
			Dose allocations (%)	6.7	6.7	6.7	6.7	6.7	12.9	12.9	12.9	12.9	12.9					
FD ₄	Equispread (log scale)	1+4	Dose levels (%)	0.000	0.001	0.055	2.3	100.0										
			Dose allocations (%)	20	20	20	20	20										
FD ₅ =ACOD _{ave}	Average optimal (log scale)	1+3	Dose levels (%)	0.00	10.3	29.8	100.0											
			Dose allocations (%)	33.3	46.7	13.3	6.7											
FD ₆ =ACOD _{med}	Median optimal (log scale)	1+3	Dose levels (%)	0.00	17.5	45.8	100.0											
			Dose allocations (%)	33.3	46.7	13.3	6.7											

*Dose levels expressed as a percentage of the MCT.

Table 3.4: Relative Efficiencies for six fixed designs for BMD_{AR} based on 8,456 data.

Design	avg RE	med RE	min RE	max RE	Q1	Q3	Proportion of singularities	Proportion of outliers
FD ₁	24.0%	21.7%	0.0%	54.3%	14.5%	35.6%	0.0%	0.0%
FD ₂	20.4%	19.1%	0.0%	50.3%	3.6%	34.6%	0.0%	0.0%
FD ₃	15.3%	4.1%	0.0%	56.7%	0.0%	30.9%	0.3%	0.0%
FD ₄	6.2%	0.0%	0.0%	65.8%	0.0%	4.5%	59.2%	7.5%
FD ₅ =ACOD _{ave}	22.1%	5.7%	0.0%	98.8%	0.1%	39.9%	1.5%	0.0%
FD ₆ =ACOV _{med}	23.9%	8.5%	0.0%	99.1%	0.7%	42.8%	2.0%	0.0%

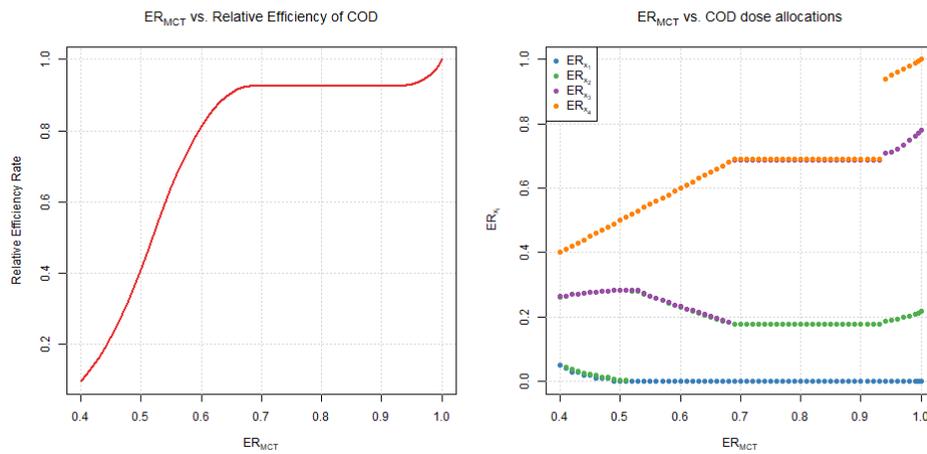


Figure 3.1: Relative efficiency of the COD for BMD_{ER} as a function of ER_{MCT} (left) and Dose allocations used in COD for BMD_{ER} as a function of ER_{MCT} (right)

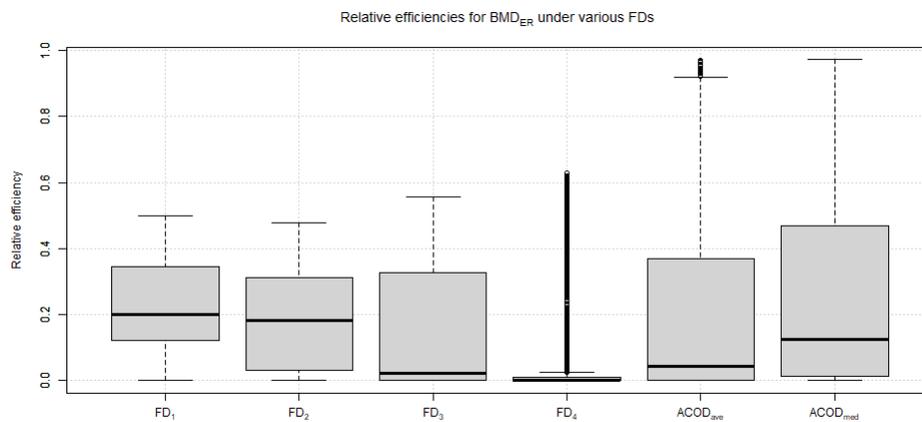


Figure 3.2: Relative efficiencies for BMD_{ER} using 8,456 data. Note that FD_3 to $ACOD_{med}$ produced non-invertible information matrices in some of the experiments, and hence only those with convergent results are included in these boxplots.

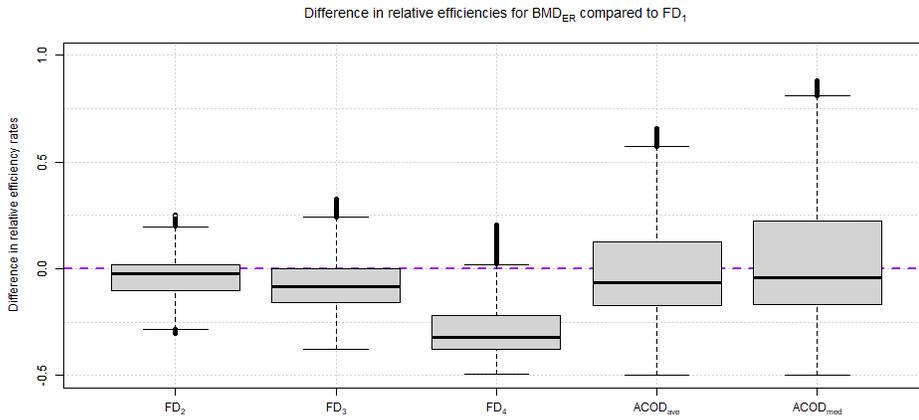


Figure 3.3: Differences in relative efficiencies for BMD_{ER} compared to FD_1 using 8,456 data. Note that FD_3 to $ACOD_{med}$ produced non-invertible information matrices in some of the experiments, and hence only those with convergent results are included in these boxplots.

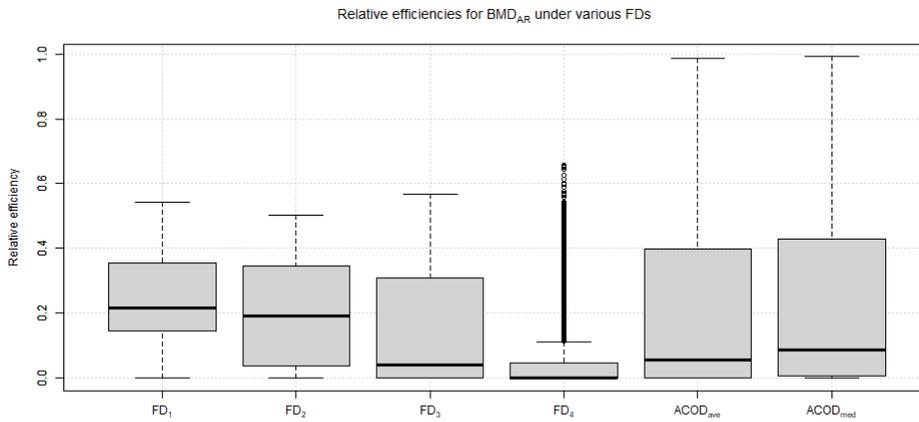


Figure 3.4: Relative efficiencies for BMD_{AR} using 8,456 data. Note that FD_3 to $ACOD_{med}$ produced non-invertible information matrices in some of the experiments, and hence only those with convergent results are included in these boxplots.

A.3. Tables and Figures for Chapter 4

Table 4.1: Parameters used to characterize illustrative VOI calculation scenarios

Parameter	Description	Value	Unit
N	Population size	100	Millions of people
TH	Time horizon	20	Years
t_{imp}	Time to implement regulation	2	Years
t_{eff}	Time for regulation to achieve a reduction in risk	2	Years
t_A	Time to conduct Test A	2	Years
t_B	Time to conduct Test B	10	Years
ICC_A	Information collection cost for Test A	0.5	\$M/test
ICC_B	Information collection cost for Test B	5	\$M/test
r	Discount rate	5	Percent
CC_{max}	Maximum annualized control cost	50	\$M/year
CC_{OH}	Annualized overhead control cost	0	\$M/year
V	Annualized value of a statistical life	0.13*	\$M/person/year
B	Annualized testing budget (Test A or Test B)	20	\$M/year
TRL	Target risk level	10^{-7}	(unitless)

* Based on a VSL of \$9.1M annualized over 70-year lifetime.

Table 4.2: Parameters governing toxicity and exposure of a chemical

Parameter Name	True Parameter (log ₁₀)	Prior Uncertainty Parameter (log ₁₀)
μ_{tox}	-2	1
σ_{tox}	1	0
μ_{exp}	-8	0
σ_{exp}	0.5	0
R	-7.40	2.15

Table 4.3: Information regarding Test A and Test B

Experiment	A	B
t_{1C}	2 years	10 years
σ_{test} for μ_{tox}	$u(\mu_{\text{tox,pr}})/2$	$u(\mu_{\text{tox,pr}})/10$
$u(\mu_{\text{tox,pos}})$	$u(\mu_{\text{tox,pr}})/\sqrt{5}$	$u(\mu_{\text{tox,pr}})/\sqrt{101}$

Table 4.4: VOI analysis under target-risk decision making

Item	Test A	Test B
$P(\text{Correct decision is made})$	7.9%	57.6%
$P(\text{Incorrect decision is made})$	1%	0%
$P(\text{No decision is made})$	91.1%	42.4%
$E(\text{ICC per correct decision})$	\$6.3M	\$8.7M
$E(\text{Number of correct decisions per year})$	3.2	2.3

Table 4.5: VOI analysis under benefit-risk decision making

Item	True ^a	Prior ^a	Test A	Test B
ORE	12.5%	93.5%	38.5%	16.0%
ETSC	\$6.1M	\$1,906.5M	\$38.5M	\$7.2M
EVPPi	NA	\$361.5M	\$321.7M	\$333.1M
EVSI	NA	NA	\$314.6M	\$333.1M (\$3.4M) ^b
ENBS	NA	NA	\$314.1M	\$328.2M
ROI	NA	NA	\$628	\$66

a. NA=not applicable.

b. differential EVSI between Test A and Test B.

Table 4.6: Parameters governing toxicity and exposure of a chemical

Parameter Name	True Parameter (log ₁₀)	Prior Uncertainty Parameter (log ₁₀)
μ_{tox}	-2.5	0.5
σ_{tox}	1	0
μ_{exp}	-8	0
σ_{exp}	0.5	0
R	-6.36	0.99

Table 4.7: Various VOI outcomes based on TRDM

Item	Test A	Test B
$P(\text{Correct decision is made})$	38.1%	100%
$P(\text{Incorrect decision is made})$	0.1%	0%
$P(\text{No decision is made})$	61.8%	0%
$E(\text{ICC per correct decision})$	\$1.3M	\$5.0M
$E(\text{Number of correct decisions per year})$	15.2	4

Table 4.8: VOI analysis under benefit-risk decision making

Item	True ^a	Prior ^a	Test A	Test B
ORE	40.5%	67.5%	48.5%	43.0%
ETSC	\$45.8M	\$282.4M	\$76.8M	\$65.2M
EVPPPI	NA	\$59.3M	\$24.0M	\$16.1M
EVSI	NA	NA	\$19.9M	\$16.0M (-\$12.5M) ^b
ENBS	NA	NA	\$19.4M	\$11.0M
ROI	NA	NA	\$38.8	\$2.2

a. NA=not applicable.

b. differential EVSI between Test A and Test B.

Table 4.9: Parameters governing toxicity and exposure of a chemical

Parameter Name	True Parameter (log ₁₀)	Prior Uncertainty Parameter (log ₁₀)
μ_{tox}	-1.5	0.5
σ_{tox}	1	0
μ_{exp}	-8	0
σ_{exp}	0.5	0
R	-8.52	1.15

Table 4.10: Various VOI outcomes based on TRDM

Item	Test A	Test B
$P(\text{Correct decision is made})$	94.7%	100%
$P(\text{Incorrect decision is made})$	0%	0%
$P(\text{No decision is made})$	5.3%	0%
$E(\text{ICC per correct decision})$	\$0.52M	\$5.0M
$E(\text{Number of correct decisions per year})$	37.9	4.0

Table 4.11: VOI analysis under benefit-risk decision making

Item	True ^a	Prior ^a	Test A	Test B
ORE	0%	18%	0%	0%
ETSC	\$0.5M	\$8.4M	\$1.0M	\$0.5M
EVPPi	NA	\$2.6M	\$1.1M	\$1.2M
EVSI	NA	NA	\$1.1M	\$1.2M (\$0M) ^b
ENBS	NA	NA	\$0.6M	-\$3.9M
ROI	NA	NA	\$1.1	-\$0.8

a. NA=not applicable.

b. differential EVSI between Test A and Test B.

Table 4.12: EVSI for various values of uncertainty reduction and timeliness of testing over 20-year time horizon under three scenarios (in \$M)

Testing Time (years)	Ratio of σ_{Test} to $u(\mu_{\text{tox,prior}})$											
	Scenario 1 ^a				Scenario 2 ^b				Scenario 3 ^c			
	0%	10%	50%	90%	0%	10%	50%	90%	0%	10%	50%	90%
0	334	334	317	254	35	35	28	18	1.2	1.2	1.1	0.8
2	334	333	314	226	31	30	20	3	1.2	1.2	1.1	0.8
4	334	333	312	202	27	26	13	-11	1.2	1.2	1.1	0.8
6	334	333	310	179	23	22	6	-23	1.2	1.2	1.1	0.8
8	334	333	308	159	20	19	1	-34	1.2	1.2	1.1	0.8
10	334	333	306	141	17	16	-5	-45	1.2	1.2	1.1	0.8

a. Higher uncertainty; optimal decision is to reduce exposure by 12.5%.

b. Lower uncertainty; optimal decision is to reduce exposure by 40.5%.

c. Lower uncertainty; optimal decision is not to reduce exposure.

Table 4.13: EVOI values for various values of the ratio CC_{max}/V^*

CC_{max}/V per capita	ORE				TSC	EVSI		ROI	
	True	Prior	Test A	Test B	True	Test A	Test B	Test A	Test B
10^{-6}	56.0%	81.0%	63.0%	58.0%	\$35.9M	\$7.6M	-\$3.1M	\$14.2	-\$1.6
10^{-5}	29.5%	57.0%	37.0%	31.5%	\$54.1M	\$29.8M	\$31.4M	\$58.5	\$5.3
10^{-4}	0.5%	28.5%	8.5%	3.0%	\$70.3M	\$47.7M	\$59.8M	\$94.4	\$11.0
10^{-3}	0.0%	0.0%	0.0%	0.0%	\$70.3M	\$0.0M	\$0.0M	-\$1.0	-\$1.0

*Based on Scenario 2

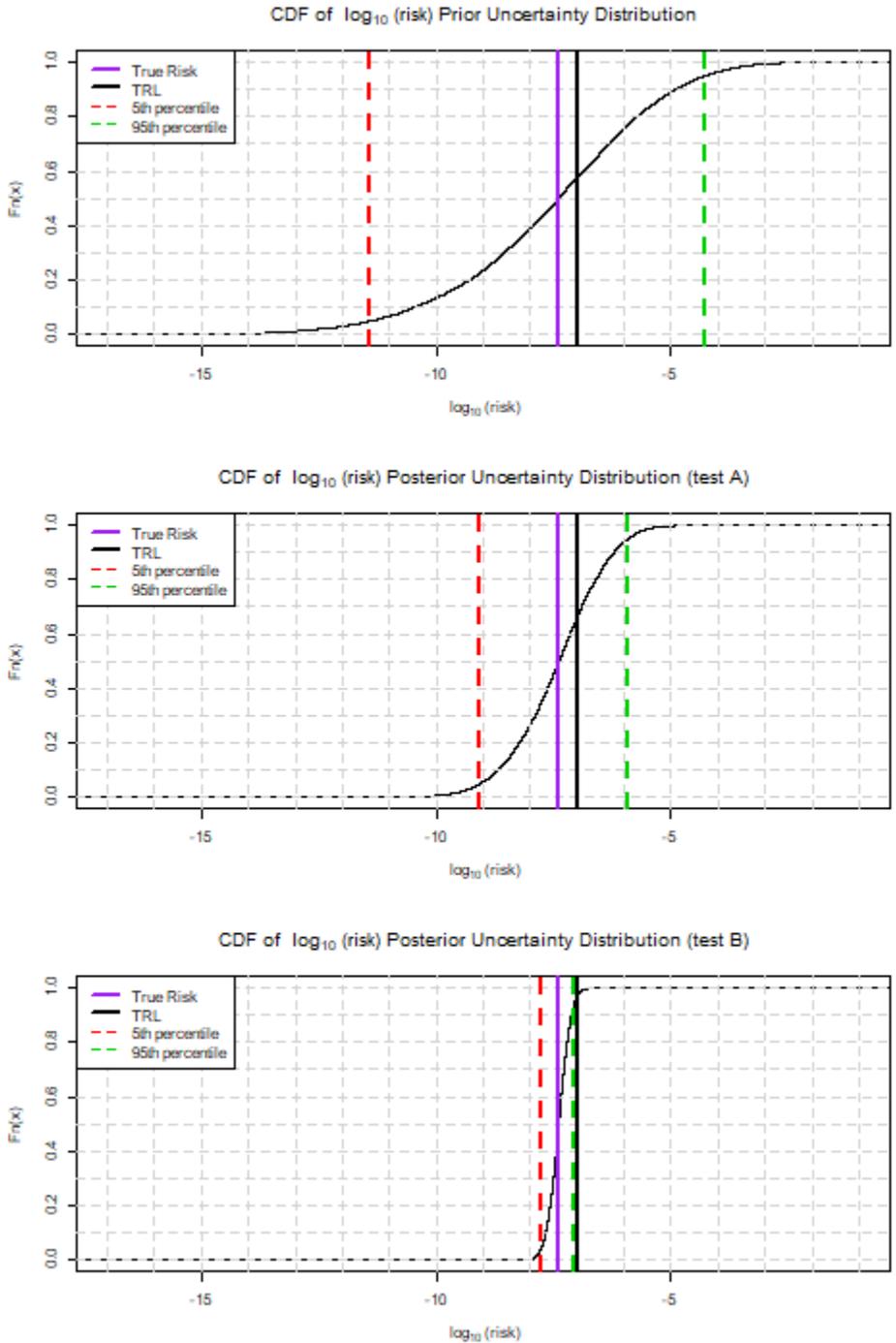


Figure 4.1: Uncertainty distributions for Scenario 1

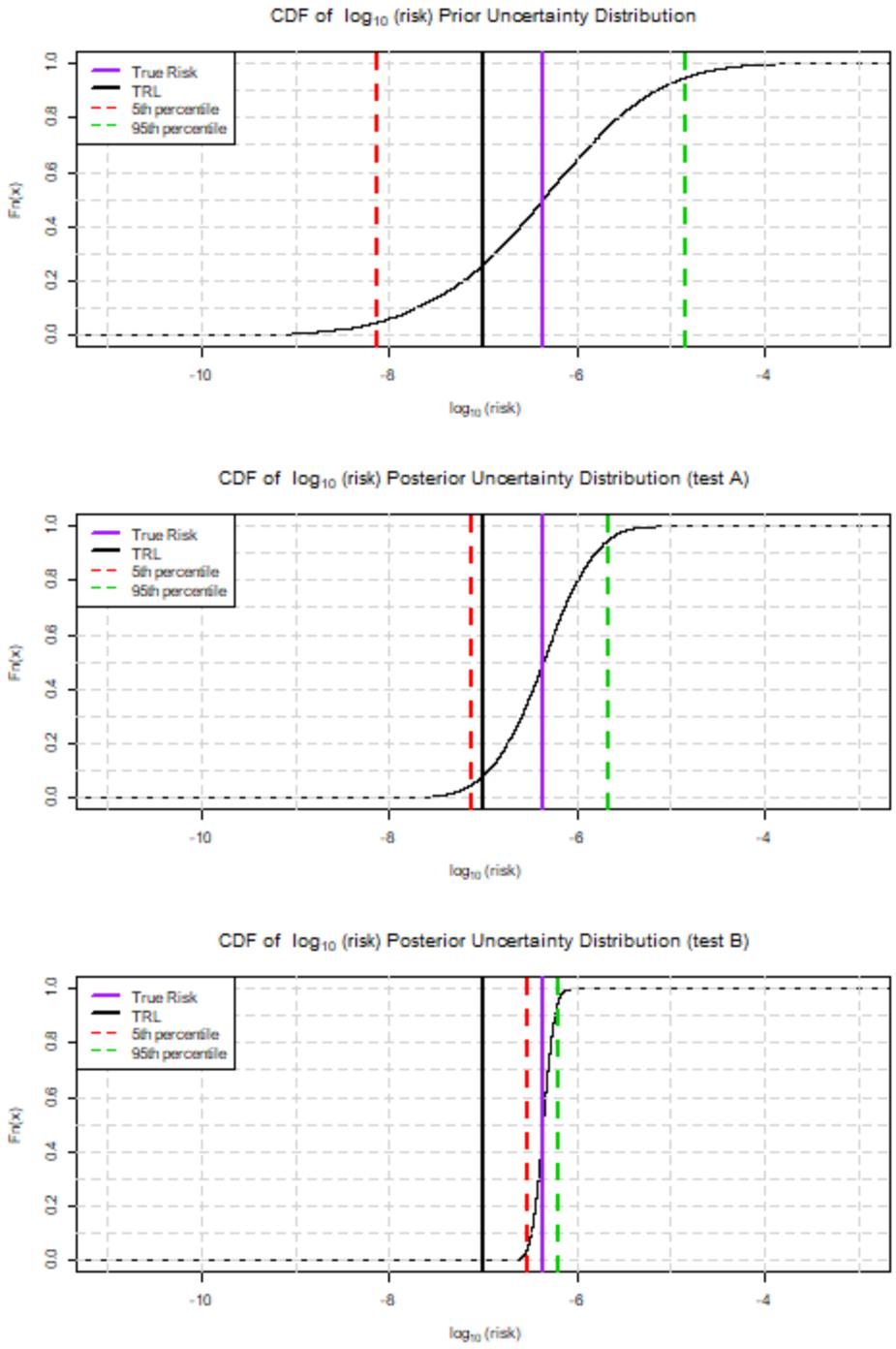


Figure 4.2: Uncertainty distributions for Scenario 2

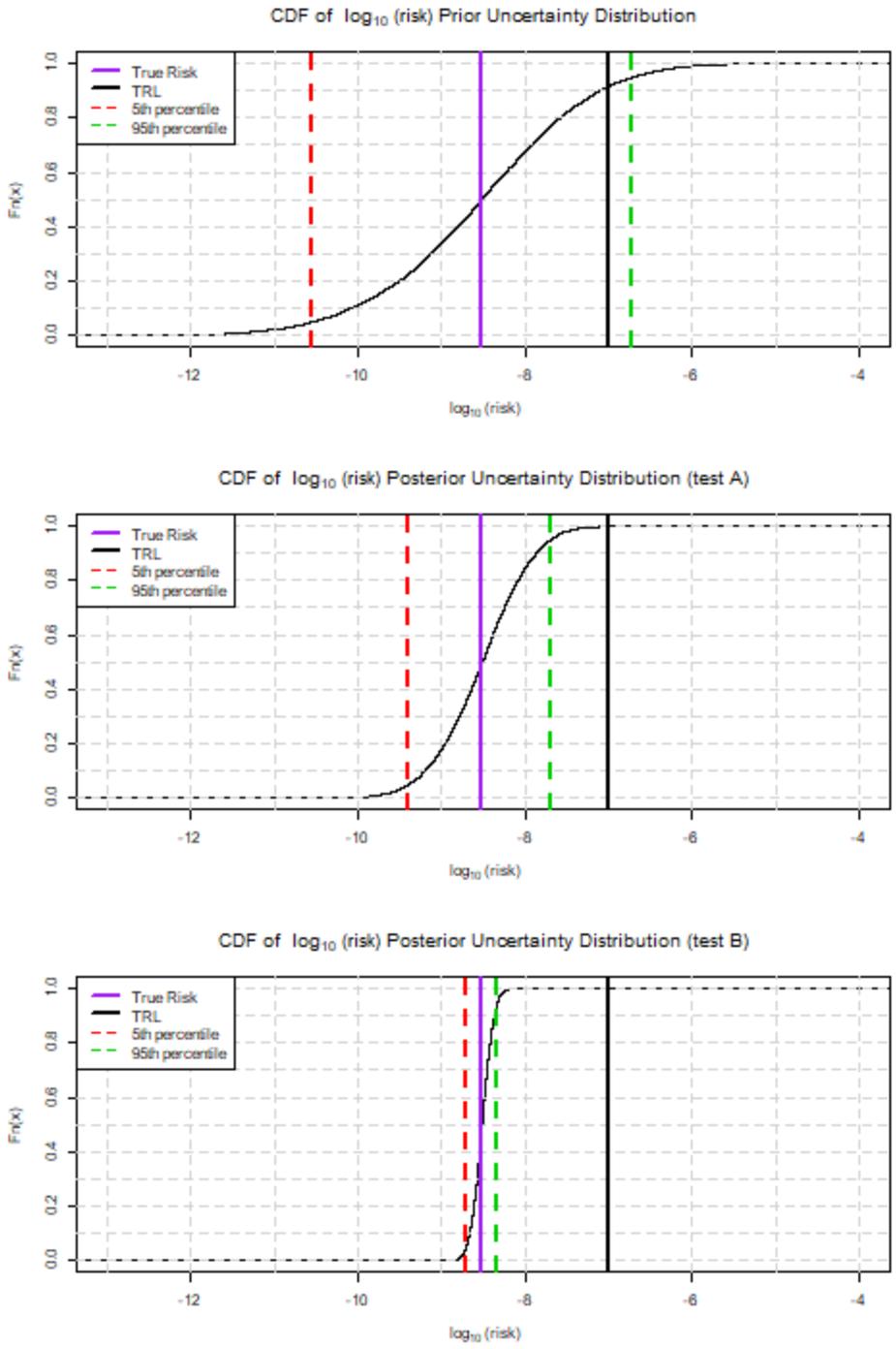


Figure 4.3: Uncertainty distributions for Scenario 3

Appendix B: Selected Proofs and Verifications

B.1. Regularity Conditions and Identifiability

B.1.1. Identifiability

The following section provides the proofs for deriving the approximate asymptotic variance of the estimated BMR_{AR} and BMR_{ER} . Note that the definitions, theorems, lemmas, and corollaries given below are quoted from Krewski and Ryzin (1981), with appropriate changes to notation.

Definition:

A concentration response model $f(x|\boldsymbol{\phi})$ is said to be identifiable if $f(x|\boldsymbol{\phi}) = f(x|\boldsymbol{\phi}') \forall x \geq 0$ implies that $\boldsymbol{\phi} = \boldsymbol{\phi}'$. If the model is not identifiable, then the maximum likelihood estimator of $\boldsymbol{\phi}$ given below may not be unique.

For the four-parameter Hill model, the proof is given as follows.

Proof:

Let

$$f(x|\boldsymbol{\phi}) = \alpha + \frac{\theta - \alpha}{(1 + (\kappa/x)^\eta)} = \alpha' + \frac{\theta' - \alpha'}{(1 + (\kappa'/x)^{\eta'})} = f(x|\boldsymbol{\phi}') \forall x \geq 0.$$

Then, it is clear that $\alpha = \alpha'$, since when $x = 0$,

$$f(0|\boldsymbol{\phi}) = \alpha = \alpha' = f(0|\boldsymbol{\phi}').$$

Similarly, $\theta = \theta'$ must hold as $x \rightarrow \infty$,

$$\lim_{x \rightarrow \infty} f(x|\boldsymbol{\phi}) = \theta = \theta' = \lim_{x \rightarrow \infty} f(x|\boldsymbol{\phi}').$$

For any x that is positive and finite, we have

$$f(x|\boldsymbol{\phi}) = f(x|\boldsymbol{\phi}') \Rightarrow$$

$$(\kappa/x)^\eta = (\kappa'/x)^{\eta'} \Rightarrow$$

$$\frac{\kappa^\eta}{x^\eta} = \frac{\kappa'^{\eta'}}{x^{\eta'}} \Rightarrow$$

$$\kappa^\eta = \kappa'^{\eta'} (x^{\eta-\eta'}) \Rightarrow$$

$$\eta = \eta' ,$$

which also implies that $\kappa = \kappa'$. Hence, the model is identifiable.

B.1.2. Regularity Conditions

The following regularity conditions are required to establish the desired asymptotic properties of the maximum likelihood estimator $\hat{\boldsymbol{\phi}}$ for $\boldsymbol{\phi}$.

C.1 $\lim_{n \rightarrow \infty} \frac{n_i}{n} = c_i$, where $c_i \in (0,1) \forall i = 1, \dots, t$ and $\sum_{i=1}^t c_i = 1$.

C.2 $f(x|\boldsymbol{\phi})$ is identifiable.

C.3 The information matrix $\boldsymbol{\Sigma}^{-1} = \mathbf{X}/n$, defined by

$$\boldsymbol{\Sigma}^{-1} = -E \left[\frac{\partial^2 l(\boldsymbol{\phi})}{\partial \phi_r \partial \phi_s} \right],$$

where $r, s = 1, 2, 3, 4$ is positive definite.

C.4 For all $x \geq 0$, the second partial derivatives $f_{r,s}(x|\boldsymbol{\phi}) = \frac{\partial^2 f(x|\boldsymbol{\phi})}{\partial \phi_r \partial \phi_s}$, where $r, s = 1, 2, 3, 4$ are continuous for all $\boldsymbol{\phi} \in \boldsymbol{\Phi}$.

C.5 $f(x|\boldsymbol{\phi})$ is continuously differentiable for $x > 0$ and all $\boldsymbol{\phi} \in \boldsymbol{\Phi}$.

Verification of C.1 through C.5:

First, C.1 is satisfied as we set $n_i \approx nc_i$ where the c_i 's are determined *a priori* to the experiment, and C.2 has been verified in section B.1.1.

To show that C.4 is satisfied, note that for all $x \geq 0$,

$$\frac{\partial^2 f}{\partial \alpha^2} = \frac{\partial^2 f}{\partial \theta^2} = \frac{\partial^2 f}{\partial \alpha \partial \theta} = 0,$$

$$\frac{\partial^2 f}{\partial \alpha \partial \kappa} = \frac{\eta(\kappa/x)^\eta}{\kappa(1 + (\kappa/x)^\eta)^2},$$

$$\frac{\partial^2 f}{\partial \alpha \partial \eta} = \frac{(\kappa/x)^\eta \ln(\kappa/x)}{(1 + (\kappa/x)^\eta)^2},$$

$$\frac{\partial^2 f}{\partial \kappa^2} = \frac{\eta(\theta - \alpha)(\kappa/x)^\eta [(\eta + 1)(\kappa/x)^\eta - \eta + 1]}{\kappa^2(1 + (\kappa/x)^\eta)^3},$$

$$\frac{\partial^2 f}{\partial \kappa \partial \eta} = \frac{(\theta - \alpha)(\kappa/x)^\eta [\eta \ln(\kappa/x) [(\kappa/x)^\eta - 1] - (\kappa/x)^\eta - 1]}{\kappa(1 + (\kappa/x)^\eta)^3},$$

$$\frac{\partial^2 f}{\partial \kappa \partial \theta} = \frac{-\eta(\kappa/x)^\eta}{\kappa(1 + (\kappa/x)^\eta)^2},$$

$$\frac{\partial^2 f}{\partial \eta^2} = \frac{(\theta - \alpha) \ln^2(\kappa/x) [(\kappa/x)^\eta - 1](\kappa/x)^\eta}{(1 + (\kappa/x)^\eta)^3},$$

and

$$\frac{\partial^2 f}{\partial \eta \partial \theta} = \frac{-\ln(\kappa/x) (\kappa/x)^\eta}{(1 + (\kappa/x)^\eta)^2},$$

which are all continuous for all $x \geq 0$ and $\phi \in \Phi$.

To verify C.5, note that

$$\frac{\partial f}{\partial x} = \frac{\eta(\theta - \alpha)(\kappa/x)^\eta}{x(1 + (\kappa/x)^\eta)^2},$$

$$\frac{\partial f}{\partial \alpha} = 1 - \frac{1}{1 + (\kappa/x)^\eta},$$

$$\frac{\partial f}{\partial \kappa} = -\frac{\eta(\theta - \alpha)(\kappa/x)^\eta}{\kappa(1 + (\kappa/x)^\eta)^2},$$

$$\frac{\partial f}{\partial \eta} = -\frac{(\theta - \alpha)(\kappa/x)^\eta \ln(\kappa/x)}{(1 + (\kappa/x)^\eta)^2},$$

and

$$\frac{\partial f}{\partial \theta} = \frac{1}{1 + (\kappa/x)^\eta},$$

which are all continuous for all $x \geq 0$ and $\boldsymbol{\phi} \in \boldsymbol{\Phi}$.

The verification of C.3 is more complicated; however, Krewski and Ryzin (1981) provide the following lemmas using the Chebyshev set.

Definition:

A set of functions $\{f_r(x); r = 1, \dots, k\}$ is said to form a Chebyshev set of $[0, X]$ if $f(x; \boldsymbol{\phi}) = \sum a_r f_r(x)$ has at most $k - 1$ zeros in $[0, X]$ for all $\boldsymbol{a}' = (a_1, \dots, a_k)$.

Lemma:

The set of functions $\{f_r(x); r = 1, \dots, k\}$ forms a Chebyshev set on $[0, X] \Leftrightarrow$ the matrix $\langle f_r(x_i) \rangle; r, i = 1, \dots, k$ is of full rank for every set of k distinct points $\{x_i\}$ in $[0, X]$.

Lemma:

The information matrix $\boldsymbol{\Sigma}^{-1}$ in C.3 is positive definite provided that $\{\frac{\partial f(x)}{\partial \phi_r}; r = 1, \dots, k\}$ forms a Chebyshev set on $[0, X_I]$ where $I + 1 > k$.

In other words, the information matrix $\boldsymbol{\Sigma}^{-1}$ is positive definite if we can show that the matrix $\langle (f_r(x_i)) \rangle$ for $r, i = 1, 2, 3, 4$ is of full rank for every set of 4 distinct points $\{x_i\}$ in $[0, X_t]$.

Proof:

Let

$$\langle (f_r(x_i)) \rangle = \begin{bmatrix} \frac{\partial f(x_1)}{\partial \alpha} & \dots & \frac{\partial f(x_1)}{\partial \theta} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(x_4)}{\partial \alpha} & \dots & \frac{\partial f(x_4)}{\partial \theta} \end{bmatrix},$$

where $\{x_i\}$ in $[0, X_t]$.

To show that the above matrix is of full rank, we need to prove that $\sum_{i=1}^4 c_i \frac{\partial f(x_i)}{\partial \phi} = \mathbf{0}$

implies that $c_1 = c_2 = c_3 = c_4 = 0$.

Now, suppose that

$$c_1 \frac{\partial f(x_1)}{\partial \alpha} + c_2 \frac{\partial f(x_2)}{\partial \alpha} + c_3 \frac{\partial f(x_3)}{\partial \alpha} = c_4 \frac{\partial f(x_4)}{\partial \alpha},$$

where at least one of the c_i 's is not zero for a given set of $\phi \in \Phi$ and $\{x_i\}$ in $[0, X_t]$.

Since $\frac{\partial f(x)}{\partial \alpha} = \frac{\kappa^\eta}{x^\eta + \kappa^\eta}$, the above relation can be written as

$$\begin{aligned} \sum_{i=1}^3 \frac{c_i \kappa^\eta}{x_i^\eta + \kappa^\eta} &= \frac{c_4 \kappa^\eta}{x_4^\eta + \kappa^\eta} \Rightarrow \\ \sum_{i=1}^3 \frac{c_i}{x_i^\eta + \kappa^\eta} &= \frac{c_4}{x_4^\eta + \kappa^\eta}. \end{aligned} \tag{B.1}$$

If $\langle (f_r(x_i)) \rangle$ is not of full rank, for the above c_1, c_2, c_3 and c_4 , we must have

$$\begin{aligned} \sum_{i=1}^3 \frac{c_i \partial f(x_i)}{\partial \kappa} &= \frac{c_4 \partial f(x_4)}{\partial \kappa} \Rightarrow \\ \sum_{i=1}^3 \frac{c_i \eta (\theta - \alpha) (\kappa/x_i)^\eta}{\kappa (1 + (\kappa/x_i)^\eta)^2} &= \frac{c_4 \eta (\theta - \alpha) (\kappa/x)^\eta}{\kappa (1 + (\kappa/x)^\eta)^2}. \end{aligned}$$

Since $\frac{\eta(\theta-\alpha)}{\kappa}$ is constant, the above relationship can be simplified to

$$\sum_{i=1}^3 \frac{c_i(\kappa/x_i)^\eta}{(1 + (\kappa/x_i)^\eta)^2} = \frac{c_4(\kappa/x_4)^\eta}{(1 + (\kappa/x_4)^\eta)^2} \Rightarrow$$

$$\sum_{i=1}^3 \frac{c_i}{x_i^\eta + \kappa^\eta} \left(\frac{\kappa^\eta}{1 + (\kappa/x_i)^\eta} \right) = \frac{c_4}{x_4^\eta + \kappa^\eta} \left(\frac{\kappa^\eta}{1 + (\kappa/x_4)^\eta} \right). \quad (B.2)$$

Substituting (B.1) in (B.2) yields

$$\sum_{i=1}^3 \frac{c_i}{x_i^\eta + \kappa^\eta} \left(\frac{\kappa^\eta}{1 + (\kappa/x_i)^\eta} \right) = \sum_{i=1}^3 \frac{c_i}{x_i^\eta + \kappa^\eta} \left(\frac{\kappa^\eta}{1 + (\kappa/x_4)^\eta} \right),$$

and $\frac{\kappa^\eta}{1 + (\kappa/x_4)^\eta}$ is constant once x_4 is specified. Hence,

$$\sum_{i=1}^3 \frac{c_i}{x_i^\eta + \kappa^\eta} \left(\frac{\kappa^\eta}{1 + (\kappa/x_i)^\eta} - \frac{\kappa^\eta}{1 + (\kappa/x_4)^\eta} \right) = 0,$$

which implies that at least one of x_1, x_2, x_3 equals x_4 , since at least one of the c_i 's is not zero. However, since this cannot be the case, this implies that $c_1 = c_2 = c_3 = c_4 = 0$, and the proof is complete.

B.2. Number of Concentration Levels Required for the Optimal Design for BMR
Chernoff (1953) gives the following theorem.

Theorem: Chernoff's theorem

Suppose that we have u estimators of interest, which may depend of $\boldsymbol{\phi} = (\phi_1, \dots, \phi_k)'$.

If R is closed and bounded, there is an element \tilde{X} of R which minimizes $v_u(X) = x^{11} + \dots + x^{uu}$ and is a convex linear combination of $t \leq k + \dots + (k + u + 1)$ elements $\tilde{X}_1, \dots, \tilde{X}_t$ of R_1 . Furthermore, $\tilde{X}_1, \dots, \tilde{X}_t$ may be chosen so that $v_u(X)$ is a continuous

function at $X = \tilde{X}$ with respect to the topology of the convex set generated by $\tilde{X}_1, \dots, \tilde{X}_t$, where

1. R_1 is a set of vectors of length $k(k+1)/2$ corresponding to the information matrix $\mathbf{X}_{\phi^{(0)}}$, and
2. R is a convex hull of R_1 , which represents the set of information matrices of the class of mixed experiments.

Here, we are interested in estimating the additional (or extra) activity level π_0 , which depends on the Hill model parameters ϕ . Chernoff's theorem states that if the information matrix $\mathbf{X}_{\hat{\phi}}$ is closed and bounded at any $\{x\}$ in $[0, X]$ under any given set of $\phi \in \Phi$, then there exists a set of concentration levels x that minimizes the variance of the estimator $\hat{\pi}_0$. Furthermore, the theorem tells us that such an optimal design would require four concentration levels.

In particular, we want to show that each element of the information matrix $\mathbf{X}_{\hat{\phi}} = -E\left(\frac{\partial^2 l}{\partial \phi_r \partial \phi_s}\right)$ is closed and bounded.

Proof:

Since $\mathbf{X}_{\hat{\phi}} = \mathbf{A}\mathbf{D}\mathbf{A}^T$, where $\mathbf{A} = \{f_\alpha(x), f_\kappa(x), f_\eta(x), f_\theta(x)\}'$ and $\mathbf{D} = \text{diag}\{n_i/\sigma_i^2\}$, we will show that $f_r(x)$ and D_i are all closed and bounded, so that $\mathbf{A}\mathbf{D}\mathbf{A}^T$ is also closed and bounded.

First, since

$$\sigma_i^2 = \sigma_0^2 f(x_i)^{2\lambda},$$

where $\lambda \geq 0$, the denominator σ_i^2 is thus bounded at the boundary $[0, x_t]$ at $\sigma_0^2 \alpha^{2\lambda}$ and $\sigma_0^2 \left[\alpha + \frac{\theta - \alpha}{(1 + (\kappa/x)^\eta)} \right]^{2\lambda}$, respectively. Furthermore, assuming $\alpha \neq 0$, since $f(x)$ is continuous for all $x \geq 0$ and σ_0^2 is a constant, σ_i^2 is closed and bounded. This then implies that all D_i 's are closed and bounded.

For each of the partial derivatives with respect to Hill model parameters,

$$f_{\alpha}(x) = 1 - \frac{1}{1 + (\kappa/x)^{\eta}}$$

is closed and bounded by $[1 - \frac{1}{1+(\kappa/x)^{\eta}}, 1]$. The partial derivative with respect to κ

$$f_{\kappa}(x) = -\frac{\eta(\theta - \alpha)(\kappa/x)^{\eta}}{\kappa(1 + (\kappa/x)^{\eta})^2}$$

has $\lim_{x \rightarrow 0^+} f_{\kappa}(x) = \lim_{x \rightarrow \infty} f_{\kappa}(x) = 0$ and is continuous for all $x \geq 0$, implying that it is closed and bounded between $[0, x_t]$.

Similarly,

$$f_{\eta}(x) = \frac{(\theta - \alpha)(\kappa/x)^{\eta} \ln(\kappa/x)}{(1 + (\kappa/x)^{\eta})^2}$$

has $\lim_{x \rightarrow 0^+} f_{\eta}(x) = \lim_{x \rightarrow \infty} f_{\eta}(x) = 0$ and is continuous for all $x \geq 0$, implying that it is closed and bounded between $[0, x_t]$.

Finally, $f_{\theta}(x) = 1 - f_{\alpha}(x)$ and hence is also closed and bounded.