Application of an Active Comparator-Based Benefit-Risk Assessment in Evaluating Clinical Trial Design Features of a New Chemical Entity Using a Bayesian Decision-Theoretic Framework

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

VARUN GOEL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

RICHARD C. BRUNDAGE, Ph.D.
ADVISOR

Jun 2010

Acknowledgements

This thesis was not possible without the help, guidance, and support of my adviser Dr. Richard Brundage. His guidance encouraged me to fearlessly pursue novel research methods. A special thanks goes to his editorial input during writing of this thesis. I am indebted to Brian Corrigan and Raymond Miller for letting me use the data from PD 0220390 as part of this thesis and providing suggestions during the course of this research. A special thanks goes to Dr Timothy Hanson for guiding me on the usage of Bayesian methodologies in this research. .

I am thankful to Dr. William Elmquist, Dr. James Cloyd, Dr. Timothy Hanson and Brian Corrigan for serving on my thesis committee. A special thanks goes to Dr. William Elmquist for his suggestions, that greatly helped in improving this thesis. I would like to thank Kyle Baron for the many thought-provoking discussions on research methods used in this thesis.

I would like to thank fellow graduate students, Wonkyung, Jae Eun, Manisha, Vijay, Amit, Hyewon, for their support and stimulating discussion during my studies at UMN. A warm thanks to my friends, Shruti, Abhijit, Sagar, Rahul, Nagdeep and Anurag for their friendship and support.

A special note of gratitude goes out to my best friend and wife, Rima Patel, who had been a pillar of support through my bad and worse times. I am forever indebted to my father, Shri. Vinod Kumar Goel, my mother, Smt. Mithlesh Goel, my loving brother, Vipul, and

sweet sister, Sugandha, for their unconditional love and support that kept me motivated and enabled me to pursue my doctoral studies.

To our parents

Abstract

During the drug development process, drug candidates are screened for their efficacy and toxicity. Dose selection is a crucial part of drug development and specifying the right dose imparts pharmacological activity while minimizing side effects. Evaluation of the benefit/risk ratio is typically done by examining the effect of a drug on efficacy and safety endpoints. However, this comparison can be difficult when there are multiple endpoints that are clinically and commercially relevant. A decision-based clinical utility is proposed and evaluated to aid in dose selection. A dose is viable if it has higher efficacy and lower toxicity than the values specified in multi-attribute decision criteria.

PD 0200390 is a ligand of the $\alpha_2\delta$ subunit of the voltage-gated calcium channel being investigated for the treatment of primary insomnia and non-restorative sleep. Wake after sleep onset and number of awakenings are the measures of sleep consolidation while ease of awakening and morning behavior following wakefulness are the measures of residual effects. The objective of this research is to select a dose that maximizes the probability of a decision criterion characterized over safety and efficacy attributes. Data is obtained from two phase II double blind, randomized, placebo controlled crossover studies in subjects with primary insomnia. Dose response models are developed as hierarchical non-linear model using NONMEM® and WinBUGS®. A Sensitivity analysis is performed to test the robustness of the selected dose with varying decision attributes.

Table of Contents

List of Tables

List of Figures

1. Introduction

The costs of drug development include an average investment of $500-$800 million, 10-12 years of development time and more than 35 trials and 4000 trial participants[1, 2]. Out of the drugs that make it to the market only 12% provide satisfactory monetary return and 70% fail to return their developmental costs[1]. A report from Food and Drug Administration (FDA) stated that there is an 8% chance that a drug entering Phase I will eventually make it to the market; which is almost a 50% reduction compared to a 14% success rate 15 years ago[3, 4]. It is estimated that almost 50% of the drugs that enter Phase III fail, a more than two-fold increase compared to a 20% failure rate 10 years ago[4]. This increased attrition is attributed to empirical methods of drug evaluation and increased scrutiny of the FDA to ascertain safety and efficacy of the new compounds[4].

The FDA called for a Critical Path Initiative to build and use "new tools to get fundamentally better answers about how the safety and effectiveness of a new drug can be demonstrated in faster time frames, with more certainty, and at lower costs"[3]. Among the methods highlighted, the transition to a model-based drug development paradigm was proposed. Model-based drug development is exemplified by the *learn-and-confirm* paradigm originally proposed by Sheiner[5]. He described drug development as two successive learn and confirm cycles[5]. The first learn-confirm cycle aims at finding the tolerated dose and the second learn-confirm cycle evaluates the benefit/risk ratio[5].

The learn-confirm paradigm rests on determining cause-effect relationships between inputs, dose, outcomes, safety and efficacy, and patient specific covariates, e.g., age, gender, weight. An efficient way to understand this relationship is by developing hierarchical pharmacokinetic and pharmacodynamic models (PK/PD). A pharmacokinetic model relates the dose of an administered drug to its exposure at the site of action. Typically, drug concentrations in the plasma are used as surrogates for drug exposure at the site of action. A pharmacodynamic model relates drug exposure to efficacy and safety endpoints. A typical PK/PD hierarchical model consists of a deterministic model that describes the PK/PD relationship and a statistical model that describes the observed variability to covariate effects and classifies the variability as between-subjects and within-subject variability. These hierarchical models are then used for prediction such as looking into what-if scenarios, finding best-dosing regimen, trial designs, and providing probabilistic assessment of efficacy and adverse events.

The hierarchical model can be developed under many statistical approaches including the maximum likelihood approach and the Bayesian approach. The maximum likelihood approach selects the model and the parameters, given which the data are most likely to be observed. Under the Bayesian approach, the prior beliefs about the model and parameters are updated to the posterior beliefs in light of evidence provided by the data. The strength of the Bayesian approach is the ability to formally incorporate prior information in the statistical inference and decision making. The Bayesian approaches are particularly useful when the objective is to learn and accumulate knowledge (induction). On the other

hand the maximum likelihood approaches are useful when the objective is to conclude only from the observed data (deduction).

It is of interest to pharmaceutical companies, patients and the FDA to quickly bring viable drug products to the market. It is equally important to terminate the development of a compound that has suboptimal efficacy and safety. Stopping the development of a suboptimal drug is not only monetarily beneficial but also ethical. Most of the go-no-go decisions about drug development have traditionally been based on the subjective assessments that are non-quantitative and sometimes inconsistent[4]. An objective approach is to use decision-theoretic framework to evaluate and choose among many possible alternatives.

This thesis explores the use of decision-theoretic framework to the problems of optimum dose and sample size selection in designing of a clinical trial. The decision-theoretic framework rests on the axioms of decision analysis and maximum utility theory to ensure consistency and maximum value of an action. Decision analysis is a formal framework for making decisions when there are multiple objectives, risk and uncertainties, decision makers from multiple disciplines and consequences that may have high economic impact. Maximum utility theory is the decision analysis framework for decision making under uncertainty and multi-attribute utility theory is a formal framework for decision making with multiple goals.

In this thesis, the concepts of Bayesian hierarchical modeling and decision theory are applied to the drug development scenario of PD 0200390, a compound under

investigation for insomnia. Data were available from two double-blind, placebo-controlled, crossover studies. Bayesian hierarchical modeling was used to describe the dose-response and covariate relationship for safety and efficacy attributes. The selected model and parameter estimates, from dose-response models, were utilized, under decision-theoretic framework, to dose finding and a future trial design.

The organization of this thesis is as follows: an introduction to insomnia, current management therapies and clinical pharmacokinetics of PD 0200390 is described in Chapter 1.1. A framework for Bayesian approaches to hierarchical modeling is discussed in Chapter 1.2. The framework of decision analysis and its use in designing experiment and clinical trials is introduced in Chapter 1.3. The development of hierarchical Bayesian dose-response models for the multiple safety and efficacy endpoints for PD 0200390 are described in Chapter 2. The use of Bayesian decision-theoretic approach to dose finding and trial design for PD 0200390 is discussed in Chapter 3. Future directions are highlighted in Chapter 4.

1.1. Insomnia and PD 0200390

Sleep

Sleep is an essential part of human life, however its effect on human functioning is little understood[6]. It is suggested that, while wakefulness utilizes body's resources, sleep provides consolidation enabling body to recover, build proteins and process information[7]. Sleep accounts for one third of the human life span, approximately 3-10 h/night. A sleep-wake cycle is established shortly after birth and changes over the individual's life time[6]. The length of sleep required changes with age. According to the National Sleep Foundation newborns should sleep for 10.5-18h, infants 9-12h, toddlers 12-14h, children 10-11h, teens 8.25 - 9.5h and adults and older adults 7-9h[87]. Loss of sleep is reported to increase risk for fatigue related automobile crashes, industrial accidents, marital problems, loss of energy during day time, metabolic and endocrine imbalance, lack of ability to maintain weight control, poor health, coronary heart disease, depression and suicidal tendencies[9-11].

It is suggested that a number of nuclei in the brainstem and hypothalamus form a sleep circuit consisting of arousal and inhibitory centers[7]. In this inhibitory-excitatory loop, excitatory nuclei are active during wakefulness and inhibitory nuclei are active during sleep[7]. Sleep is influenced by several neurotransmitters: norepinephrine is important in maintaining normal sleep pattern, serotonin is important for sleep regulation, acetylcholine is involved in rapid-eye-movement sleep (REM), and dopamine reduces sleep and promotes wakefulness[6]. Neuropeptides like galanin increases rapid-eye-movement sleep and slow-wave sleep and grehlin promotes sleep[7]. Other neurochemicals

like histamine, substance P and corticotropin modulates neuronal activity during wakefulness[6].

Sleep Cycle

Sleep consists of stages which are characterized by changes in brain's electrical activity, muscle tone and eye movements. Electroencephalography (EEG) monitors the electrical activity of the brain, electrooculography (EOG) monitors the movement of the eye and electromyography (EMG) monitors the skeletal muscle tone[6]. Sleep is differentiated into rapid-eye-movement sleep, characterized by rapid eye movements and reduced skeletal muscle tone, and the non-REM (NREM) sleep. The NREM sleep is further differentiated into four stages; stages 1 and 2 are the stages of light sleep and stages 3 and 4 are the stages of deep sleep. The differentiation is based on the frequency and amplitude of the neuro-physiological signals. Stage 1 sleep is the transition between wakefulness and sleep and lasts for 0.5-7 minutes[6]. Stage 2 sleep consists of low voltage EEG with spindle shaped waves and high voltage spikes and is a deeper sleep compared to Stage 1[6]. Stage 3 and 4 are known as delta sleep and consist of high amplitude, slow waves[6]. REM sleep is characterized by low voltage, mixed frequency EEG, low muscle tone, and variable autonomic functioning[6]. In REM sleep the brain seems to be as active as during wakefulness and is the time at which dreaming occurs[7].

During normal sleep, one transitions through various stages of sleep. The NREM sleep stages comprise of 75% of total sleep and are prominent in early hours of sleep[12, 13]. The REM sleep comprises of 25% of sleep time, lasts 5-30 minutes, and increases in duration and frequency in later part of the sleep[12, 13]. The deep sleep stages of the NREM sleep and

REM sleep are thought to be important for memory consolidation. Different stages of sleep are reported to affect different types of memory consolidation. Slow wave or deep sleep stages are thought to be responsible for consolidation of declarative memories such as facts whereas REM sleep is associated with consolidation of emotional and procedural memories[12].

The duration of REM sleep decreases with age, with 50% REM sleep in newborn babies to 20% in adults. The sleep pattern changes with age. In the elderly, there is also a decrease in delta sleep and total sleep time leading to increased number of awakenings and wake time[6]. Wakefulness is characterized as low voltage EEG, rapid eye movement and blinks, and a high muscle tone[6].

Diagnostic Tools

Polysomnography

Polysomnography (PSG) is a tool for the continuous monitoring of multiple physiological parameters during a sleep study. The PSG is usually done in a dedicated sleep lab that provides a comfortable environment with a dark sound proof room and a comfortable bed. Input from EEG, EOG and EMG, oxygen saturation, limb movements, chest and abdominal movements are used to characterize normal or disturbed sleep patterns[13]. Sleep stages are scored for a period of 20-30 seconds, called an epoch of EEG, EOG and EMG , based on the system of rules defined in "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects" by Rechtschaffen and Kales[14]. The data acquired through EEG, EMG and EOG were traditionally recorded on a printer paper, but recently computer-assisted PSG has replaced the old technique[13].

The data analysis can be fully automated using software that is interactive or semi-automatic depending upon the requirement and nature of the clinical research[13]. Sleep data from PSG are summarized as the time periods and the latencies of various sleep stages. A description of commonly used time and latency variables, derived from PSG, along with their definitions is given below. A representation of sleep architecture based on sleep and wakefulness is shown in Figure 1.

Latency to persistent sleep (LPS): Time from the beginning of the PSG recording until the initiation of 10 minutes of continuous sleep;

Total sleep time (TST): Total time spent in stage 1, 2, 3, 4, NREM and REM sleep;

Wake time after sleep onset (WASO): The number of minutes of wake time after the initiation of persistent sleep;

Number of awakenings after sleep onset (NAASO): The number of times the subject awakens for at least 2 epochs after at least 2 epochs of sleep Stage 1, 2, 3, 4 or REM;

Number of shifts to Stage 1: Number of times the subject enters Stage 1 sleep;

Time in each sleep Stage: The time (in minutes) spent in each stage of sleep (1, 2, 3, 4, Non-REM, and REM;

Latency to sleep onset (LSO): Time from beginning of the recording to the first epoch of Stages 2, 3, 4 or REM;

Latency to REM: The number of minutes from initiation of sleep until the occurrence of the first epoch of REM sleep;

Wake time after sleep (WTAS): The amount of time awake after the final awakening until the end of the PSG recording.

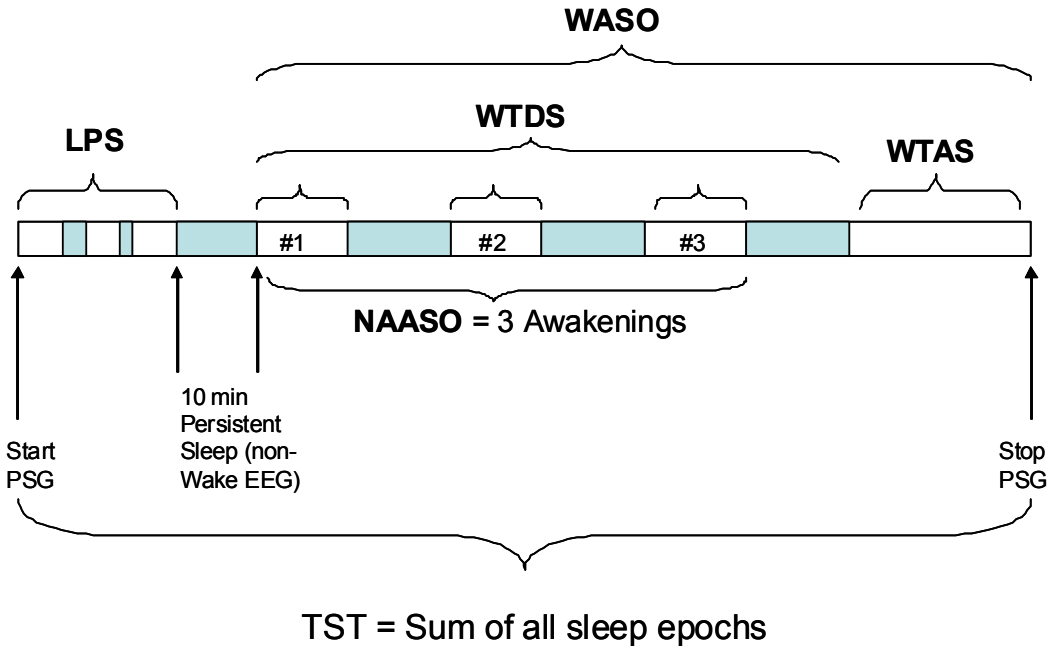

Figure 1. Some of the sleep parameters derived from PSG recordings, where, WASO: Wake After Sleep Onset; LPS: Latency to Persistent Sleep; NAASO: Number of Awakenings After Sleep Onset; TST: Total Sleep Time; WTDS: Wake Time During Sleep; WTAS: Wake Time After Sleep. The shaded areas in the figure represent sleep phases.

Other methods for diagnosing insomnia include sleep logs and actigraphy[15]. A sleep log is a record kept by a patient about the amount and quality of sleep. Typical inputs are bedtime, sleep time, number and duration of awakenings and naps taken during the day. Actigraphy is a technique of recording movement during sleep. An actigraph is a watch

size instrument worn on the wrist and keeps log of the patient's movement. The rest-activity obtained by actigraphy has been reported to be in close agreement with the sleep-wake scoring by PSG[15].

Approximately one third of the population experiences sleep disorders in their life-time[6]. The sleep disorders are broadly classified into, insomnia, excessive day-time sleepiness and abnormal sleep behaviors[6]. The definition, classification and management of insomnia are described below.

Insomnia

Insomnia is defined as sleep that is inadequate or abnormal, despite sufficient opportunity for uninterrupted sleep[16]. Insomnia is not defined by total sleep time, but inability to get sufficient sleep required for feeling refreshed on wakefulness[15]. It is characterized by significant distress or impaired functioning occurring with any of the following:

- Difficulty falling asleep
- Difficulty staying asleep
- Early awakening
- Non-restorative sleep

Classification of Insomnia

According to the American Psychiatric Association's (APA) *Diagnostic and Statistical Manual on Mental Disorders, fourth edition (DSM-IV)*[16], insomnia can be classified as primary or secondary.

The APA requires a diagnostic criterion for primary insomnia as one month with difficulty initiating, maintaining or nonrestorative sleep[16]. The condition must be associated with significant distress and day time impairment that cannot be attributed to any other sleep disorder, psychiatric or medical disorder, or substance abuse[16]. It is characterized by a consistent set of symptoms that follow a general disease course[17]. Among patients diagnosed with insomnia, 25% to 30% are classified with primary insomnia. The occurrence of primary insomnia is attributed to endocrine, behavioral and neurological factors[17]. Secondary insomnia is defined as insomnia related to co-existing physical, psychiatric or medication use. However, there seems to be a limited understanding of the cause-effect relationship between insomnia and other existing conditions[17].

Insomnia is also classified as chronic (long term) and acute (transient) based on the duration of symptoms. Transient insomnia lasts for few days and can result from air travel, noise, odd working hours or other environmental and social events. The duration of chronic insomnia can range from 30 days to six months and is related to an intrinsic sleep disorder, primary insomnia, medical or psychiatric conditions[15].

Epidemiology of Insomnia

Insomnia is the most prevalent sleep disorder and chronically occurs in 10–15% of the United States population, with an additional 25–30% experiencing intermittent sleep problems[17]. In the primary care setting, it is the second most common overall complaint after pain[15]. It is reported that elderly are more affected by insomnia and that women are at higher risk of developing sleep abnormalities compared to men[17]. Individuals, who are

unemployed, divorced, separated, and with low socio-economic status are reported to have a higher incidence of insomnia[6, 15]. Even with the high incidence, insomnia remains largely untreated. Only 5% of insomnia patients seek medical assistance and approximately 10%-20% of insomnia patients use non-prescription drugs or alcohol to alleviate symptoms[6]. Untreated primary insomnia is linked to increased incidence of clinical depression, panic disorder, alcohol abuse, and possibly, coronary artery disease in men[15]. Insomniacs have a decrease in quality of life and have high health care consumption[15]. The cost of insomnia to the US public is estimated to be between $92.5 to $107.5 billion, which includes medical expenses, ramifications of accidents, decreased work efficiency, and decreased work productivity due to employee absence[18].

Management of Insomnia

The management or treatment of insomnia can be classified into non-pharmacological and pharmacological therapies[19]. Non-pharmacological therapy includes cognitive behavioral therapy and treatments such as yoga and acupuncture. The cognitive behavioral therapy is a collective term for intervention that includes sleep-hygiene education, stimulus-control therapy, sleep-restriction therapy, and relaxation training[19]. Sleep-hygiene focuses on environmental and habitual factor that adversely affects sleep, recommendations include a regular sleep schedule, avoiding stimulants like caffeine and nicotine, having a comfortable sleep environment and avoiding loud noise and bright light[19]. Stimulus-control therapy aims at reducing stimuli in the bedroom and reinforces the association of the bedroom with sleep[9]. Sleep-restriction focuses on inducing sleep by temporary sleep deprivation[19]. It is reported that 70% to 80% of people treated with non-

pharmacological methods report a positive outcome with effect lasting for at least six months after discontinuation of treatment[6]. The non-pharmacological therapies such as sleep deprivation, relaxation training are empirically supported to meet the American Psychological Association criteria for psychological treatment of insomnia[6]. A web search conducted on November 15, 2009 on clinicaltrials.gov for the search term "insomnia" resulted in 356 studies that included 69 studies that were indexed as "randomized" trials with "behavior" interventions. A large percentage of trials with behavioral intervention suggest a wide acceptance of behavioral therapies for the treatment of insomnia.

Pharmacological interventions include benzodiazepines, benzodiazepine-receptor agonists, melatonin-receptor agonists, antidepressants, antipsychotics, barbiturates, anti-histamines, melatonin and valerian[19]. The level of wakefulness, arousal, REM and NREM sleep stages are controlled by the interplay between the wakefulness- and sleep-promoting hypothalamic and brainstem nuclei[20]. The pharmacological interventions treat insomnia by interacting with wakefulness-and sleep-promoting neuronal system[20]. Sedative drugs act by increasing the effectiveness of the sleep-promoting GABAergic system[20]. The activation of the sleep promoting GABAergic system is through activation of the ionotropic $GABA_A$ receptor that regulates chloride channels. The $GABA_A$ receptor is modulated by several drug classes like benzodiazepines (e.g., estrazolam, flurazepam, quazepam, temazepam and triazolam), barbiturates, neurosteroids, ethanol and general anesthetics (e.g., isoflurane and propofol)[20]. Other non-benzodiazepines that selectively interact with the α_1 (sedative) subunit of $GABA_A$ are zolpidem, zopiclone, eszopiclone

and zaloplon[19, 20]. First-generation antihistamines like diphenhydramine and doxylamine act by blocking the wakefulness-promoting histaminergic system[20]. Melatonin, a hormone produced by pineal gland, has hypnotic properties and is sold as a nonprescription sleep agent[20]. Melatonin interacts with MT1 and MT2 receptors at different locations in the brain and its action on suprachiasmatic nucleus is involved in initiating and maintaining sleep[20]. Ramelteon is a melatonin-receptor agonist that targets melatonin receptors and promotes sleep[19].

It is reported that a large portion of insomnia patients remain untreated and as much as 25% -30% self medicate with over the counter (OTC) hypnotics and alcohol[15]. Insomnia treatment is dominated by benzodiazepine and related hypnotic compounds that interact with the GABAergic system. These compounds have shown efficacy in inducing and maintaining sleep; however, they present significant risks for dependence, substance abuse and withdrawal symptoms[7]. The side effects include next-day sedation, ataxia, and memory and cognitive impairment[7].

In recent years, the trend for insomnia compounds is shifting from GABAergic system to the endogenous pathways that restore sleep-wake mechanisms[7]. One such approach is enhancing of slow-wave sleep (SWS)[7]. Increased SWS is associated with the feeling of a good night sleep, memory consolidation and enhanced cognitive functions the following day[7]. Compounds that promote SWS, like gabaxadol, ritanserin, ketanserin, act through different pathways than the benzodiazepines and related compounds and have shown to increase sleep quality and day time performance[7].

The α2δ subunit modulators related to gabapentin are another class of compounds reported to increase SWS. PD 0200390 is a α2δ subunit binding compound currently under development for insomnia and is the compound that was studied in this thesis. A brief introduction and the clinical pharmacokinetics of PD 0200390 are described in the following sections.

PD 0200390

PD 0200390 (PD) [(3S,4S)-(1-aminomethyl-3,4-dimethyl-cyclopentyl)-acetic acid] is a first-in-class voltage-sensitive calcium channel (VSCC) α2δ subunit binding compound being developed as a treatment for patients with insomnia at Pfizer, Inc[21]. The rationale for studying PD in the treatment of insomnia is based on preclinical and clinical evidence of a sleep enhancing effect with PD and other VSCC α2δ subunit binding compounds[22, 23]. Studies suggest that binding with the VSCC α2δ subunit causes modulation of the Ca^{2+} channel resulting in the reduction of excitatory neurotransmitters such as noradrenaline, dopamine and 5-hydroxytryptamine[24]. The oral administration of 10 mg/kg PD to male Sprague Dawley rats showed increased slow-wave sleep[22]. Other VSCC α2δ subunit binding compounds are gabapentin (Neurontin®) and pregabalin (Lyrica®) and have shown effectiveness in epilepsy, neuropathic pain, fibromyalgia and generalized anxiety disorder[22]. Gabapentin and pregablin both have exhibited inhibitory effects on stimulation-evoked release of neurotransmitters[24]. As PD is an investigational compound, extensive information about animal and in-vitro pharmacology is not publically available.

Clinical Pharmacokinetics

The pharmacological action following PD administration is due to the activity of the parent compound. PD is not metabolized appreciably in humans, is highly permeable across membranes and is primarily eliminated unchanged in urine[25].

Following oral administration to humans, PD was rapidly absorbed with the time to the maximum plasma concentration, Tmax, ranging from 1.66 to 3.24 h[21, 25]. There is no clinically relevant difference on the area under the plasma concentration-time curve (AUC) of PD when administered with food. The fed/fasting ratio (90% confidence interval) for C_{max} and AUC is 86.3% (82.8%-91.1%) and 91.9% (89.1%-94.7%), respectively[26]. However, T_{max} of PD is prolonged when given with food or at night[26]. The capsule formulation of PD was found to be bioequivalent to the oral solution[26].

In a study done in normal, mild, moderate and severe renal impaired subjects, oral clearance, renal clearance, and AUC,were correlated with renal function[27]. After a 25 mg oral administration of PD, C_{max} was observed in the range of 0.31 to 0.74 μg/mL and Tmax was observed in the range of 1 to 4 h[27]. The apparent volume of distribution/F was estimated to be 48 L and was not affected by changing renal function[27]. The amount of orally administered drug recovered in urine was 92.1% and 95.6% in healthy and mild renally-impaired subjects, respectively[27]. After an oral administration of ^{14}C labeled PD to healthy subjects, almost 91% of radioactivity was recovered in urine. PD elimination is proportional to renal function as measured by creatinine clearance. The clearance/F of PD after oral administration (124-132 mL/min) is close to net glomeruler filtration rate[25]. Oral clearance/F is decreased in elderly subjects, consistent with the age-related decrease

in creatinine clearance, and in renally-impaired subjects[21]. The pharmacokinetics of PD was observed to be linear in a multiple dose escalating study (dose range 50 to 200 mg/day) done in healthy volunteers[28]. PD has a mean terminal half-life of 5.36 hours that increases with decreasing renal function. The therapeutic exposure range for PD is currently under investigation; an exposure related response curve may require dosing based on renal function. The low hepatic metabolism of PD suggests a lower potential for drug-drug interaction with hepatically metabolized compounds.

Summary

Insomnia is a common disorder, and a large patient population remains untreated. Most of the insomnia subjects self medicate with alcohol as the most common hypnotic. Insomnia drug market is largely comprised of benzodiazepines and related compounds, however they posses a significant risk for dependency, drug abuse and next-day residual effects. PD is a first-in-class α2δ subunit binding compound that shows promise in enhancing slow wave sleep. PD is largely excreted unchanged in the urine with mean reported half-life of 5.36h. These pharmacokinetic characteristics result in a low potential of drug-drug interactions and minimal next day residual effects. Therefore, PD 0200390 is currently undergoing clinical studies for the treatment of insomnia.

1.2. Bayesian Inference

The traditional methods of statistical procedures are referred to as frequentist or classical methods. Frequentist methods are widely used in health care evaluation to make statistical inference and in study design. In the frequentist view, probability is an inherent property of the system and a more exact value can be determined by repeated examination. Bayesians on the other hand conceptualize probability as the *degree of belief*[29]. Frequentists are also called as objectivists and Bayesians as subjectivists[29].

The Bayesian framework can be cast in the following way. Let x be a set of data and θ be the parameter of interest. In Bayesian analysis beliefs about the parameter θ prior to observing data x is represented by having a probability distribution on the parameter θ, given as $P(\theta)$. The updated beliefs about θ posterior to observing x, are represented by probability distribution, $P(\theta|x)$. The vertical bar, '|', means 'conditioned upon'. The updating of prior beliefs about θ, $P(\theta)$ to posterior belief about θ, $P(\theta|x)$, is obtained using Bayes theorem,

$$P(\theta|x) = \frac{P(x|\theta) \cdot P(\theta)}{P(x)}, \tag{1.1}$$

where, $P(x|\theta)$ is known as the likelihood function evaluated at the parameter value θ, and $P(x)$ is a normalizing constant obtained by averaging the numerator over the parameter space of θ, i.e., $P(x) = \int_{\Theta} P(x|\theta)\ P(\theta)\ d\theta$. Bayes theorem was published as a posthumous work by Thomas Bayes, a Nonconformist minister from a small English town of Turnbridge Wells, in 1763[30]. Soon after, Bayesian analysis was eagerly adopted by Laplace and other leading probabilists of that time; however, it was soon disregarded because of an inability to handle prior distributions[31]. Since the first half of 20th century

the frequentist's methods have been the primary methods for data analysis[31]. However, Bayesian approaches have regained momentum in the later part of the 20th century owing to advances in computing methods[31].

Bayesians vs. Frequentists

The philosophical controversy between frequentist and Bayesian methods in science is a matter of active debate. Some particularly thoughtful discussions on this issue are presented by David Freedman[29], Bradley Efron[32], Branden Fitelson[33] and Howson and Urbach[34].

The most important distinction is that Bayesian results depend upon previous evidence or opinions, whereas frequentist results are obtained solely from data at hand[35]. Another important distinction between Bayesian and frequentist approaches is the interpretation of results. Suppose a clinical trial is carried out to test if the test drug differs from the standard therapy. The null hypothesis is that the treatment difference does not exist whereas the alternative hypothesis is that the treatment difference exists. The frequentist result, a p-value, is the probability of observing data as extreme as observed in the trial given the null hypothesis is true, $P(Data|H0)$. A Bayesian result is the probability that there is no treatment difference given the data, $P(H0|Data)$[35]. In the frequentist approach, the difference either exists or not and a probability cannot be assigned to the difference.

In frequentist approaches, the parameter values are assumed to be fixed and unknown however, can be precisely estimated by repeated experimentation. In Bayesian analysis, the parameter values are assumed to be random and have a probability distribution[36].

In the frequentist approach, the interval estimate is referred to as 'confidence interval' and in Bayesian setting it is known as 'credible interval'. A 95% confidence interval means that in a long series of repeated experiments, 95% of the computed confidence intervals will have the true value of the parameter. In frequentist approach, the probability that the computed interval contains the parameter is either 0 or 1. In Bayesian setting, a 95% credible interval refers to the probability that the true parameter mean lies in that interval[30].

Bayesian Approaches in Clinical Trials

The use of Bayesian approaches in clinical trials has increased in recent years. Berry argues that an important use of Bayesian approaches is in prediction and decision making[37]. The purpose of a clinical trial is to evaluate health care interventions. The decision to choose an intervention for a future patient, e.g., choosing a treatment over standard of care, should take into account evidence at hand and the outcome of that decision. The Bayesian decision-theoretic approach facilitates decision making based on approaches of the maximization of expected utility. A discussion on decision theory and the principles of maximization of utility in decision making and experimental design is discussed in Chapter 1.3. Berry argues that Bayesian approaches to clinical trials are more ethical than the frequentist approaches[35]. In a frequentist trial design, the data should be observed for the planned sample size in order to make any inference. On the

other hand, Bayesian approaches allows one to look at data as the trial progresses and make decisions about stopping of the trial based on futility or efficacy[35].

Bayesian Modeling in Pharmacokinetics and Pharmacodynamics

Pharmacokinetic/Pharmacodynamic (PK/PD) modeling is now routine in drug development. Population PK/PD modeling refers to developing hierarchical models where the data from the population is treated as a single entity and the subject specific parameters are assumed to have a probability distribution in the population. The estimation procedures for population PK/PD models are not straight forward due to nonlinearity in structure, noisy data and sparse sampling. The following section reviews the use of Bayesian analysis of PK/PD data. A detailed discussion can be obtained from Lunn *et al*.[38], Best *et al*.[39] and Dufful *et al*.[36] Since the denominator in equation (1.1)is a normalizing constant, the posterior distribution of the parameter vector is proportional to the likelihood times the prior beliefs about the parameters given as, equation (1.2),

$$posterior \propto prior \times likelihood. \tag{1.2}$$

The posterior distribution of the parameter can also be viewed as the weighted average of prior beliefs and likelihood. Consider two extreme cases. If the prior beliefs about the parameter are strong, represented by a point mass of probability at a given value, then the posterior distribution of the parameter will represent the prior and likelihood evidence from data will have no effect. If, on the other hand, there are no prior beliefs about the parameter, represented by a non-informative prior, then the posterior distribution of the parameter is proportional to the likelihood evidence obtained from data, and Bayesian

results are similar to the maximum likelihood approaches. In which case, the mode of the posterior distribution of the parameter from Bayesian analysis is same as the maximum likelihood estimate obtained from frequentist analysis.

The population or hierarchical models are specified in stages. In Bayesian approaches, any number of hierarchies can be incorporated in the model. A typical PK/PD hierarchical model is specified in three stages. Stage 1 specifies a distribution model for the data, stage 2 is the model for inter-individual variability and stage 3 is the specification of the priors. It is possible to add more levels, e.g., to specify inter-occasion variability when data is observed within a subject at multiple occasions or to specify inter-study variability when data are pooled from several studies. The following section describes a three-level hierarchical model for a typical pharmacokinetic model.

Stage 1

Let y_{ij} denote the observation at the j^{th} time, (where $j = 1, \ldots, J$) in the i^{th} individual, (where $i = 1, \ldots, I$), at the and let t_{ij} denote the time of observation in i^{th} subject at the j^{th} time point. In the data presented above, a balanced design is assumed, whereby time points are same for all subjects in the population. The distribution of data can be specified as,

$$y_{ij} \sim N(f(\boldsymbol{\theta}_i, t_{ij}), \sigma^2), \tag{1.3}$$

where, y_{ij} is assumed to have a normal distribution, with a mean function, $f(\boldsymbol{\theta}_i, t_{ij})$, and variance σ^2. The mean function, $f(\boldsymbol{\theta}_i, t_{ij})$, also referred to as the structural model, is a

function of the subject specific vector of parameters, $\boldsymbol{\theta}_i$, and time, t_{ij}. In pharmacokinetic models, data are typically specified as normal or log-normal distributions. In PD models, the data are diverse and the probability distribution can be Poisson, Bernoulli, exponential or some other distribution depending upon the type of observations measured. In a PK model, an example of the structural model, $f(\boldsymbol{\theta}_i, t_{ij})$, is a one-compartmental model, whereas in a PD model, an Emax or linear model may be most appropriate.

Stage 2

The second stage assumes that the individual parameters for the structural model arise from a probability distribution. Since, these parameters are not directly observed they are also referred to as latent (hidden) variables. Most often in PK/PD modeling, random effects distributions are specified as normal or log-normal distributions,

$$\boldsymbol{\theta}_i \sim MVN_p(\boldsymbol{TV\theta}, \Sigma), \tag{1.4}$$

where $MVN_p(\cdot, \cdot)$ is a p-dimensional multivariate normal distribution with the population mean vector, or the typical value of parameter given as $\boldsymbol{TV\theta}$, and variance-covariance matrix Σ. For example, in a one-compartment pharmacokinetic model the random effects on clearance and volume can be specified as a two-dimensional multivariate normal distribution. The population mean vector, $\boldsymbol{TV\theta}$, can also be specified as a function of covariates,

$$\boldsymbol{TV\theta} = g(\boldsymbol{\mu}, \boldsymbol{z}_i), \tag{1.5}$$

where, $g(\cdot,\cdot)$ is a function of the population fixed effects vector, $\boldsymbol{\mu}$, and the individual covariate vector $\boldsymbol{z}_i$, e.g., age, weight and gender.

Stage 3

At the third level of the hierarchical Bayesian model, the priors for the parameters in the models are specified. The priors are the information at hand about the parameters in the model. In case historical data are not available, or when one wishes to be objective while only using the evidence from the data, non-informative priors are specified. Lunn *et al*. suggest non-informative prior distributions for the parameters in the analysis of PK/PD data[38]. The choice of prior distribution has been traditionally chosen for mathematical convenience so that the posterior distributions can be derived analytically[38]. The prior distribution of the precision parameters ($1/\sigma^2$) is specified as $G(a,b)$ where G is a gamma distribution with parameters a and b being the shape parameters. Those parameter can be arbitrarily small to specify non-informativeness; i.e., $a = b = 0.001$. The prior for the inverse of the variance-covariance matrix (Σ^{-1}) is specified as $W(\rho R, \rho)$, where W is a Wishart distribution, R is the scale matrix and ρ is the degrees of freedom. The Wishart distribution ensures that the samples drawn from it are positive-definite matrices. The choice of R is chosen to be centered on the prior precision and the parameter ρ is set equal to the dimension of R to get a vague prior. For the fixed effects parameters in β and the means of the random effects, $\boldsymbol{TV\theta}$, the priors are specified as normal distributions $N(q, H)$, where q and H are the mean and variance of a normal distribution. Non-informative priors are specified by imposing a large variance, e.g., $H = 1000$

centered at mean, $q = 0$. The choice of parameters to obtain informative prior distributions is discussed by Dufful *et al*[36].

Model Fitting

In a Bayesian paradigm the inferences are drawn from the posterior distribution of the parameters. Let $\boldsymbol{\theta} = \{\boldsymbol{TV\theta}. \boldsymbol{\theta}_1 \ldots . \boldsymbol{\theta}_I . \boldsymbol{\mu}. \Sigma. \sigma^2\}$ be a vector consisting of random effects and population model parameters for the hierarchical pharmacokinetic model specified in the previous section and let $\mathbf{y} = (y_{11}, \ldots , y_{IJ})$ be a response vector for I individuals at J time points. The posterior distribution of the parameter vector $\boldsymbol{\theta}$ is obtained by Bayes rule using priors and the likelihood evidence provided by the data[40],

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(\mathbf{y})}, \tag{1.6}$$

where, $p(\boldsymbol{\theta}|\mathbf{y})$ is the posterior distribution of parameters, $f(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood function and $\pi(\boldsymbol{\theta})$ is the prior distribution of the parameter vector $\boldsymbol{\theta}$. The denominator $m(\mathbf{y})$ is the marginal distribution of the observed data $\mathbf{y}$. It is a proportionality constant obtained by integrating the numerator over all possible values of the parameter in the parameter space Θ,

$$m(\mathbf{y}) = \int_{\Theta} f(\mathbf{y}|\boldsymbol{\theta})\ \pi(\boldsymbol{\theta})\ d\boldsymbol{\theta}. \tag{1.7}$$

For nonlinear PK/PD models an analytical solution to this integral is not possible. Markov chain Monte Carlo (MCMC) integration methods are used to obtain the joint posterior distribution of the parameters. The integration method works by sequentially

sampling values from a Markov chain that eventually reaches a stationary distribution that is also the posterior distribution of the parameters[41]. The MCMC integration methods that are available include Gibbs Sampling[42, 43] and Metropolis-Hastings algorithm[44, 45]. Use of MCMC methods in PK/PD is reviewed by Lunn *et al*.[38], Best *et al*.[39] and Dufful *et al*.[36]. These MCMC methods are readily available in Bayesian softwares, e.g., WinBUGS[46] and BRugs[47].

Convergence Diagnostics

Convergence diagnostics are used to check if the Markov chain has converged to a stationary distribution. A useful graphical diagnostic is the "trace" plot, where the line connecting successive samples are plotted against iteration number[38]. A visual inspection of this plot is then used to assess if the chain has converged to a stationary distribution. Sometimes, the appearance of a stationary distribution is possible when the samples are monitored for a small number of iteration, or when the chain converges extremely slowly. To be certain that the samples are indeed from the stationary distribution, multiple chains are initiated from different starting values and monitored in the trace plot. The chains should mix well if they have reached a common stationary distribution, represented by the appearance of a "fat hairy caterpillar". The appearance of a "wiggly snake" represents autocorrelation, in which case the number of samples should be increased[38]. Increasing the samples and re-plotting them in a trace plot will eventually look like the required appearance of a "fat hairy caterpillar"[38]. This visual method also provides informal assessment on the number of sufficient samples and "burn in". The burn-in samples are the samples from the posterior distribution obtained before the MCMC chains converges

to the stationary distribution. These burn-in samples are discarded and not used for making inference. The samples collected after the burn in are used for making inferences about the parameters. Another diagnostic measure of convergence is the Brooks-Gelman-Rubin (bgr) statistic. The bgr statistic assesses the convergence using within and between chain variability[42, 48]. The bgr statistic is also available in WinBUGS and the BRugs package in R. The bgr plot presents three lines in red, green and blue as shown in figure below (Figure 2).

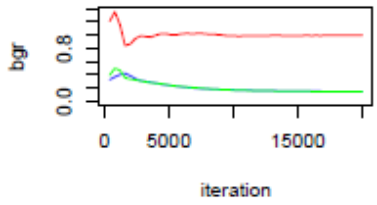

Figure 2 Figure showing a typical plot for Brooks-Gelman-Rubin (bgr) statistic

The red line is the ratio of the between-chain credible interval to the within-chain credible interval. This ratio should reach a value of 1 at convergence. The credible interval for the MCMC samples pooled from all chains is shown as a green line and the average credible interval across chains is shown in blue. At convergence, both the green and blue lines should become stable[46, 48]. Normally, these chains are run for 10,000-40,000 iterations with the initial samples being discarded as burn in. The MCMC samples from the chains (after burn in) are pooled to obtain inference on the posterior distribution of the parameters.

Model Selection

Often, Bayesian model selection is based on computation of the Bayes factors (BF)[49]. In the case of two competing models, the BF for model M_1 against model M_2 is the ratio of posterior odds of M_1 vs. M_2 to the ratio of the prior odds of M_1 vs. M_2[41],

$$BF = \frac{P(M_1|y)/P(M_2|y)}{P(M_1)/P(M_2)}. \tag{1.8}$$

Applying Bayes theorem, the above equation (1.8) can also be written as

$$BF = \frac{\frac{P(y|M_1)P(M_1)}{P(y)} / \frac{P(y|M_2)P(M_2)}{P(y)}}{P(M_1)/P(M_2)} = \frac{P(y|M_1)}{P(y|M_2)}. \tag{1.9}$$

The marginal distribution of data y under model, $P(y|M_i)$, where M_i, $i = 1.2$, is obtained by integrating over the parameter space under model, M_i,

$$P(y|M_i) = \int P(y|\theta_i,\ M_i)P(\theta_i|M_i)d\theta_i, \tag{1.10}$$

Consider two competing models that are a priori equally probable, i.e. $P(M_1|y) = 1 - P(M_2|y)$ and $P(M_1)/P(M_2) = 1$, then the BF is the ratio of posterior probability of models given data y,

$$BF = \frac{P(M_1|y)}{P(M_2|y)}. \tag{1.11}$$

In a restrictive case, where one model is a subset of another, the BF is reduced to likelihood ratio test statistic. A criticism of Bayes factors is that when improper non-informative priors are used, the resulting BF is also improper[50]. To address this issue, several formulations of Bayes factors are suggested including pseudo Bayes factors[51], posterior Bayes factors[52], fractional Bayes factors[53] and intrinsic Bayes factors[54]. Dey *et al*. proposed a pseudo Bayes factors (PsBF) approach for model comparisons in non-linear random effects models when non-informative priors are used. The PsBF approach compares models based on cross-validated predictive densities[51]. The approach computes

the predictive density of an observation deleted from the analysis dataset. The posterior-predictive density of the deleted observation is known as the conditional predictive ordinate (CPO).

Conditional Predictive Ordinate using Monte Carlo Estimates

Consider an analysis dataset from a pharmacodynamic study where J dose levels are available from I subjects. In the notation described below, the subscripted bracket, $_{()}$, means 'does not include' the variables specified in the bracket. Let $\mathbf{y}$ be a response vector of all data and $\mathbf{y}_{(ij)}$ be a vector of response observations without the observation at the j^{th} dose level in the i^{th} subject. The conditional predictive density for the j^{th} observation in the i^{th} individual, y_{ij}, deleted from the analysis dataset, $y_{(ij)}$, is given as

$$f(y_{ij}|\mathbf{y}_{(ij)}) = \int_{\Theta} f(y_{ij}|\boldsymbol{\theta})\ \pi(\boldsymbol{\theta}|\mathbf{y}_{(ij)})\ d\boldsymbol{\theta}. \tag{1.12}$$

where, $\pi(\boldsymbol{\theta}|\mathbf{y}_{(ij)})$ is the posterior distribution of the parameters obtained when the model is fitted to the data $\mathbf{y}_{(ij)}$ under the prior $\pi(\boldsymbol{\theta})$,

$$\pi(\boldsymbol{\theta}|\mathbf{y}_{(ij)}) = \int_{\Theta} f(\mathbf{y}_{(ij)}|\boldsymbol{\theta})\ \pi(\boldsymbol{\theta})\ d\boldsymbol{\theta}. \tag{1.13}$$

The log of the product of CPO for all observations in the dataset is termed as the log pseudo marginal likelihood (LPML) given as

$$LPML = \sum_{i=1}^{I}\sum_{j=1}^{J} \log f(y_{ij}|\mathbf{y}_{(ij)}). \tag{1.14}$$

The log pseudo Bayes factor for the competing models $M_1 \text{vs.} M_2$ is given as,

$$\log PsBF = LPML_{M1} - LPML_{M2}. \tag{1.15}$$

The model selection paradigm shifts from selecting the model that best fits the data (BF approach) to selecting the model that best predicts a future dataset that originates from the same process as the original data (PsBF approach). Jeffery's scale of evidence with modifications suggested by Raftery was used in comparison between models as shown in Table 1[55].

Table 1 Values of PsBF and 2logPsBF suggested for model comparisons

PsBF	2log PsBF	Evidence for M1
<1	<0	Negative
1 to 3	0 to 2	Weak
3 to 12	2 to 5	Positive
12 to 150	5 to 10	Strong
>150	>10	Very strong

The computation of CPO requires prediction of the density of the observation that is not utilized in the fitting process. This may require fitting the model repeatedly so that each observation is predicted by the model that is not included during model development. Gelfand *et al*. provided a MCMC integration procedure which bypasses the need to repeatedly fit the data[42]. The procedure uses the posterior distribution of the data as the importance sampling density for the MCMC integration of the equation (1.12). The derivation below illustrates the computation of the CPO statistic, $f(y_{ij}|\mathbf{y}_{(ij)})$, in equation (1.12) by MCMC integration[56, 57].

Let $h(\boldsymbol{\theta})$ be an importance sampling density for $f(\mathbf{y}_{(ij)}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$, and $\{\boldsymbol{\theta}^{(iter)}\}$, $iter = \{i, \ldots, n\}$ be the MCMC samples from $h(\boldsymbol{\theta})$. The weight, $w^{(iter)}$, is defined as $w^{(iter)} = f(\mathbf{y}_{(ij)}|\boldsymbol{\theta}^{(iter)})\pi(\boldsymbol{\theta}^{(iter)})/h(\boldsymbol{\theta}^{(iter)})$ and the Monte Carlo integration for (1.12) is

$$f(y_{ij}|\mathbf{y}_{(ij)}) = \frac{\sum_{iter=1}^{n}\left[f(y_{ij}|\boldsymbol{\theta}^{(iter)})w^{(iter)}\right]}{\sum_{iter=1}^{n} w^{(iter)}}. \tag{1.16}$$

If the importance sampling density, $h(\boldsymbol{\theta})$, is approximated by the posterior density, $\pi(\boldsymbol{\theta}|\mathbf{y})$, then the resulting weight, $w^{(iter)}$, for the MCMC integration is defined as

$$w^{(iter)} = \frac{f(\mathbf{y}_{(ij)}|\boldsymbol{\theta}^{(iter)})\pi(\boldsymbol{\theta}^{(iter)})}{f(\mathbf{y}|\boldsymbol{\theta}^{(iter)})\pi(\boldsymbol{\theta}^{(iter)})/f(\mathbf{y})} \quad = \quad \frac{f(\mathbf{y})f(\mathbf{y}_{(ij)}|\boldsymbol{\theta}^{(iter)})}{f(\mathbf{y}|\boldsymbol{\theta}^{(iter)})},$$

Since, $f(y_{ij}|\boldsymbol{\theta}^{(iter)}) = f(\mathbf{y}|\boldsymbol{\theta}^{(iter)})/f(\mathbf{y}_{(ij)}|\boldsymbol{\theta}^{(iter)})$ this simplifies to

$$w^{(iter)} = \left[\frac{f(y_{ij}|\boldsymbol{\theta}^{(iter)})}{f(\mathbf{y})}\right]^{-1}. \tag{1.17}$$

Using $w^{(iter)}$, defined in (1.17), in equation (1.16), the Monte Carlo estimate of CPO (equation (1.12)) is

$$f(y_{ij}|\mathbf{y}_{(ij)}) = \frac{\sum_{iter=1}^{n}\left[f(y_{ij}|\boldsymbol{\theta}^{(iter)}) \cdot \left[\frac{f(y_{ij}|\boldsymbol{\theta}^{(iter)})}{f(\mathbf{y})}\right]^{-1}\right]}{\sum_{iter=1}^{n}\left[\frac{f(y_{ij}|\boldsymbol{\theta}^{(iter)})}{f(\mathbf{y})}\right]^{-1}}, \tag{1.18}$$

which simplifies to

$$= n\left[\sum_{i=1}^{n} f(y_{ij}|\boldsymbol{\theta}^{(iter)})^{-1}\right]^{-1}. \tag{1.19}$$

The expression in equation (1.19) is the harmonic mean of the conditional densities at the j^{th} dose level in the i^{th} subject evaluated at the posterior sample values[56]. This can be easily evaluated from the posterior samples of the parameter vector $\boldsymbol{\theta}^{(iter)}$ obtained during the MCMC integration. Geisser and Eddy first proposed the use of the PsBF for model comparison[51]. Given the conditional predictive distribution, Dey *et al.* suggests using CPO plots and deviance plots for comparing between models[50].

AIC, BIC and DIC

Two of the common criteria used in model comparison are Akaike information criteria[58] (AIC) and Bayes information criteria[59] (BIC). The Akaike information criteria is given as $\mathrm{AIC} = -2(\log \text{ maximum likelihood}) + 2(\text{number of parameters})$; and Bayes information criteria is given as

$$\mathrm{BIC} = -2(\log \text{ maximum likelihood}) + (\log n)(\text{number of parameters}).$$

For both model selection criteria, a lower value is preferred for model selection. As can be seen, the two criteria only differ in the coefficient that is multiplied by the number of parameters. BIC is more conservative in selection of larger models compared to the AIC. The AIC and BIC methods require knowledge of the number of parameters for model comparison. Determination of the effective number of parameters in a hierarchical model is not straight forward. The number of parameters in a one-compartment pharmacokinetic model with first order oral absorption, fitted to data from 100 subjects with inter-individual variability on CL, V and Ka, specified as a 3×3 covariance matrix, and with an additive residual variability, the total number of parameters is three hundred and ten[36].

However, not all of these parameters contribute equally to the likelihood function, therefore a deviance information criteria has been proposed for hierarchical models[60].

The deviance information criterion, (DIC), is a generalization of the Akaike information criteria and is used to compare competing models[60]. Similar to the AIC, the DIC is comprised of a component representing model fit $\bar{D}$ and a penalty function, p_D for model complexity,

$$DIC = \bar{D} + p_D, \tag{1.20}$$

where $\bar{D} = E_{\theta|y}[D]$ and is the mean of the posterior distribution of deviance statistic. The deviance statistic is given as

$$D(\boldsymbol{\theta}) = -2\log f(\mathbf{y}|\boldsymbol{\theta}) + 2\log h(\mathbf{y}), \tag{1.21}$$

where $f(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood function evaluated at the parameter $\boldsymbol{\theta}$ and $h(\mathbf{y})$ is some standardizing constant of the data that does not impact model selection[41]. The penalty function, p_D, represents the effective number of parameters that is often less than the number of parameters because of borrowing of strengths between the individual specific latent variables in the hierarchical models[41]. The variable p_D can be reasonably defined as the mean of the posterior deviance, $\bar{D}$, minus the deviance evaluated at the posterior means of the parameter, $D(E_{(\theta|y)}[\boldsymbol{\theta}])$,

$$p_D = \bar{D} - D(\bar{\boldsymbol{\theta}}). \tag{1.22}$$

A smaller value of DIC indicates a better fit. Similar to the AIC and the BIC, the DIC values are not a measure of model validity and are only used to compare between two

models. By convention, a difference of 3 to 5 is considered as the smallest meaningful difference for model selection[41].

Posterior Predictive Checks

A posterior predictive check (PPC) evaluates the prediction ability of the model. The PPC compares the posterior predictive discrepancy statistic $D(\mathbf{y}^*)$ to the observed data discrepancy statistic $D(\mathbf{y})$[41]. The term "posterior predictive" refers to the predictive distribution of the data from the posterior distribution of parameters.

Let $\mathbf{y}^*$ be the simulated prediction value from the final model, $\mathbf{y}$ be the observed data and $p(\boldsymbol{\theta}|\mathbf{y})$ is the posterior distribution of the parameter vector $\boldsymbol{\theta}$, conditioned on the observed data vector $\mathbf{y}$. The posterior-predictive distribution of the simulated data vector $\mathbf{y}^*$, conditioned on the observed data vector $\mathbf{y}$, is obtained by integrating out the parameter $\boldsymbol{\theta}$ over the parameter space Θ as shown in equation (1.23),

$$P(\mathbf{y}*|\ \mathbf{y}) = \int_{\Theta} p(\mathbf{y}*|\boldsymbol{\theta})\ p(\boldsymbol{\theta}|\mathbf{y})\ d\boldsymbol{\theta}. \tag{1.23}$$

Computation of the discrepancy statistic for the simulated data $\mathbf{y}^*$, $D(\mathbf{y}^*)$, requires simulating a vector of observations $\mathbf{y}^*$ from the posterior distribution of parameter $p(\boldsymbol{\theta}|\mathbf{y})$, simulating $p(\mathbf{y}^*|\boldsymbol{\theta})$, and then computing $D(\mathbf{y}^*)$. Yano *et al.* compared several discrepancy measures for evaluation of pharmacokinetic models[61]. A choice of the discrepancy statistic would relate to some aspect of the model in question, e.g., the maximum or minimum concentration for PK measurements[36], or number of responders for a pharmacodynamic model.

Summary

In a Bayesian framework, the beliefs about parameters prior to observing data are updated to posterior beliefs using Bayes rule. Bayesian approaches are becoming increasingly popular in health care evaluation and PK/PD modeling. The primary advantage of using Bayesian is that it provides a formal framework for incorporating prior evidence into data analysis. The use of Bayesian methods has been argued to be more ethical in clinical research because it allows changes to the conduct of the trial as the trial progresses. The use of Bayesian approaches in modeling pharmacodynamic data for an insomnia drug, PD, is discussed in Chapter 2, but prior to that, Chapter 1.3 describes a framework for a Bayesian decision-theoretic approach that is subsequently utilized in Chapter 3 for dose finding and clinical trial design for PD.

1.3. Decision theory

The science of decision making has been classified into three research categories, the normative model, the descriptive model and the prescriptive model for decision making[62]. The normative model explores how decisions should be made, a descriptive model explores how decisions are made and prescriptive model explores ways to improve decision making.

Decision theory is the science of normative decision making. The expected utility theory is the normative model for decision making under uncertainty and the multi-attribute utility theory is the normative model for decision making with multiple goals[62]. The use of decision-theoretic approaches in clinical setting is reviewed by Chapman[62], in statistical decision problems by Berger[63], in econometrics by Chamberlain[64] and in pharmacokinetic/pharmacodynamic modeling by Sheiner[65].

The application of decision theory to specific decision making problems is called decision analysis[62]. Keeney gives an intuitive definition of decision analysis as "a formalization of common sense for decision problems which are too complex for informal use of common sense" and a technical definition that decision analysis is "a philosophy, articulated by a set of logical axioms, and a methodology and collection of systematic procedures, based upon those axioms, for responsibly analyzing the complexities inherent in decision problems"[66].

Axioms of Decision-Analysis

Central to decision analysis is a set of logical axioms originally proposed by von Neumann and Morgenstern[67]. The axiomatic foundations suggest that attractiveness among decisions should depend upon: 1) the likelihood of consequences among decisions and 2) the preference of the decision maker to those consequences[66]. Kenney suggests that all decision problems require subjective assessments and the axioms assert that the likelihood of the consequences and the preferences among the consequences should be treated separately[66]. The likelihood of the consequences is estimated by the probabilities and the preferences among consequences are estimated by the utilities. The utility of a consequence is expressed as a numerical value. A consequence with higher preference is assigned a higher numerical value. The probabilities and utilities are combined, using the axioms of decision analysis, to obtain expected utility of alternatives. According to the axioms of decision analysis, the alternative with highest expected utility should be preferred[66]. A decision problem has the following features: 1) a need to obtain a certain objective, 2) the existence of multiple alternatives, 3) the consequences associated with each alternative, 4) uncertainty associated with the consequences of each alternative and 5) preference among consequences[66]. Decision analysis is structured into following steps: 1) structure of the decision problem, 2) consequence of each alternative, 3) preference among consequences, 4) comparison and evaluation of alternatives.

The axioms for decision analysis presented here are adapted from texts from Keeney[66] and Robert[68]. Formal statements of these axioms are discussed by Savage[69], Pratt *et al.*[70], Fisburn[71], Anscombe and Aumann[72] and Robert[68].

Axiom 1a. There should be at least two alternatives that can be specified.

Axioms1b. The resulting consequence of each alternative can be identified.

Axiom2. The likelihood or probability of each consequence that could result from each alternative can be identified.

Axiom3. The decision maker should be able to assign preferences for all possible consequences. Let C be a set of consequences that is assumed to be completely known. The preference relation 'is preferred to' is symbolized by '$\prec$' and 'is equivalent to' is symbolized as '$\sim$'. For any two consequences, $c1$and $c2$, the decision maker must be able to specify either one of the following: $c1$ is preferred to $c2$ ($c1 \prec c2$), $c2$ is preferred to $c1$ ($c1 \succ c2$) or $c1$ and $c2$ are equally preferred ($c1 \sim c2$).

Axiom4a: If two alternatives result in the same two consequences, the alternative that yields the higher probability of the preferred consequence should be preferred.

Axiom4b: (Transitivity) The axiom asserts that if $c1$ is preferred to $c2$ and $c2$ is preferred to $c3$ then $c1$ is preferred to $c3$: if $c1 \prec c2 \prec c3$ then $c1 \prec c3$.

Axiom4c: (Archimidean) If $c1$ is preferred to $c2$ then for any $c3$ in C the following holds: $p \cdot c1 + (1-p) \cdot c3 \prec p \cdot c2 + (1-p) \cdot c3$, where p is a probability. This axiom implies that there should be conservation of ordering when one of the consequences is replaced by a set of consequences with associated probabilities that is indifferent to the consequences being replaced.

Axiom 1a and 1b relates to structuring of the decision problem, Axiom 2 assesses the impact of each alternative, Axiom 3 determines preferences among consequences and Axiom 4 is necessary for comparison between alternatives[66].

In relation to drug development, one can relate the process of pharmacokinetic/pharmacodynamic modeling as a method of assigning probabilities of consequences (safety and efficacy outcomes) to alternatives (dose, covariates). The process of assigning utilities is to assign numerical values for preference among consequences. For example, a safe and efficacious outcome will have a higher utility compared to efficacious outcome with adverse events.

Maximum Utility Theory

The theory of normative decision making under uncertainty is presented in texts from Savage[69], Wald[73], Von Neumann[74] with further work from Raiffa and Schleifer[75] and Degroot[76]. The foundation for Bayesian decision theory is the principle of maximization of expected utility. Consider that a decision maker, DM, has to take an action/decision, d, from available set of decisions, D. The consequences of DM's decision d will depend upon the unknown value of parameter, θ, that belongs to a set, Θ, of all possible parameter values. The unknown parameter θ, also referred to as *unknown states of nature/world*, relates an action/decision, to the consequences. For a decision, $d \in D$, (where "$\in$" is read as 'belongs to') and a state of nature $\theta \in \Theta$, a consequence $c(\theta, d)$ is observed. If the uncertainty in the value of θ is represented by a probability distribution,

$\pi(\theta)$, then the result of a decision, d, will have a probability distribution of consequence $c(\Theta, d)$ on the set C of all possible consequences.

The decision maker will have preferences over the consequences in C. Let, u, denote a real-valued function that assigns a numeric value, representing preference of the decision maker, to each consequence in C. According to decision theory, if the decision maker satisfies the axioms of decision analysis, he/she will choose an action/decision that maximizes the expected value of the function, u. The function, u, which satisfy this property is known as a utility function and the value assigned to each consequence is called its utility. The expected utility of an action d, $U(d)$, is the weighted average of utilities of possible consequences of an action weighted by the probability of each state of nature $\pi(\theta)$[77],

$$U(d) = \int_{\Theta} u(d, \theta)\ \pi(\theta)\ d\theta. \tag{1.24}$$

The optimal decision, or Bayes action, is to choose an action, d^{OPT}, that maximizes the expected utility as given in equation (1.25). In the equation (1.25), the term $argmax$ stands for "argument of the maximum" and returns the value of the argument, d, for which the expression, $U(d)$, is maximum,

$$d^{OPT} = argmax \quad U(d). \tag{1.25}$$

In application to drug development, the selection of a therapeutic dose is a decision problem. For example, administration of a dose of a drug to an individual can be

considered as an action/decision that results in safety and/or efficacy event. The consequences depend upon the selection of dose and the *unknown states of nature*, or more appropriately, the unknown PK/PD model and its parameters in an individual. A real-valued utility function can be defined that puts preference on the consequence of safe and efficacious outcomes. The normative approach to the dose selection would be to choose a dose that maximizes the utility of safe and efficacious outcome in the individual. An efficient way to reduce the uncertainty in the *states of nature* is by developing pharmacokinetic/pharmacodynamic models. Hence, an efficient dose selection is achieved by taking into account patient characteristics (e.g., age, weight, gender, renal and hepatic function) and prior information generated by modeling.

Decision Theoretic Experimental Design

The concept of Bayesian decision theory also applies to designing experiments. The objective of the decision maker is to choose an experimental design that maximizes the goal or purpose of the experiment. In terms of drug development, prior information is often available from historic controls and/or proof of concept trials. There are budget constraints and specific goals of the trial, e.g., finding ED50 in a Phase II trial, or demonstrating efficacy in a confirmatory Phase III trial. The Bayesian decision-theoretic approach provides a coherent framework to optimize trial designs by incorporating prior learning and cost constraints in selecting a trial that maximizes the goal of the experiment[78].

A decision-theoretic experimental design is a two part decision problem. A terminal decision based on the goal of the trial and an initial decision to select the best experimental design.

As pointed out by Lindley[79], Chaloner and Verdinelli[80], and Clyde[78], the decision maker has to choose an experiment design, e, from a set of possible experimental designs E that will lead to observation of data, y, from sample space Y. After observing the data, y, a decision d will be chosen from a set of decision space D. The utility of making a decision, d, after observing y following an experiment e is given as $u(d, \theta, e, y)$. The experimental design problem consists of two decisions sequenced in time[79].

Choose a trial/experiment design e from set of possible trial designs E

Conclude a decision d from set of possible decision space, D.

The two part decision problem is solved in reverse time order. After observing data, y, from a particular trial design, e, a terminal decision, d^{OPT}, will be the one that maximizes the posterior expected utility function, $u(d, \theta, e, y)$,

$$U^{opt}(e, y) = \max_{d \in D} \int_{\Theta} u(d, \theta, e, y)\ p(\theta|y, e) p(y|e)\ d\theta. \tag{1.26}$$

In equation (1.26), $U^{opt}(e, y)$ is the expected utility of the terminal decision, d^{OPT}, after observing data from an experiment, e, and is obtained by averaging the utility function $u(d, \theta, e, y)$ over the posterior distribution of parameters $p(\theta|y)$. The posterior distribution of parameter $p(\theta|y, e)$ is obtained after observing data y from a future

experiment, e, using the Bayes rule. Since the experiment has not been performed yet, the data are not fixed and the evaluation of the utility function should consider uncertainty in observing data, y, given an experiment design, e. The pre-posterior utility is obtained by integrating (1.26) over the sample space, Y. The sample space, Y, consists of all possible values of Y,

$$U^{opt}(e) = \int_{Y} U^{opt}(e, y)\ p(y|e)dy\ = \int_{Y}\int_{\Theta} U^{opt}(e, y)p_e(y|\theta, e)p(\theta)d\theta dy. \quad (1.27)$$

In equation (1.27) $p(y|e)$ is the marginal probability distribution of data conditioned on experiment, e. This marginal distribution is obtained by integrating over the prior distribution of parameter θ, $p(\theta)$. The prior probability distribution of parameter θ, $p(\theta)$, represents our prior knowledge about θ. Lindley proposes that the best experiment design e^{opt} is the one that maximizes the pre-posterior expected utility[81],

$$U(e^{opt}) = \max_{e\in E}\int_{Y}\ \max_{d\in D}\int_{\Theta}\ u(d, \theta, e, y)\ p(\theta|y)\ p(y|e)\ d\theta dy. \quad (1.28)$$

According to Lindley, for designing an optimal experiment design, a utility function should be defined that reflects the purpose of the experiment. The experimental design is then a decision problem and the design that maximizes the utility function is the optimal design[80, 81].

Utility Functions

A utility function assigns a numerical value to consequences based on the preferences of the decision maker. In the health industry, there are multiple decision makers and these may have different preferences, attributes and utilities. For the drug industry, profit and

the return on investment is a critical goal; for a regulatory agency, safety is of primary importance; and for the patient, the decision among treatment interventions is based upon the relative cost, efficacy and safety profile. Lindley and Singpurwalla consider design methods where a manufacturer should design an experiment to convince the consumer or the regulatory agency about the viability of their product[79]. They suggest that manufacturer should design an experiment that maximizes the utility of the consumer/regulatory agency, taking into account the cost incurred for experimentation[79]. The consumer's utility may be "is the drug better than standard of care" or "does the drug provide substantial efficacy with limited risk of side effects". In drug development, the drug is evaluated for safety and efficacy in multiple sequentially designed Phase I and Phase II trials. An optimal drug development strategy will be to constantly evaluate the consumer's utility throughout the development process. An early assessment of this utility will enable a quicker Go/No Go decision and monetary savings.

Multi-Attribute Utility

Multi-attribute utility functions are used when the decision has to consider more than one goal. In drug development or clinical therapy one may wish to find a dose that maximizes over multiple safety and efficacy attributes. Several authors have used multi-attribute utility concepts to problems of dose/dosage regimen selection and drug development decisions[82-85]. Poland and Wada combined drug-disease, economic, pharmacokinetic and pharmacodynamic models of an HIV protease inhibitor to guide the drug development decision of a once daily versus twice daily dosing regimen for experienced or treatment naïve subjects[82]. Ouellet *et al*. combined safety and efficacy attributes as a clinical utility

index to select between pharmacologically similar compounds for insomnia[84]. Graham *et al*. combined adverse events and efficacy data in a Bayesian decision-theoretic framework to choose between immediate and controlled-release formulations of oxybutynin for urinary urge incontinence[85]. Jonsson and Karlsson used a utility approach considering both beneficial and adverse effects of oxybutynin for individualizing dosing regimens[86]. In the examples mentioned above, the utility is derived under the assumption of utility independence among attributes. The utility independence assumes that there is no synergy, interaction or tradeoff between the attributes of efficacy and toxicity[87] and the joint utility is a weighted sum of utilities of safety and efficacy, equation (1.29),

$$U(\text{eff},\text{tox}) = w_e \cdot u_e(\text{eff}) + w_t \cdot u_t(\text{tox}), \tag{1.29}$$

where, w_e and w_t are the weights of efficacy and toxicity and sum to 1 reflecting relative importance of each attribute, and u_e and u_t are utilities of efficacy and toxicity.

Conway and Petroni proposed a probability trade off technique (PTT) to jointly account for efficacy and toxicity. The method evaluates the maximum threshold of toxicity that can be tolerated for added efficacy and the minimum threshold efficacy that is acceptable with no incidence of toxicity. They used the probability trade off technique to determine stopping rules and a composite test statistic for Phase II trial of an anti-cancer drug[88]. Troche *et al*. proposed the use of PTT to optimize therapeutic intervention for patients considering their subjective preference of toxicity/efficacy thresholds[87]. Yin *et al*. used the odds ratio tradeoffs, to simultaneously account for efficacy and toxicity, in a Bayesian adaptive trial aimed at dose-finding of an anti-cancer drug in subjects with metastatic

breast cancer[89]. Stallgard *et al*. combined utilities for efficacy, toxicity and trial costs in a Bayesian decision-theoretic framework for optimizing Phase II trial designs[90].

Application in Designing PK/PD Experiments

The designs for PKPD studies can be improved by using optimal design approaches. Most optimal design approaches in PKPD uses D-optimality criteria available in softwares such as PFIM[91] and PopED[92]. The D-optimality refers to finding an optimal design that maximizes the determinant of the Fisher information matrix and thus focuses on the precision of the parameter estimates[93]. However, the literature on the use of formal Bayesian decision-theoretic approaches is limited. Chaloner and Verdinelli give a review of Bayesian decision-theoretic experimental designs for linear and non-linear models[80]. They suggest that the decision-theoretic framework can be justified for many non-Bayesian optimality criteria. For example, when inference on model parameters is the main goal of the analysis, a utility function based on Shannon information leads to D-optimality[80]. Atkinson *et al*. compared the D-optimal designs to Bayesian decision-theoretic designs for finding optimal sampling time points for a three compartment model. The objective of the experimental design was to minimize asymptotic variance of the area under the curve, time to maximum concentration and the maximum concentration[94]. Stroud *et al.* used a decision-theoretic utility approach in designing optimal sampling times for paclitaxel. The sampling times were chosen that maximized the utility function of the posterior precision of AUC/time above a critical value while minimizing the inconvenience of the patients at the chosen time points[95]. Han and Chaloner used Bayesian approaches to optimize sampling time points for a HIV dynamic

model that maximized the function of the posterior precision of parameters[96]. Palmer and Muller optimized the apheresis schedule in cancer patients prior to high dose chemotherapy. The utility function minimized the posterior predictive loss function of the estimated parameters[97].

In some experimental designs, prediction of the future observations might be more important than precision of parameters. Wakefield individualized dosing regimens with the utility that minimizes the posterior predictive loss function. The loss function penalizes the doses that resulted in model predicted concentrations that did not fall within a pre-specified therapeutic concentration range[98]. Rosner *et al*. describe a Bayesian decision theoretic approach to individualize dosing regimen for high dose chemotherapy in leukemia patients[99]. The study used historic data to derive prior information on the pharmacokinetic model and dose linearity. Following a relatively lower test dose to a new patient, an optimal therapeutic dose for the patient was obtained that minimized a loss function of the predicted AUC not falling in a target therapeutic range[99].

An important feature in Bayesian experimental design is the use of prior information. The drug industry may wish to design an experiment using all prior information, but a regulatory agency might look at the results with different prior beliefs about the safety and efficacy of such compounds. Etzioni and Kadane discusses sample size estimation for an experiment where the manufacturer and the consumer have different prior beliefs about the product[100]. Han and Chaloner considered the possibilities of different priors at both the design and analysis stages while designing optimal sampling times for a HIV model[80].

Decision-Theoretic Approaches in Clinical Trials

The use of decision analysis in context of PKPD modeling and clinical trials is discussed by Sheiner[65]. In Sheiner's view, the purpose of the empirical investigation of the dose-response model is to obtain the probability distribution of outcomes as a function of dosage regimen[65]. The probabilities that affect the outcome are the individual PK/PD parameters and a set these are collectively known as the response surface. Sheiner suggested that mapping of the response surface is sufficient for making optimal dosing decisions. This view is in contrast to the empirical hypothesis testing approach for determining efficacy and drug labeling decisions. Sheiner's view of the optimal decision is based on the normative decision-theoretic approach of maximizing utility. He suggested that the regulatory agencies and drug developers should not focus on suggesting a dose, but instead provide information on probabilities and outcomes that will allow individuals to make an optimal dose decision based on their subjective preference for utilities to risk/benefit[65].

Sheiner described the drug development process as cycles of learning and confirming. The goal of the learning phase is to generate information and map the response surface, and the confirming phase aims at falsifying a specific hypothesis[65]. The design, analysis and goals of the two phases differ significantly. In the learning phase, there are a large numbers of design variables, whereas the confirming phase is typically done in a homogenous population with few design variables[65]. The design and analysis in the learning phase can be most suitably Bayesian, while the confirming phase is typically frequentist.

The Bayesian decision theoretic approach to experimental design has been used to accommodate both learning and confirming phases of drug development. The learning and confirming phases are sequentially spaced clinical trials and the results from one trial can affect the future course of experimentation. Muller *et al.* discuss simulation-based approaches for the maximization of the expected utility in sequential decision problems. The approach has been used for dose-finding trial for a new therapeutic intervention. The results obtained at a particular time was used to decide if the trial should be abandoned for futility, continued for dose finding, or moved forward to a pivotal confirmation trial[101]. A key objective of early phase I/II and proof of concept studies is to characterize the dose-response curve. Evaluation of the dose response curve is essential in determining the optimal dosing regimen. A standard dose-response curve study evaluates sets of pre-specified dose levels with nearly equal numbers of subjects assigned to each dose level. Berry *et al.* argues that such a trial may be inefficient because the dose selected may or may not have dose-response relationship in the interval of doses considered for the trial[37]. Moreover, if the curve has a sigmoidal Emax relationship, the dose response may be flattened at higher doses and the observations may seem to be wasted. Another drawback is the fixed sample size in the trial at each assigned dose. The authors suggest a Bayesian adaptive trial design using Bayesian updating and Bayesian decision analysis. Using a pre-specified termination rule, the adaptive trial either continues patient allocation until the required precision of ED95 (dose required to get 95% of maximum possible response) is obtained, or stops the trial for futility, or can switch to confirmatory mode. This seamless Phase II/III trial design minimizes patient

resources and adapts to variability in accumulating data, and saves time and resources by stopping early for futility or switching to confirmatory mode[37]. Smith *et al.* used a Bayesian adaptive design in a proof-of-concept study for dose-finding of a new candidate therapy for neuropathic pain[102]. They used a Bayesian decision-theoretic approach to drop the ineffective doses or doses that were not tolerated from future randomization[102]. Smith and Marshall describe a Bayesian dose-response study where prior information on an existing, therapeutically similar, compound can be used in designing a study for a novel compound[103]. Krams *et al.* used a Bayesian adaptive trial for dose finding of a therapeutic agent for acute stroke therapy. The objective of the trial was to determine a dose that gave a clinically relevant effect in patients[104].

Summary

The normative model of decision making is known as decision theory. The principles of decision theory are applied in a wide variety of disciplines like clinical practice, econometrics, statistics, economics, theory of games and artificial intelligence. Central to the idea of decision making under uncertainty is the principle of maximization of expected utility. An optimal decision or action is the one that maximizes the utility of the consequences. A Bayesian decision-theoretic experimental design provides a formal framework for incorporating prior information and decision theoretic approaches to finding optimal designs. The flexible framework within Bayesian experiment design allows multiple goals to be maximized simultaneously; including parameter precision, budget constraints, and patient's convenience and contrasting prior opinions at design and analysis stages. The extension of utility theory over multiple goals is referred to as multi-

attribute utility theory. Recently multi-attribute utility has gained attention in the pharmacometrics community in benefit/risk assessment and dose selection. In this thesis, Bayesian decision-theoretic approaches are used to optimize dose and experimental design for PD 0200390. The design variables dose and sample size were maximized over multiple attributes of efficacy and safety.

The objective of this thesis is to explore the relationship between the safety and efficacy endpoints of sleep and dose of PD 0200390 and to use this relationship for optimal clinical trial design of PD. There are two specific aims in this thesis. First specific aim is to develop hierarchical Bayesian dose-response models for the multiple safety and efficacy endpoints for PD 0200390 that is described in Chapter 2. The second specific aim is to use Bayesian decision-theoretic approach to dose finding and trial design for PD 0200390 and is described in Chapter 3.

2. Modeling Objective

PD 0200390 (PD) is an alpha-2-delta subunit binding compound that is being considered for the treatment of insomnia. There were two proof-of-concept studies for which data were available. Multiple safety and efficacy endpoints were available from each individual. The first objective of this analysis was to evaluate dose-response relationships for these efficacy and safety endpoints. Safety and efficacy endpoints can be correlated, and estimating this correlation can help in understanding the pharmacology of the drug action and in future trial simulations. The second objective of the modeling work was to compare the simultaneous-outcome-fitting of the endpoints to the individual-outcome-fitting of the endpoints.

2.1. Methods

Data sources

Data were pooled from two Phase II randomized, double-blind, placebo and/or active controlled trials. Study A4251003 and study A4251068 were carried out in subjects with primary insomnia. The details of trial design and the inclusion/exclusion criteria are described below.

Study A4251003

This study was a Phase IIa, randomized, double-blind, placebo-controlled, active-comparator controlled, 4-way crossover trial in subjects with primary insomnia. The four treatments consisted of two doses of PD, 25 mg and 75 mg, placebo, and as an active comparator, zolpidem (10 mg). Each treatment was administered for two consecutive

nights with approximately one week between each treatment. For each treatment, study medication was administered thirty minutes prior to bed time and polysomnography (PSG) was carried out for 8 h. Forty three subjects were enrolled in this study.

Inclusion Criteria

Healthy subjects, age 18-64, males or females (non-pregnant and non-lactating) with three month history of primary insomnia as defined by DSM –IV diagnostic criteria were enrolled in the trial. There were two screening visits prior to study enrollment to ascertain if a subject met the inclusion criteria. The diagnostic criteria for primary insomnia were: subjective average sleep latency of >=30 minutes, subjective average total sleep time of <=6.5 hr, mean total sleep time between 4h to 7h at the two PSG screening nights and total sleep time of <435 minutes on both nights, mean latency to sleep of >=20 minutes at the two PSG screening nights, and neither screening nights with latency of persistent sleep of <15 minutes.

Exclusion Criteria

Subjects that presented any of the following were excluded from the study: current use of medication that were not considered acceptable by the clinical investigators, consumption of >3 alcoholic drinks per day, consumption of >5 caffeine doses per day, smoking >15 cigarettes per day, suffering from chronic pain that can interfere sleep, and taking regular naps during day time.

Study A4251068

This study was a single dose, randomized, double-blind, placebo-controlled, trial with a 5-way crossover Williams design (a type of Latin-square design, where each treatment is followed by another at least once to account for carry over effects). Eighty-three subjects with primary insomnia were randomized to four doses of PD 0200390 (5 mg, 15 mg, 30 mg and 60 mg) and the placebo group.

Inclusion Criteria

Males and females (non-pregnant and non-lactating) aged 18-64 years with three month history of primary insomnia according to the DSM-IV-TR criteria. There were two screening visit prior to the study enrollment to ascertain if subject meets the inclusion criteria. The diagnostic criteria for primary insomnia were: at screening visits the subject must have subjective wake after sleep onset of >=60 minutes, latency to sleep onset of >=45 minutes, and subjective total sleep time of <=6.5 h. Subjects must meet the PSG criteria of mean wake after sleep onset of >=60 minutes, and total sleep time between 3h and 7h on both the screening nights and mean latency to persistent sleep of >=20 minutes with neither screening night with latency to persistent sleep of <15 minutes.

Exclusion Criteria

Subject that presented any of the following were excluded from the study: history or presence of psychiatric disease, history or presence of breathing related sleep-disorder, narcolepsy or any other dissomnias, parasomnias or restless legs syndrome, history or presence of medical or neurological condition interfering with sleep, current use of medication with psychotropic effects (drugs prescribed for depression, mania, anxiety or

psychosis), alcohol or substance use in last six months, or excessive consumption of caffeine or cigarettes.

In both the trials, treatments were separated by 7 treatment-free nights. For each treatment, the medication was administered for two consecutive nights, 30 minutes prior to the habitual bedtime. Subject demographics for the two trials are shown in Table 2.

Table 2 Subject demographic characteristics stratified by study group

Demographics	A4251003	A4251068	Total
Age (years) Median [Range]	46 [26, 63]	47 [19,64]	47 [19, 64]
Gender (M/F)	18/25	28/55	46/80
Weight (kg) Median [Range]	76 [46, 121]	74 [46, 113]	75 [46, 121]

Response Variables

There were four response variables, two efficacy and two safety, available for the data analysis. The safety and efficacy variables are described below.

Efficacy Variables

The number of awakenings from sleep (NAASO) and wake time after sleep onset (WASO) are the efficacy variables. The efficacy measures are obtained from the PSG study following drug administration and are described in Chapter 3.

Safety Variables

Ease of awakening from sleep (AFS) and early morning behavior following wakefulness (BFW) are the measures of residual effects of the drug. These residual measurements are obtained through the Leeds sleep evaluation questionnaire (LSEQ)[99], consisting of visual analog scale (VAS) questions, measured on a 100 mm line. It was administered to each subject 30 to 60 minutes after completion of the PSG recording, the morning after treatment administration. High between-subject and within-subject variability was observed in the plot of VAS score vs. dose; therefore these scores were converted to a binary variable, 0/1, whereby a decrease of 20% and 30% from placebo in the AFS and BFW VAS score, respectively, was considered as an adverse event.

Data Analysis Procedure

A Bayesian hierarchical model was fitted in WinBUGS[44] 1.4 (MRC Biostatistics Unit, Cambridge, UK) and BRugs[45] package in R (version2.8.1): A language and Environment for Statistical Computing[100]. Generalized hierarchical mixed models were developed to describe the dose-response relationships for WASO, NAASO, AFS and BFW. The explanatory variables available were age, gender, dose and study group. The data analysis procedure consisted of the following steps.

Hierarchical non-linear mixed effects models were fitted to each safety and efficacy endpoint separately to identify dose-response models and covariate relationships (individual-outcome-fitting)

After reasonable models were identified, the final dose-response models were then simultaneously fitted together to account for correlation (simultaneous-outcome-fitting).

The final model was evaluated using posterior predictive and visual predictive checks.

Model Specification

The hierarchical Bayesian model consisted of three levels. In the first level the distribution assumptions were made for the pharmacodynamic response variables. In the second level the distribution assumptions were made for the subject-specific random effects and in the third level of the hierarchy, prior distributions for the parameters were specified.

The WASO observations are always positive and the log-transformed WASO values were described by a normal distribution. The NAASO is a count measure and a Poisson distribution was used to describe the number of awakenings. The residual measures, AFS and BFW, are binary and were described as Bernoulli random variables. These probability distribution functions, (Log-Normal, Poisson and Bernoulli), belong to a mathematical class of functions called the exponential family of distributions[101].

In the first level of hierarchy, let Y_{kij} denote the k^{th} type of response ($k = 1, \ldots, 4$, for WASO, NAASO, AFS and BFW) in the i^{th} individual, (where $i = 1, \ldots, I$), at the j^{th} dose, (where, $j = 1, \ldots, J_i$, J_i is the total number of doses received by the i^{th} individual) and let ds_{ij} represent the j^{th} dose level in the i^{th} subject. This notation is used to account for different levels of doses that were administered in study A4251068 and study A4251003. The observation, Y_{kij}, conditioned upon the parameter vector ν_{kij} and

dispersion parameter ψ_k, was modeled to arise from an exponential family of distributions with probability density function written in general form as

$$f(Y_{kij}|\ \nu_{kij}.\psi_k) = \exp\left\{\frac{Y_{kij}\cdot\nu_{kij} - d(\nu_{kij})}{a_k(\psi)} + c_k(Y_{kij}.\psi)\right\}, \tag{2.1}$$

where, $d(\nu)$ and $c(Y,\psi)$ are known functions that are specified in Table 3. The conditional mean and variance of the response variable Y_{kij} are given as

$$\begin{aligned} E(Y_{kij}|\nu_{kij}.\psi_k) &= d'(\nu_{kij}) \\ Var(Y_{kij}|\nu_{kij},\psi_k) &= a_k(\psi)\cdot d''(\nu_{kij}) \end{aligned} \tag{2.2}$$

The properties of the observations from the exponential family of distributions are further described by MuCullagh and Nelder[101]. Similar methods for modeling multiple responses from the exponential family are used by O'Malley *et al.*[102] and Zhu *et al*.[103] and Gueorguieva[104].

Table 3 Characteristics of some distributions in the exponential family

Response	**Log(WASO)** ($k=1$)	**NAASO** ($k=2$)	**AFS** ($k=3$)	**BFW** ($k=4$)
Distribution	Normal	Poisson	Binomial	Binomial
PDF	$\frac{1}{2\pi\sigma^2}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$	$\frac{\mu^y e^{-\mu}}{y!}$	$\binom{n}{y}\mu^y(1-\mu)^{n-\mu}$	$\binom{n}{y}\mu^y(1-\mu)^{n-\mu}$
$\boldsymbol{\nu}$	μ	$\log\mu$	$\log\left(\frac{\mu}{1-\mu}\right)$	$\log\left(\frac{\mu}{1-\mu}\right)$
$d(\boldsymbol{\nu})$	$\frac{\nu^2}{2}$	e^{ν}	$-n\log(1+e^{\nu})$	$-n\log(1+e^{\nu})$
$a(\psi)$	σ^2	1	1	1
$c(\boldsymbol{Y},\psi)$	$-\frac{1}{2}\left(\frac{y}{\sigma^2}-\log 2\pi\sigma^2\right)$	$-\log y!$	$\log\binom{n}{y}$	$\log\binom{n}{y}$
$\boldsymbol{E(Y)}$	μ	μ	$n\mu$	$n\mu$
$\boldsymbol{Var(Y)}$	σ^2	μ	$n\mu(1-\mu)$	$n\mu(1-\mu)$

PDF: Probability density function

In generalized models, a link function relates the prediction from the model to the expected value of the datum Y. In cases, when the natural parameter, ν, is used as a link function, it is also referred to as the canonical link function. In this modeling approach, canonical link functions were used against predictors. The canonical link function, $\nu_k(\cdot)$, transforms the mean, μ_{kij}, for the k^{th} response on to the interval $(-\infty, +\infty)$, which is then regressed to identify dose-response and covariate relationships. The log and logit transformations are types of link functions used for transforming the means of Poisson and Bernoulli random variables on to the interval$(-\infty, +\infty)$. For example, in a pharmacokinetic model where the concentrations are assumed to be log-normal, the log-transformation is the link function. The transformed mean function $\nu_k(\mu_{kij})$ was modeled as a function of dose, ds_{ij}, the fixed effects parameter vector β_k and the random-effects parameter vector ϕ_{ki}, as shown in equation (2.3),

$$\nu_k(\mu_{kij}) = f_k(ds_{ij}, \beta_k, \phi_{ki}), \tag{2.3}$$

where, the function $f_k(\ )$ denotes a structural model for the k^{th} response variable.

The parametric distribution and covariate relationships for the latent subject-specific random effects, ϕ_{ki}, were specified in the second stage of the hierarchy. The log-normal distribution of latent subject-specific parameter ϕ_{ki} was specified as in equation (2.4) and a normal distribution was specified as in equation (2.5).

$$\phi_{ki} = g_k(\boldsymbol{\theta}_k, \boldsymbol{z}_i) \cdot e^{\eta_{ki}} \tag{2.4}$$

$$\phi_{ki} = g_k(\boldsymbol{\theta}_k, \boldsymbol{z}_i) + \eta_{ki} \tag{2.5}$$

where $g_k(\ )$ is a parametric function of the covariate model for the k^{th} response and reduces to $\boldsymbol{\theta}_k$ if no covariate model is specified, $\boldsymbol{z}_i$ is the covariate vector for the i^{th} subject, $\boldsymbol{\theta}_k$ are the fixed effects parameters and $\boldsymbol{\eta}_{ki}$ is the vector of individual specific random effects, in the i^{th} subject for the k^{th} response variable, that have a multivariate normal distribution with a mean vector fixed to $\mathbf{0}$.

In this stage, a dual modeling strategy was applied: an individual-outcome-fitting approach was applied to determine a structural model for each outcome, followed by a simultaneous-outcome-fitting approach.

Individual-Outcome-Fitting

In the individual-outcome-fitting approach, the second level of the hierarchical Bayesian model for the k^{th}outcome was independent of the other dose-response models. Therefore four separate models were developed, whereby the random effect vector, $\boldsymbol{\eta}_{ki}$, for the k^{th} outcome in the i^{th} individual had a $p_{(k)}$-dimensional multivariate normal distribution with a $p_{(k)} \times 1$-dimensional mean vector of $\mathbf{0}$, and a $p_{(k)} \times p_{(k)}$-dimensional variance covariance matrix, $\Sigma_{p(k)}$, as shown in equation (2.6),

$$\boldsymbol{\eta}_{ki} \sim MVN_{p(k)}(\ \mathbf{0}\ ,\ \Sigma_{p(k)}), \tag{2.6}$$

where $MVN_{p(k)}(\cdot,\cdot)$ denotes a p-dimensional multivariate normal distribution for the k^{th} response model. In the individual-outcome-fitting there was no borrowing of information between the four response types.

Simultaneous-Outcome-Fitting

The four responses were fitted simultaneously in the simultaneous-outcome-fitting approach[104]. Hence, the latent variable vector for the four responses, $\boldsymbol{\eta}_i = \{\boldsymbol{\eta}_{1i}, \ldots, \boldsymbol{\eta}_{4i}\}$, in the i^{th} individual, had a $\sum_{k=1}^{4} p_{(k)}$-dimensional multivariate normal distribution (equation (2.7)).

$$\begin{pmatrix} \boldsymbol{\eta}_{1i} \\ \boldsymbol{\eta}_{2i} \\ \boldsymbol{\eta}_{3i} \\ \boldsymbol{\eta}_{4i} \end{pmatrix} \sim MVN_{\sum_{k=1}^{4} p_{(k)}} (\mathbf{0} , \Sigma) \tag{2.7}$$

In equation (2.7), the dimension of the multivariate normal distribution, $\sum_{k=1}^{4} p_{(k)}$, is the sum of the dimensions of the $p_{(k)}$ random effects for the four response types. The simultaneous fitting approach is particularly useful if the four responses are correlated. Accounting correlation between the response variables will reduce uncertainty in the future simulation studies. Further, the correlation estimates can also suggest latent processes and possible dependencies.

The major strength of Bayesian analysis is the incorporation of prior information in making inference. However, in this analysis non-informative priors were specified in the third stage of hierarchy because of unavailability of prior data. The choice of probability distributions to specify non-informative priors is also discussed in Chapter 1.2. The prior distribution of the precision parameters ($1/\sigma^2$) was specified as $G(a, b)$ where G is a gamma distribution with parameters a and b being the shape parameters. The prior for the inverse of the variance-covariance matrix (Σ^{-1}) was specified as $W(\rho R, \rho)$, where W is a

Wishart distribution, R is the scale matrix and ρ is the degrees of freedom. For the fixed effects parameters in β and the means of the random effects in $\boldsymbol{TV\theta}$, the priors were specified as normal distributions $N(q, H)$, where q and H are the mean and variance of a normal distribution. The specific values of the parameters of the prior distribution are provided in the Appendix A.

Model Fitting

Bayesian model selection and inference was performed in WinBUGS and BRugs. To perform MCMC integration, two chains were run for 40,000 iterations and initial 10,000 iterations (samples) from both chains were discarded during the burn-in phase. Convergence diagnostics (the trace plot and bgr statistics), were used to check if the Markov chain converged to a stationary distribution. A description of convergence diagnostic tools is given in Chapter 1.2. The MCMC samples after burn-in from the two chains (60,000 samples) were pooled to obtain inference on the posterior distribution of the parameters. The posterior means of the random effects for each individual were plotted against the covariates to elucidate possible relationships.

Covariate Model Selection

The covariates age, gender and study (A4251003/A4251068) were tested for incorporation in the model. Age was a continuous covariate, whereas gender and study were categorical; where males and study A4251068 were assigned label 1 else 0. In the model selection process the DIC and PsBF were used. A description of model selection tools, DIC and PsBF, is given in Chapter 1.2. The PsBF was calculated for the vector of

observation for the i^{th} subject at the j^{th} dose level to compare between the individual-outcome-fitting and the simultaneous-outcome-fitting models.

A detailed description of PsBF is given in Chapter 1.2. In summary, the PsBF approach compares models based on cross-validated predictive densities[49]. The approach computes the predictive density of an observation deleted from the analysis dataset. In the PsBF approach, the model selection paradigm shifts from selecting the model that best fits the data to selecting the model that best predicts a future dataset that originates from the same process as the original dataset. The computation of PSBF for univariate observations is discussed in Chapter 1.2. The following section shows the computation of PsBF for a vector of K responses in the i^{th} subject receiving the j^{th} dose. Let $\mathbf{y}$ be response vector of all data, $\mathbf{y}_{(kij)}$ be a vector of all observations *without* the k^{th} response in the i^{th} subject receiving the j^{th} dose and $\mathbf{y}_{(\cdot ij)}$ be a vector of all observations *without* a vector of K responses in the i^{th} subject receiving the j^{th} dose, (where, $j = 1, ..., J_i$, J_i is the total number of doses received by the i^{th} individual) and $\mathbf{y}_{\cdot ij} = \{y_{1ij}, \ldots, y_{4ij}\}$. Then the conditional predictive density for a vector of K outcomes, with an assumption of independence between observations, is given by the following equation (2.8),

$$f(\mathbf{y}_{\cdot ij}|\mathbf{y}_{(\cdot ij)}) = \int_{\Theta} \prod_{k=1}^{K} f(y_{kij}|\boldsymbol{\theta})\ \pi(\boldsymbol{\theta}|\mathbf{y}_{(\cdot ij)})\ d\boldsymbol{\theta}, \tag{2.8}$$

where, $\pi(\boldsymbol{\theta}|\mathbf{y}_{(\cdot ij)})$ is the posterior distribution of parameters when the model is fitted to data $\mathbf{y}_{(\cdot ij)}$. The log pseudo marginal likelihood (LPML) can be given as

$$LPML = \sum_{i=1}^{I} \sum_{j=1}^{J_i} \log f(\mathbf{y}_{\cdot ij} | \mathbf{y}_{(\cdot ij)}). \tag{2.9}$$

Using the posterior density of parameter $\pi(\boldsymbol{\theta}|\mathbf{y})$, and the Monte Carlo estimate of the CPO, $f(\mathbf{y}_{\cdot ij}|\mathbf{y}_{\cdot ij})$, is given as

$$f(y_{\cdot ij}|\mathbf{y}_{\cdot ij}) = \left[\frac{1}{n} \sum_{i=1}^{n} \left[\prod_{k=1}^{K} f(y_{kij}|\boldsymbol{\theta}^{(iter)}) \right]^{-1} \right]^{-1}. \tag{2.10}$$

For inclusion of covariates, the LPML values were compared between the models. A higher LPML indicates better prediction ability of the model. An increase in LPML by 2 to 5 units was considered positive (Table 1).

Model Evaluation

Posterior predictive and visual predictive checks were used to assess if the model fits the data in a reasonable way, The model evaluation tools, posterior predictive checks and visual predictive checks, are discussed in Chapter 1.2 .

A good discrepancy measure for the posterior predictive check for the dose-response model is the prediction of the percentage of subjects at each dose level that achieves a clinically significant difference from placebo (CSD). A reduction from placebo of 40% and 20% for WASO and NAASO, respectively, is a clinically significant difference. The magnitude of the change from placebo that defines a clinically significant difference for the endpoints of WASO, NAASO, AFS and BFW were obtained through communication with project team members at Pfizer Inc. The posterior-predictive check was used to compare the spread of the 95% prediction interval of the simulated proportion of subjects with CSD to the observed proportion of subjects with CSD. Model evaluation through

PPC was performed for WASO and NAASO response variables. The algorithm used for the calculation of the posterior predictive check is given below.

For the i^{th} subject, obtain a sample of model parameter vector $\boldsymbol{\theta}^*$ from the posterior distribution of the parameters, $p(\boldsymbol{\theta}|\mathrm{y})$. The parameter vector $\boldsymbol{\theta}^*$ consists of a vector of fixed effects parameter, $\boldsymbol{\beta}_k^*$, and the individual random effect vector $\boldsymbol{\phi}_{ki}^*$ and the variance estimates.

Compute the mean response, μ_{kij}^*, for the k^{th} response in the i^{th} subject at the j^{th} dose level, ds_{ij}

$$\mu_{kij}^* = \nu_k^{-1}\left[f_k(ds_{ij}, \boldsymbol{\beta}_k^*, \boldsymbol{\phi}_{ki}^*)\right]. \tag{2.11}$$

Simulate response $y_{ijk}^* \sim \ dist_k(\cdot)$ from the appropriate family of distributions, $dist_k$. Log-normal for WASO or Poisson for NAASO.

The measure of interest is the proportion of subjects in the trial with a clinically significant reduction from placebo. Let $\mathrm{t}_{kij}^* = \begin{cases} 1 & \left\{\dfrac{y_{kij}^* - y_{ki(placebo)}^*}{y_{ki(placebo)}^*}\right\} \geq CSD_k \\ 0 & \text{otherwise} \end{cases}$,

where $j \neq placebo$. The new variable t_{kij}^* defines the occurrence of a clinically significant event in the i^{th}individual after taking j^{th} dose level of PD.

Repeat steps 1-4 for $i = 1, \ldots, I$ subjects taking $j = 1, \ldots, J_i$ doses as in the dataset.

Calculate the discrepancy measure $D_{kj}(\mathbf{y}^*)$ as the proportion of subjects in the simulated trial with a clinical significant event for the k^{th} response at the j^{th} dose level.

$D_{kj}(\mathbf{y}^*) = \frac{1}{I}\sum_{i=1}^{I} t^*_{kij}$.

Repeat steps 1-6 it for $n = 1, \ldots, 200$ simulated trials.

Compare the plot of $D_{kj}(\mathbf{y}^*)$ from N simulated trials to the observed data discrepancy statistic $D_{kj}(\mathbf{y})$.

The posterior predictive check was analyzed visually for model selection and criticism. The visual predictive check compares the observed data to the model predicted data (step 1 through 3 above). Model evaluation for AFS and BFW was done using visual predictive checks.

2.2. Results

A complete set of observations for the four response variables, were available for 118 subjects from the two Phase 2a studies, Two dose levels (25 mg, 75 mg) and placebo data from 40 subjects in study A42510003, and four dose levels (5 mg, 15 mg, 30 mg and 60 mg) and placebo data from 78 subjects in A4251068, were used in the modeling process.

The convergence diagnostics, bgr statistics and trace plots, for the first 20,000 iterations of the final model parameters, (fixed effects, typical values of the random effects and precision of WASO), from the simultaneous-outcome-fitting of endpoints are shown in Figure 3 and Figure 4, respectively. The red line in the plot of the bgr statistic (Figure 3) is the ratio of the between-chain 80% credible interval to the within-chain 80% credible

interval, and reaches to one at convergence at around 5000 iterations. The green line in the plot of the bgr statistic (Figure 3) represents the normalized width of the 80% credible interval for the pooled samples from the two chains. The blue line in the plot of bgr statistic is the average of the normalized width of the 80% credible interval for the individual chains. Both, the red and the green line, converge to stability at around 5000 iterations representing convergence (Figure 3). In the trace plots for these parameters (Figure 4), the appearance of a "fat hairy caterpillar" is seen for most of the parameter except $\theta 1$, $\theta 3$ and $\theta 4$ that have appearance of a "wiggly snake". The appearance of a "fat hairy caterpillar" represents samples from a stationary distribution, whereas a wiggly snake represents auto-correlation between samples. Owing to auto-correlation the samples were collected up to 80000 iterations. The posterior density estimates of these parameters, (20,000-80,000 iterations), are shown in Figure 5. The density plots shows that parameters have unimodal distributions that are symmetric around the mode. Summary statistics (mean, median, 95% CI) of the parameters were computed from 60,000 iterations obtained after 20,000 "burn-in" iterations. The parameter estimates and the correlation matrix from the simultaneous-fitting of the endpoints are reported in Table 4 and Table 5, respectively.

The relationship of WASO and NAASO to dose was described as an inhibitory Emax model and the inter-individual variability was incorporated as log-normal on the baseline $E0$, defined as the response at baseline or placebo, and $ED50$, defined as the dose that brings about 50% of the maximum inhibition, (equation (2.12) and equation (2.13), respectively). The Emax values, defined as the maximum possible inhibition, in the two

models, were fixed to the value of 100% of baseline. This was done because upon estimation, Emax values were estimated to be close to one: 0.99, (95% CI: [0.96, 0.99]), for WASO and 0.93, (95% CI: [0.78, 0.99]), for NAASO. This is further supported by the pharmacology of the drug; a high enough dose of PD can sedate a person sufficiently to not have an awakening or wake time during the eight hour of the PSG recording. The relationship of the logistic mean functions of AFS and BFW to dose was described as linear with an additive normally-distributed inter-individual variability (equation (2.14) and equation (2.15), respectively). The WinBUGS code for individual-outcome-fitting and simultaneous-outcome-fitting approach is given in Appendix A. The final model equations for the WASO, NAASO, AFS and BFW variables are given in equation (2.12), equation (2.13), equation (2.14) and equation (2.15).

WASO Model

$$\begin{aligned} \log(Y_{1ij}) &\sim Normal(\mu_{1ij}, \sigma^2) \\ \mu_{1ij} &= E0_{1i} \cdot \left(1 - \left(\frac{dose_{ij}}{ED50_{1i} + dose_{ij}}\right)\right) \\ E0_{1i} &= TVE0_1 \cdot e^{(\eta_{1i})} \\ ED50_{1i} &= TVED50_1 \cdot e^{(\eta_{2i})} \end{aligned} \tag{2.12}$$

NAASO Model

$$\begin{aligned} Y_{2ij} &\sim Poisson(\mu_{2ij}) \\ \mu_{2ij} &= E0_{2i} \cdot \left(1 - \frac{dose_{ij}}{ED50_{2i} + dose_{ij}}\right) \\ E0_{2i} &= \exp(\theta 1) \cdot \left(\frac{age_i}{47}\right)^{\theta 2} \cdot \exp(\theta 3)^{gen_i} \cdot \exp(\theta 4)^{std_i} \cdot e^{(\eta_{3i})} \\ ED50_{2i} &= ED50_2 \cdot e^{(\eta_{4i})} \end{aligned} \tag{2.13}$$

Where, gen_i and std_i are gender and study effects, respectively. Males and study A4251068 is labeled 1.

AFS Model

$$Y_{3ij} \sim Bernoulli(\mu_{3ij})$$
$$\log\left(\frac{\mu_{3ij}}{1-\mu_{3ij}}\right) = \theta 5 + \theta 6 \cdot dose_{ij} + \eta_{5i} \tag{2.14}$$

BFW Model

$$Y_{4ij} \sim Bernoulli(\mu_{4ij})$$
$$\log\left(\frac{\mu_{4ij}}{1-\mu_{4ij}}\right) = \theta 7 + \theta 8 \cdot dose_{ij} + \eta_{6i} \tag{2.15}$$

The covariates, age, gender and study, were introduced in the model in a proportional manner as shown in equation (2.13) for the $E0$ variable. For the NAASO model, age and study effects were marginally significant covariates for the baseline E0 parameter. For model comparison, a drop in DIC by 3-5 units and an increase in 2logPsBF by more than 2-5 units are considered significant. Inclusion of age resulted in a drop in DIC by 1 unit and an increase in 2logPsBF by 4 units. Addition of study effect on E0 further resulted in a drop in DIC by 3 units and an increase in 2logPsBF by 2 units. The covariate gender was not significant in the model assessment for the NAASO variable but was a significant covariate during previous individual-outcome-fitting in NONMEM, a non-linear mixed effects modeling program for population PK/PD analysis that uses maximum likelihood approach. A reason could be that this dataset uses eight less individuals, 118, than in the NONMEM analysis dataset, where 126 individual were used.

The reason that 118 individuals were used was the availability of the complete set of observation for the four endpoints in these individuals. It is however acknowledged that missing data can be handled by analysis through WinBUGS but was not done in this modeling exercise. The covariate effect of gender was still kept in the model and is not believed to adversely affect model prediction as noted by a drop in DIC by 0 unit and increase in LPML by 0.6 units. The inclusion of age, gender and study on the parameters of the dose-response models for WASO, AFS and BFW were not significant.

A visual predictive check for the model for WASO, NAASO, AFS and BFW are shown in Figure 6, Figure 7, Figure 8 and Figure 9, respectively. The plots show that both the trend and the spread of the data are well captured by the model. The posterior predictive check for the variables of WASO and NAASO (Figure 10and Figure 11, respectively) shows that model captures the overall trend of the data. For the WAASO variable, the discrepancy statistic of the simulated data is over predicted at the dose of 5 mg (Figure 10). For the NAASO variable, the discrepancy statistic of the observed data lies within the 95% prediction interval of the discrepancy statistic of the simulated data for all doses (Figure 11).

The simultaneous-outcome-fitting approach was highly significant compared to the individual-outcome-fitting approach (Table 6). The DIC for the simultaneous-outcome-fitting approach (3761) is 103 points less than that of the individual-outcome-fitting approach (3864). Similarly, the difference in the LPML for the simultaneous-outcome-fitting (-1915) is 36 points higher than that of individual-outcome-fitting (LPML, -1951), i.e. a 2log PsBF of 72 points.

The mean baseline E0 value for WASO was estimated to be 65.4 min/night (95% CI: 58.6 min/night - 73.2 min/night) with an inter-individual variability of 51.5% (95% CI: 42.9% - 60.7%). For the NAASO dose-response model, the mean baseline E0 value for a 47 year old female in the study A4251003 is estimated to be 8.24 counts/night (95% CI: 7.21 counts/night - 9.42 counts/night) with an inter-individual variability of 28.5% (95% CI: 22.8% - 34.8%). There was significant correlation (mean: 0.52, 95% CI: 0.28 - 0.71) between the random effects for the E0 values of NAASO and WASO (Table 5). For the NAASO variable, the model estimated an increasing trend in the number of awakenings with age. For a unit increase in log (age/47), the proportionate change in NAASO counts is 1.22 [exp(0.195)] with a 95% CI between 0.96 [exp(-0.0373] and 1.52 [exp(0.423)]. Males were estimated to have higher a number of awakenings: 1.15 times (95% CI: 1.01 to 1.29) the number of awakenings in females. The subjects in study group A4251068 were estimated to have 1.40 times (95% CI: 1.23 - 1.60) the number of awakenings in the A4251003 study group. The ED50 values for the WASO and the NAASO variables were estimated to be 41.0 mg (95% CI: 32.8 mg - 51.9 mg) and 60.5 mg (95% CI 49.5 mg - 75.2 mg), respectively. The inter-individual variability on the ED50s for the WASO and the NAASO variables were 85.4% (95% CI: 64.8% - 107%) and 56.1% (95% CI: 37.4% - 76.6%), respectively, and were significantly correlated (mean correlation: 0.58, 95% CI: 0.26 - 0.80). The residual effects AFS and BFW were highly and significantly correlated with each other, (mean: 0.81, 95% CI: 0.63 - 0.92) and were negatively correlated with the ED50 values of NAASO and WASO, (-0.27 to -0.35), however, for most of the 95%

CI of correlation estimate, 0 was included in the credible set and should be interpreted with caution.

2.3. Discussion

Hierarchical dose-response models were fitted and evaluated for the response variables of WASO, NAASO, AFS and BFW obtained from the combined data from the two Phase II studies of PD, an investigational compound for insomnia. The dose-response models for WASO and NAASO were fitted as inhibitory Emax models. The logistic models for AFS and BFW were fitted as linear. The analysis showed that an increase in the dose of PD leads to decrease in the wake time after sleep onset and reduces the number of awakenings after sleep onset. The analysis also showed that like other insomnia drugs, an increase in the dose of PD leads to higher incidence of residual effects morning after awakening.

Simultaneous PK/PD modeling is often encountered in the pharmacometrics literature; however reports of simultaneous modeling of correlated safety and efficacy endpoints are limited. A reason is that safety and efficacy data are generally discrete or odd-type data, e.g., time-to-event, count, binary. In the current data analysis approach, a multivariate distribution of latent variables was used to accommodate correlation between the response variables of WASO, NAASO, AFS and BFW. The use of latent variables to simultaneously model continuous and discrete data has been previously mentioned in the literature[103-105]

Simultaneous fitting of the endpoints was estimated to be significantly better at predicting a future vector of observation outcomes compared to the individual-outcome-fitting approach. This is particularly useful in simulation studies, when all four outcomes are simulated, and can lead to lower sample sizes in designing clinical trials. The correlations observed from simultaneous-outcome-fitting were intuitive and pharmacologically meaningful. There were positive correlations between the E0s and ED50s for WASO and NAASO variables. It is assumed that a subject with high tendency for awakening may have longer wake times during his sleep. The measures of AFS and BFW are highly correlated which is plausible because the residual measures are related to each other. If an individual feels reduction in the ease of awaking he/she may also feel the reduction in alertness as assessed by early morning behavior.

In a previously published report, the PD dose-response model for WASO was also fitted as an inhibitory Emax model with Emax parameter fixed at 100% of E0[81]. For the WASO model, Ouellet *et al*. estimated the maximum likelihood estimates for E0 and ED50 values of the WASO model to be 78.6 minutes/night and 20.8 mg, respectively. In the present work, the posterior mean estimates of E0 and ED50 for the WASO model are estimated to be 65.4 minutes/night and 41.0 mg, respectively. This difference may be due to the differences in the data sets. In the previous report, the data were pooled from a study in subjects with primary insomnia and another study in healthy subjects with a phase-shift design. A phase-shift design triggers insomnia by requiring healthy subjects to go to bed 4 h prior to their habitual bedtime. The forced insomnia in healthy subjects

may lead to higher WASO baseline value. Healthy subjects may be more sensitive to drug effect which will lead to lower ED50 values.

In the modeling approach the variables of AFS and BFW, originally collected as millimeters on a 100 mm VAS scale, were converted to binary variables. The reason for doing that was the high variability observed in the VAS score (Figure 12). A conversion from the VAS score to the probability of adverse event was assumed to improve data visualization and interpretation. The value of 20% reduction as an adverse event for AFS was chosen on the basis of two factors, the proportions of subjects with adverse events should be similar for the 25 mg (dose level in Study A4251003) and the 30 mg dose group (dose level in A4251068), and the high overall incidence of adverse events. The impact of categorization on AFS variable is shown in Figure 13. Similarly, a 30% decrease was a decided as a significant event for the BFW variable.

It is worth mentioning the reason to use Bayesian methods for this dose-response modeling exercise. The population model development was initially carried out in NONMEM. The simultaneous-outcome-fitting in NONMEM gave erroneous results with program termination. It is speculated that, a reason for this behavior is that NONMEM uses linearization methods in evaluation of its objective function whereas Bayesian approaches uses MCMC methods in evaluation of integral. The reason and nature of discrepancies in the results from NONMEM and the results from WinBUGS is not explored in this Thesis.

This model will be useful for prediction purposes. For example, it may be of interest to know the percentage decrease in WASO and/or NAASO or the probability of residual effects at a particular dose of PD. On the other hand, one may use the model for finding a dose of PD that provides the desired efficacy or limited probability of residual events. The use of this multivariate model in the decision-theoretic approach to dose-finding of PD is the objective of the next chapter.

Table 4 Final parameter estimates and 95% credible intervals [CI] from the final dose-response model using the simultaneous-outcome-fitting approach. (IIV: Inter-individual variability).

Parameter (units)	Response Variable	Mean	Median	95% CI
$TVE0_1$ (minutes/night)	WASO	65.4	65.4	[58.6, 73.2]
$TVED50_1$ (mg)	WASO	41.0	40.9	[32.8, 51.9]
$TVED50_2$ (mg)	NAASO	60.5	60.3	[49.5, 75.2]
exp($\theta 1$) (counts/night)	NAASO	8.24	8.26	[7.21, 9.42]
$\theta 2$ (power on age)	NAASO	0.195	0.196	[-0.0373, 0.423]
exp($\theta 3$) (gen=1)	NAASO	1.15	1.15	[1.01, 1.29]
exp($\theta 4$) (std=1)	NAASO	1.40	1.40	[1.23, 1.60]
$\theta 5$ (intercept AFS)	AFS	-2.73	-2.71	[-3.52, -2.04]
$\theta 6$ (slope AFS)	AFS	0.0361	0.0359	[0.0234, 0.0492]
$\theta 7$ (intercept BFW)	BFW	-4.00	-3.97	[-5.02, -3.12]

$\theta 8$ (slope BFW)	BFW	0.0416	0.0414	[0.0269, 0.0574]
IIV on $E0_1$ (%)	WASO	51.5	51.1	[42.9, 60.7]
IIVon $ED50_1$ (%)	WASO	85.4	84.3	[64.8, 107.0]
IIVon $E0_2$ (%)	NAASO	28.5	28.2	[22.8, 34.8]
IIVon $ED50_2$ (%)	NAASO	56.1	54.6	[37.4, 76.6]
IIV on AFS (Additive)	AFS	1.76	1.73	[1.26, 2.34]
IIV on BFW (Additive)	AFS	1.61	1.57	[1.07, 2.21]

Table 5 Final estimates and the 95% credible intervals [CI] of the correlation matrix obtained from the simultaneous-outcome-fitting approach.

	$E0_1$	$ED50_1$	$E0_2$	$ED50_2$	η_3	η_4
$E0_1$	1					
$ED50_1$	-0.25 [-0.49, 0.04]	1				
$E0_2$	0.52 [0.28, 0.71]	0.13 [-0.20,0.45]	1			
$ED50_2$	0.16 [-0.21, 0.51]	0.58 [0.26, 0.80]	0.35 [-0.04, 0.66]	1		
η_3	0.06 [-0.22, 0.34]	-0.33 [-0.61, -0.004]	-0.15 [-0.44, 0.17	-0.35 [-0.69, 0.06]	1	
η_4	0.03 [-0.28, 0.32]	-0.27 [-0.59, 0.10]	-0.19 [-0.49, 0.15]	-0.35 [-0.70, 0.10]	0.81 [0.63, 0.92]	1

Table 6 Comparison of the deviance information criteria (DIC) and the log-pseudo marginal likelihood (LPML) values between the individual-outcome-fitting and the simultaneous-outcome-fitting approaches.

	DIC	LPML
Simultaneous-outcome-fitting	3761	-1915
Individual-outcome-fitting	3864	-1951
Difference	-103	36; 2LogPsBF=72

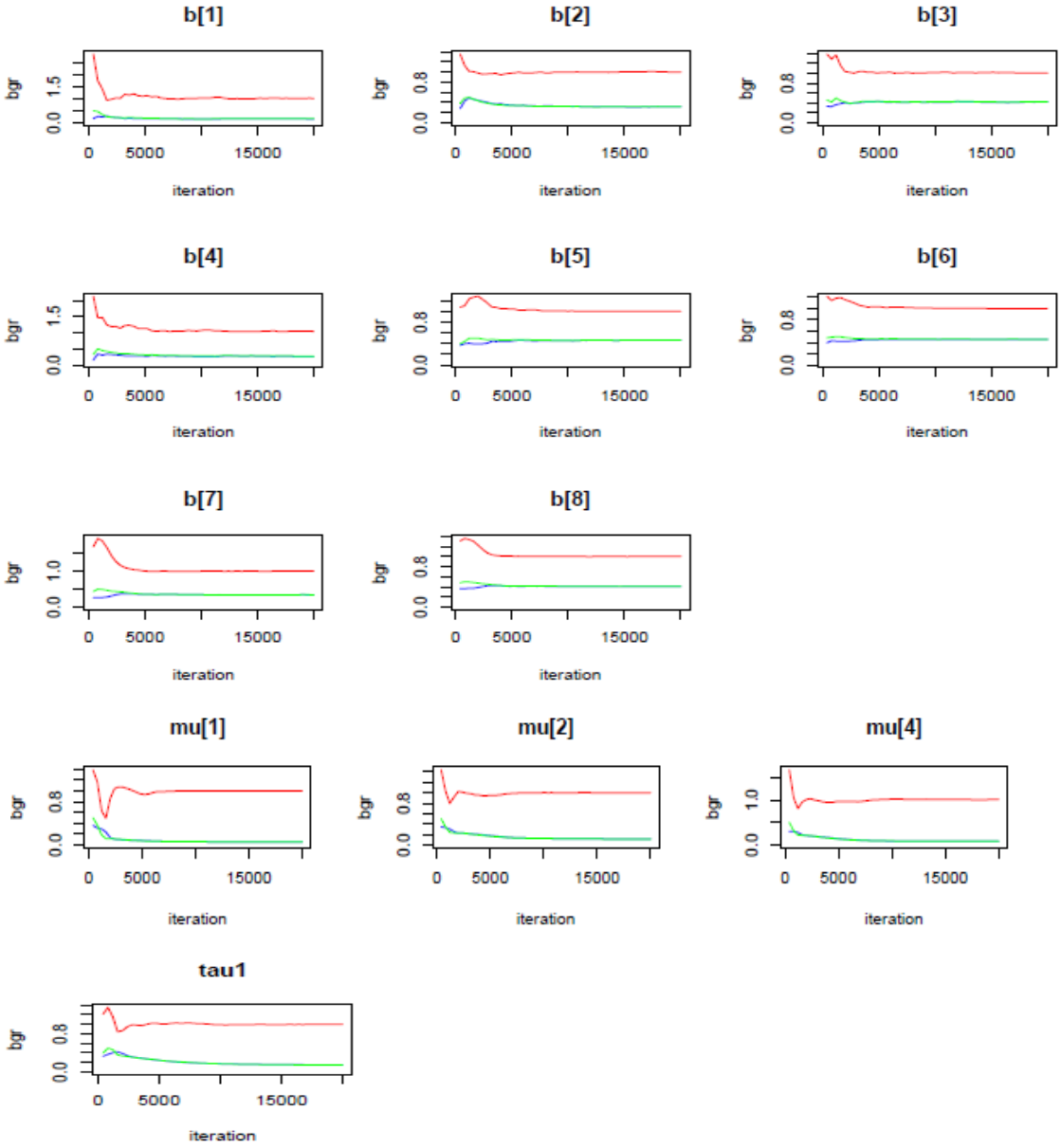

Figure 3 The plot of Brooks-Gelman-Rubin statistic for the final simultaneous-outcome-fitting model. The variables b[1] to b[8] represent $\theta 1$ to $\theta 8$ in equation (2.12)-equation(2.15). The variables mu[1] , mu[2] and mu[4] are $TVE0_1$, $TVED50_1$ and $TVED50_2$ and tau1 is the precision, given as $1/\sigma^2$ in equation (2.12) for WASO. The red line is the ratio of the between-chain 80% credible interval to the within-chain 80% credible interval, and reaches to one at convergence at around 5000 iterations. The green line represents the normalized width of the 80% credible interval for the pooled samples from the two chains. The blue line is the average of the normalized width of the 80% credible interval for the individual chains. Both, the red and the green line, converge to stability at around 5000 iterations representing convergence.

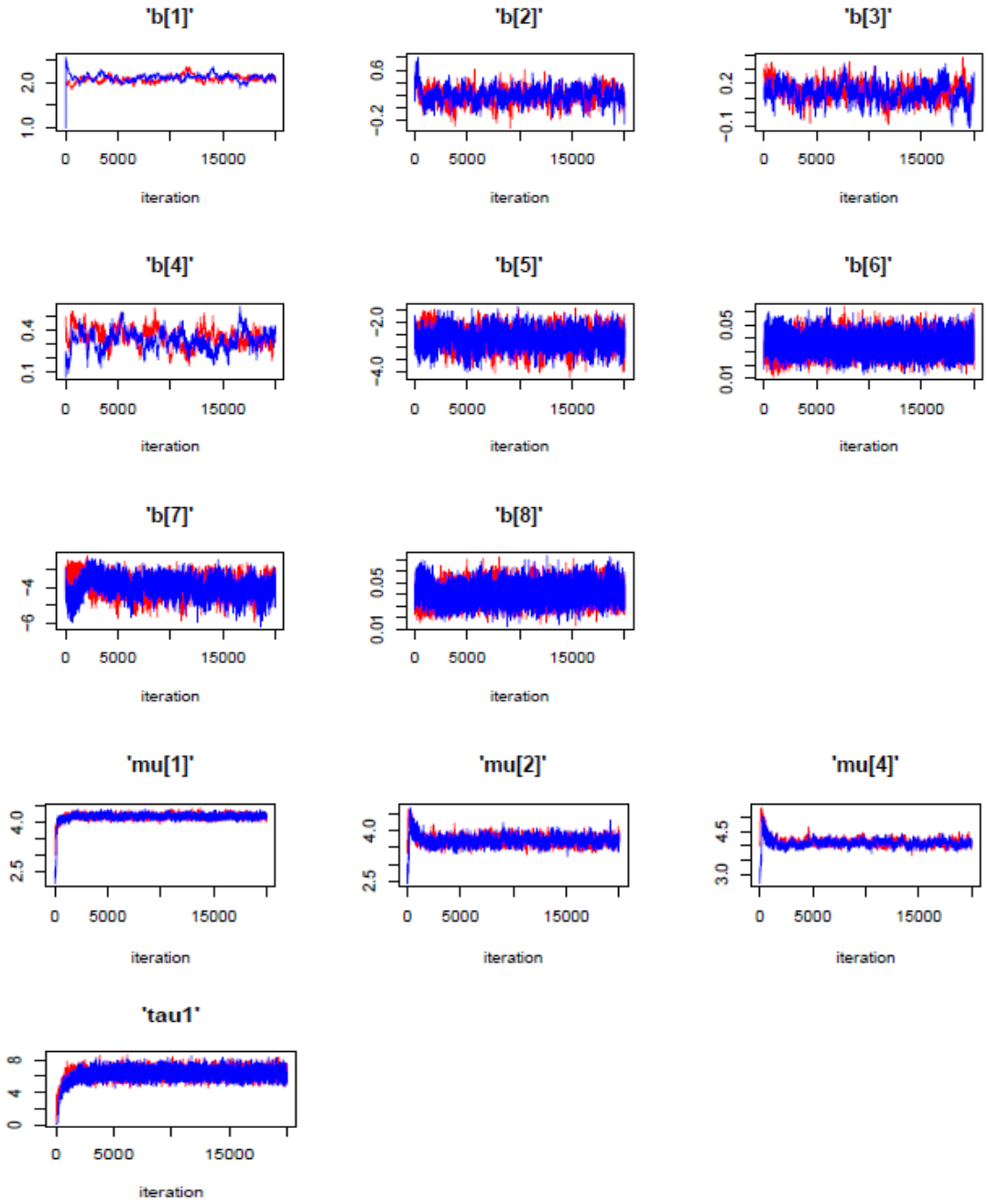

Figure 4 The trace plots for the final simultaneous-outcome-fitting model. The variables b[1] to b[8] represent $\theta 1$ to $\theta 8$ in equation (2.12)-equation(2.15). The variables mu[1] , mu[2] and mu[4] are $TVE0_1$, $TVED50_1$ and $TVED50_2$ and tau1 is the precision, given as $1/\sigma^2$ in equation (2.12) for WASO. In the trace plots for these parameters shown above, the appearance of a "fat hairy caterpillar" is seen for most of the parameter except for b[1], b[3] and b[4] that have appearance of a "wiggly snake". The appearance of a "fat hairy caterpillar" represents samples from a stationary distribution, whereas a wiggly snake represents auto-correlation between samples. Owing to auto-correlation the samples were collected up to 80000 iterations.

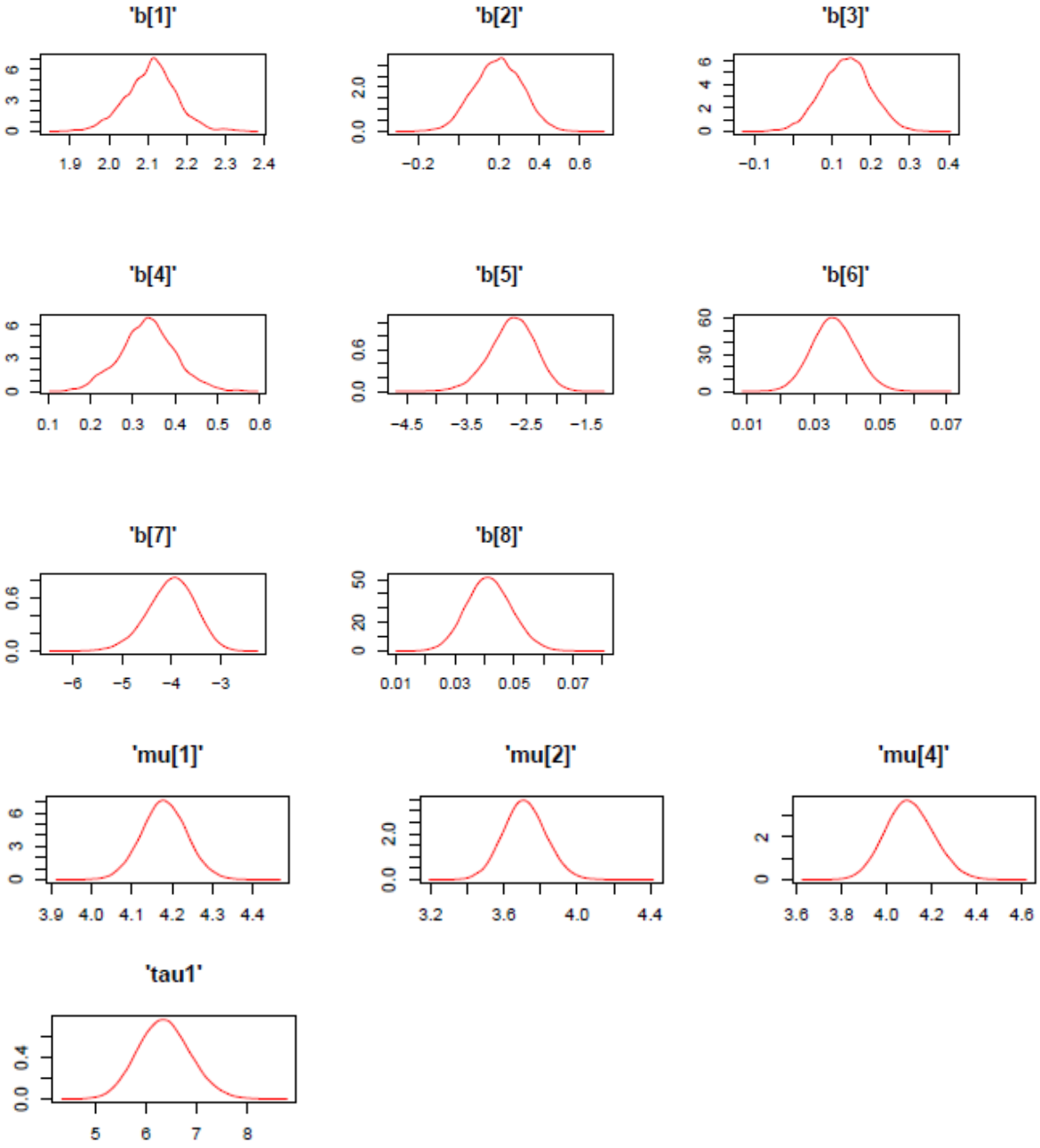

Figure 5 The posterior density plots for the final simultaneous-outcome-fitting model. The variables b[1] to b[8] represent $\theta 1$ to $\theta 8$ in equation (2.12)-equation(2.15). The variables mu[1] , mu[2] and mu[4] are $TVE0_1$, $TVED50_1$ and $TVED50_2$ and tau1 is the precision, given as $1/\sigma^2$ in equation (2.12) for WASO. In the density plots shown above, the parameters have a unimodal distribution that is symmetric around their modes.

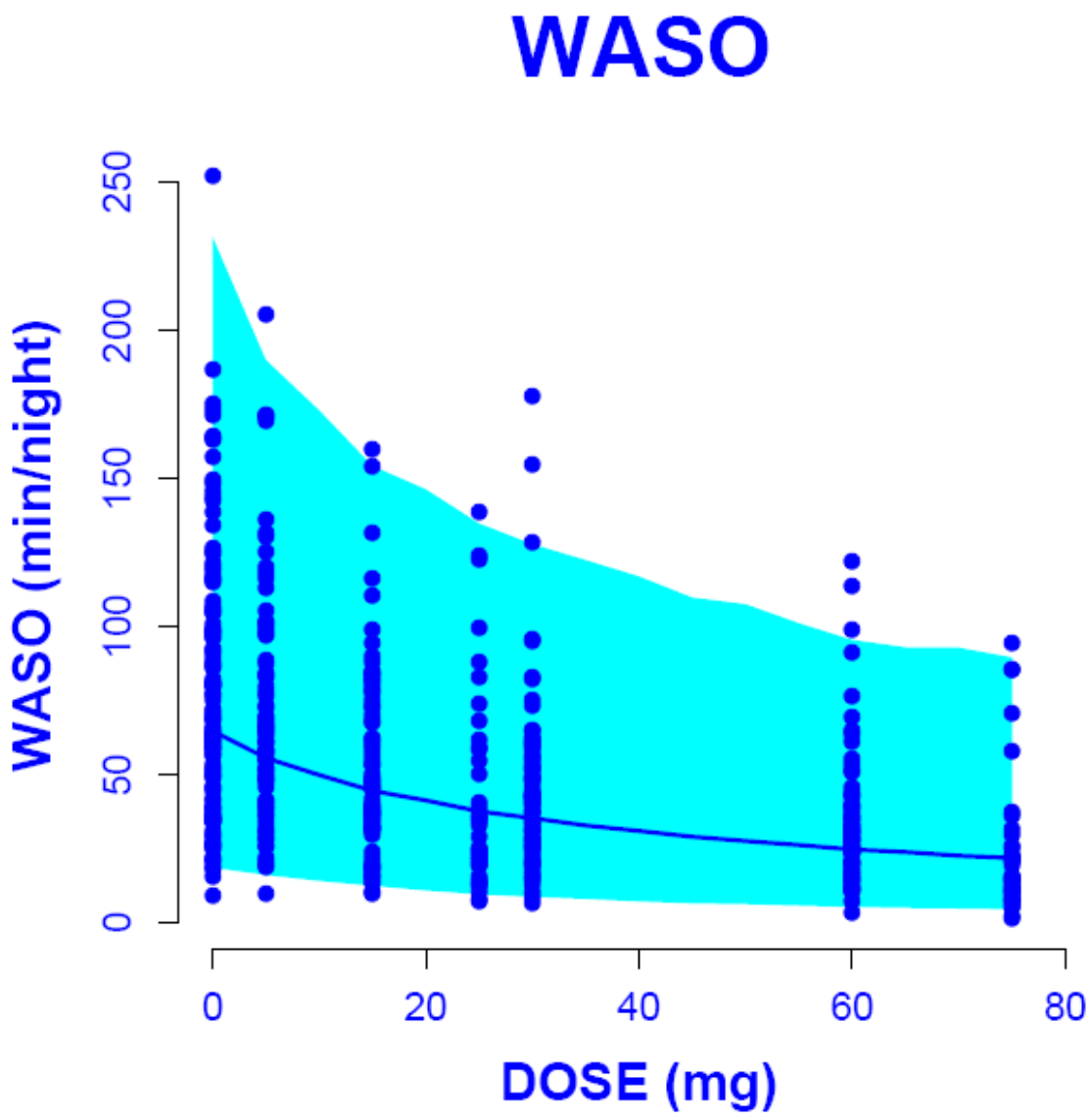

Figure 6 Visual predictive check of the model predictions (shaded interval) and observed data (dots) for the WASO dose-response model. The shaded area represents a 95% prediction interval, the solid line is the median of model predictions, observed data points are shown as dots.

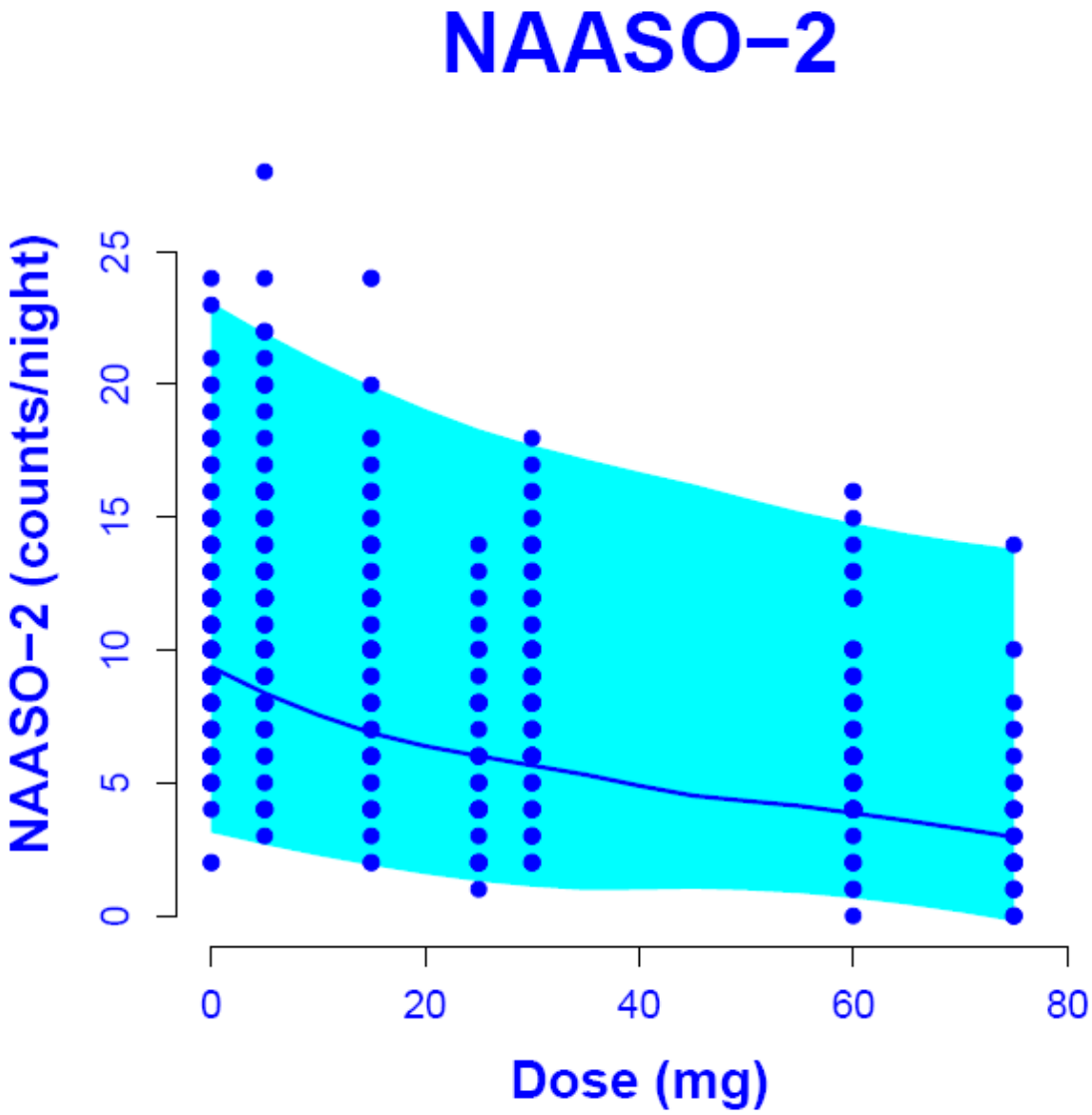

Figure 7 Visual predictive check of the model predictions (shaded interval) and observed data (dots) for the NAASO dose-response model. The shaded area represents a 95% prediction interval, the solid line is the mode of model predictions, observed data points are shown as dots.

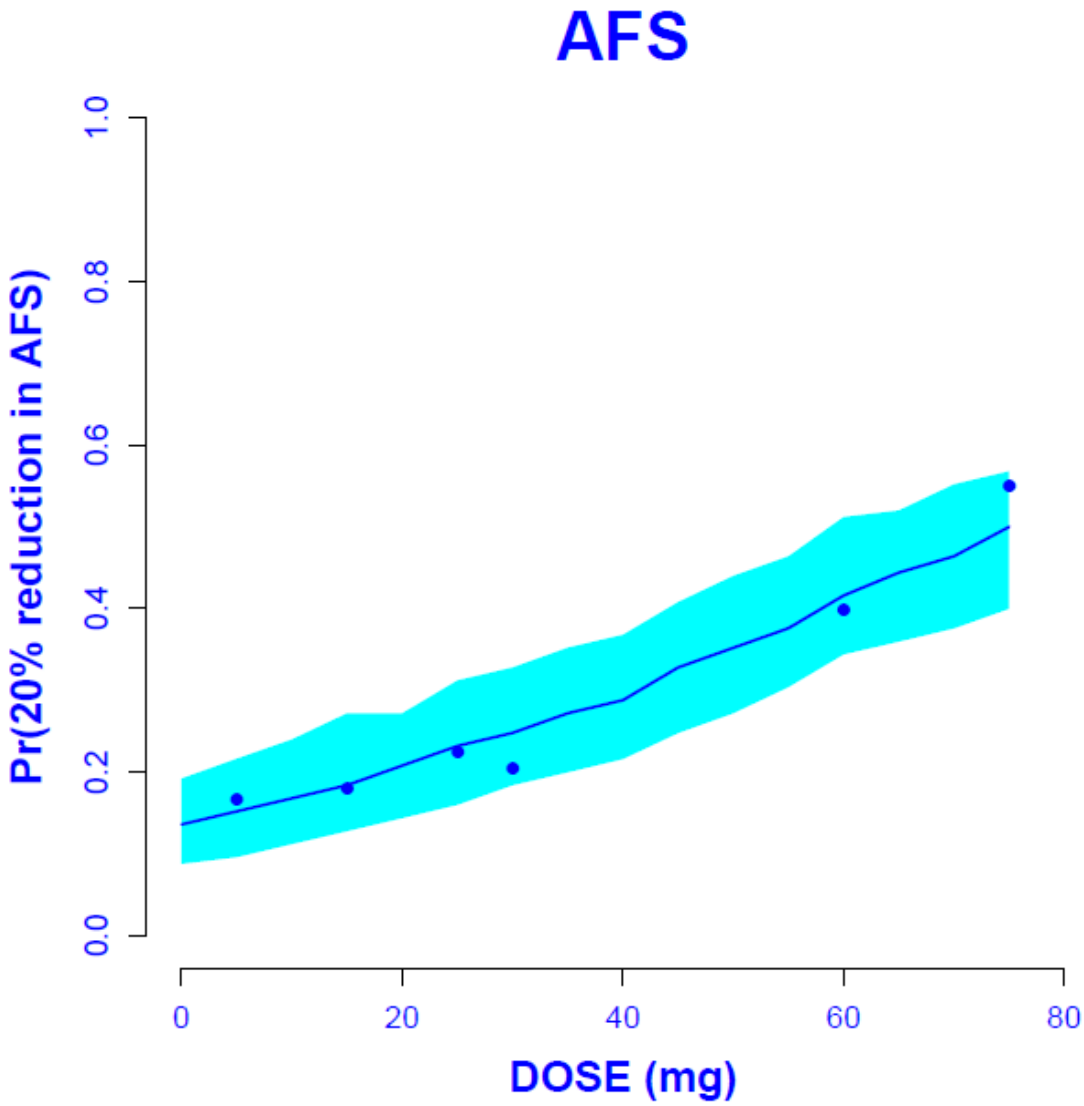

Figure 8 Visual predictive check of the model predictions (shaded interval) and observed data (dots) for the AFS dose-response model. The shaded area represents a 95% prediction interval, the solid line is the median of model predictions and observed proportions are shown as dots.

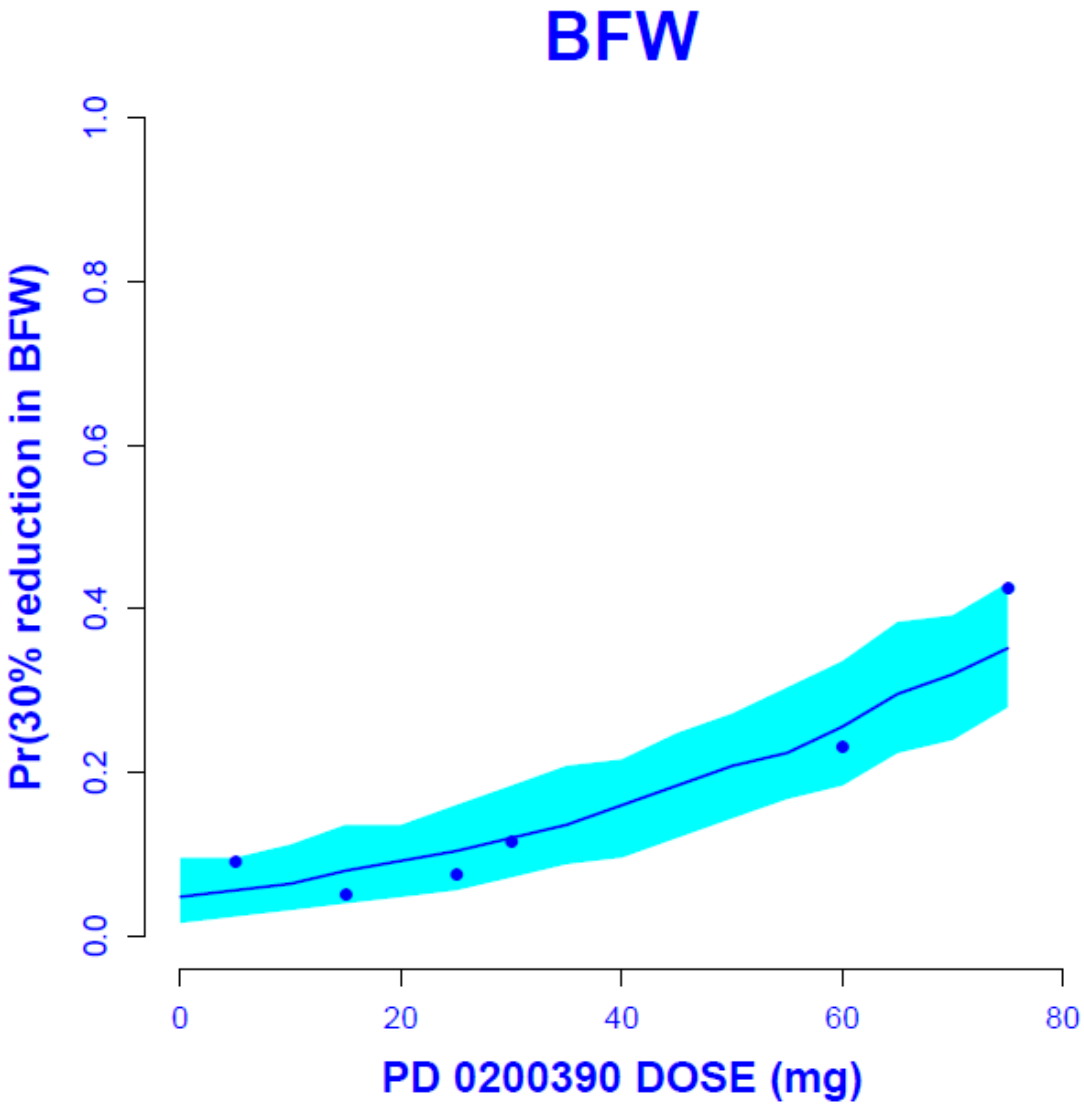

Figure 9 Visual predictive check of the model predictions (shaded interval) and observed data (dots) for the BFW dose-response model. The shaded area represents a 95% prediction interval, the solid line is the median of model predictions and observed proportions points are shown as dots.

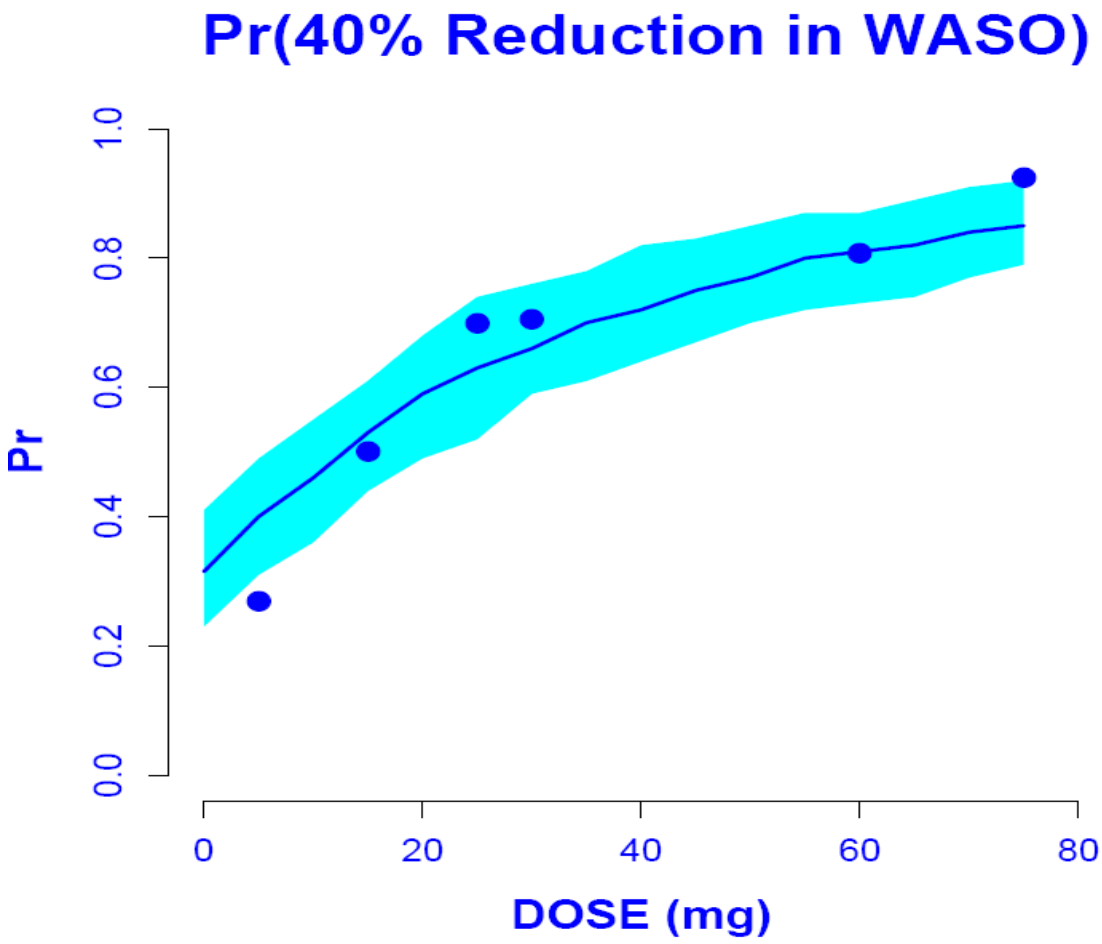

Figure 10 Posterior predictive check for the WASO dose-response model. The shaded area represents a 95% prediction interval of the simulated discrepancy statistic, the solid line is the median of the predictions, and the observed data discrepancy statistics are shown as dots.

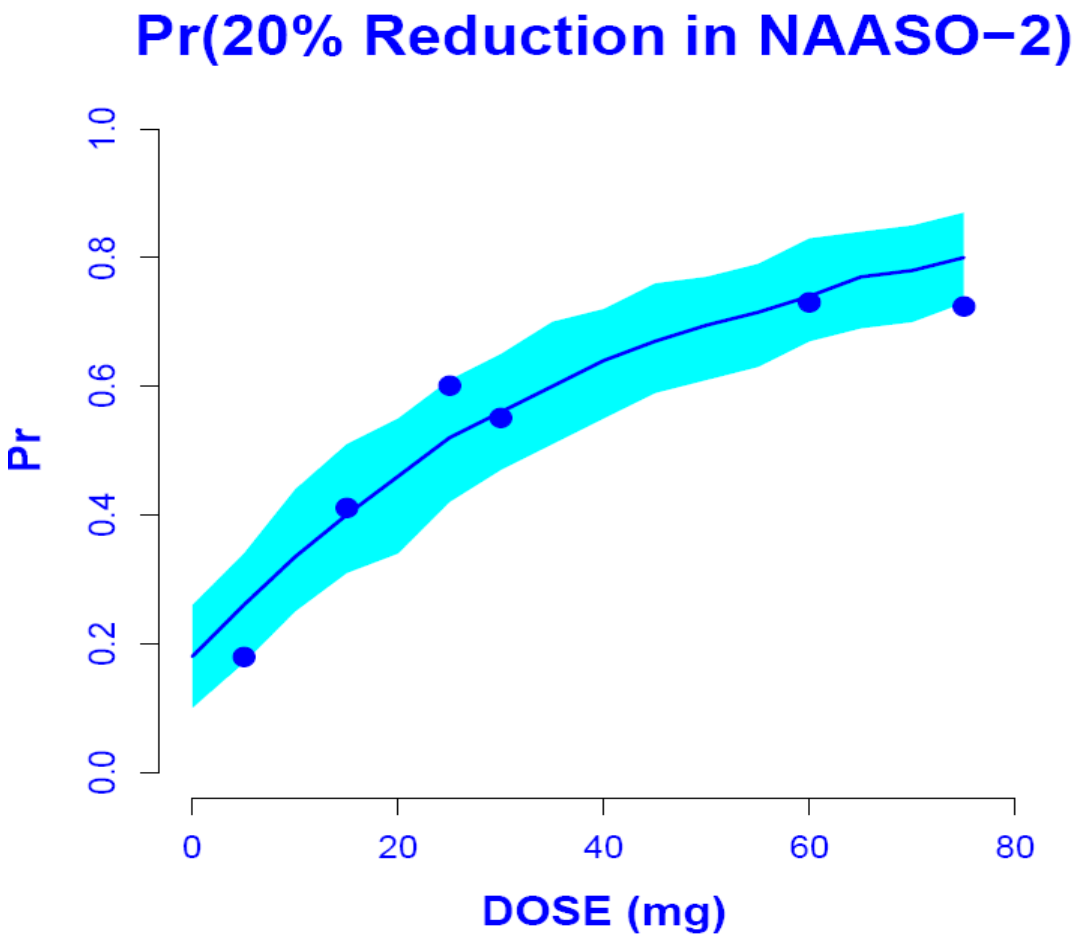

Figure 11 Posterior predictive check for the NAASO dose-response model. The shaded area represents a 95% prediction interval of the simulated discrepancy statistic, the solid line is the median of the predictions, and the observed data discrepancy statistics are shown as dots.

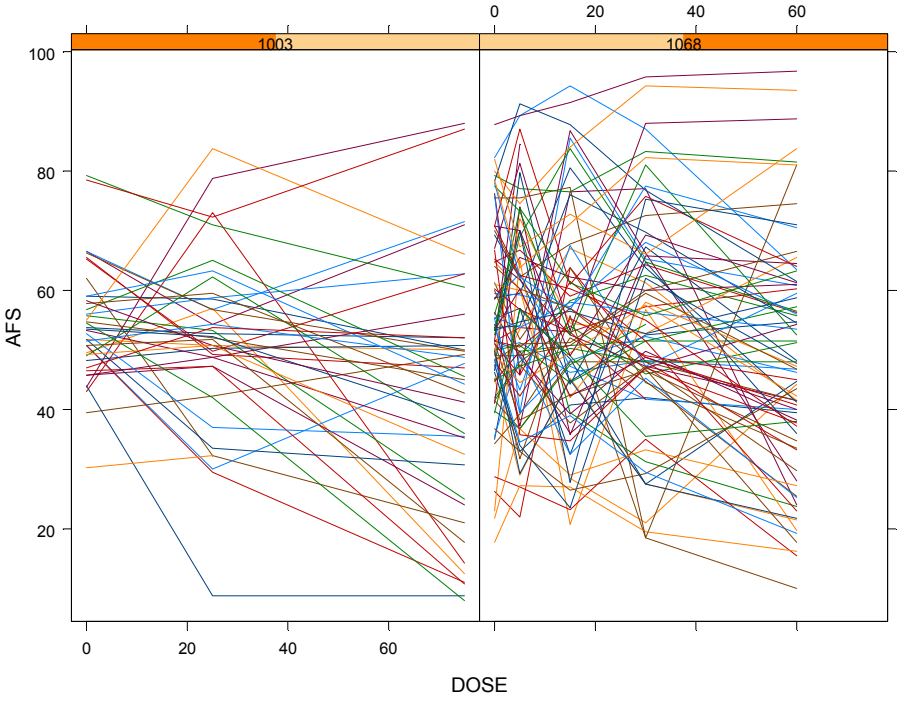

Figure 12 Spaghetti plot for the AFS, visually analog scale, scores from the two studies. Left panel shows observations from study A4251003 and right panel shows observations from study A4251068.

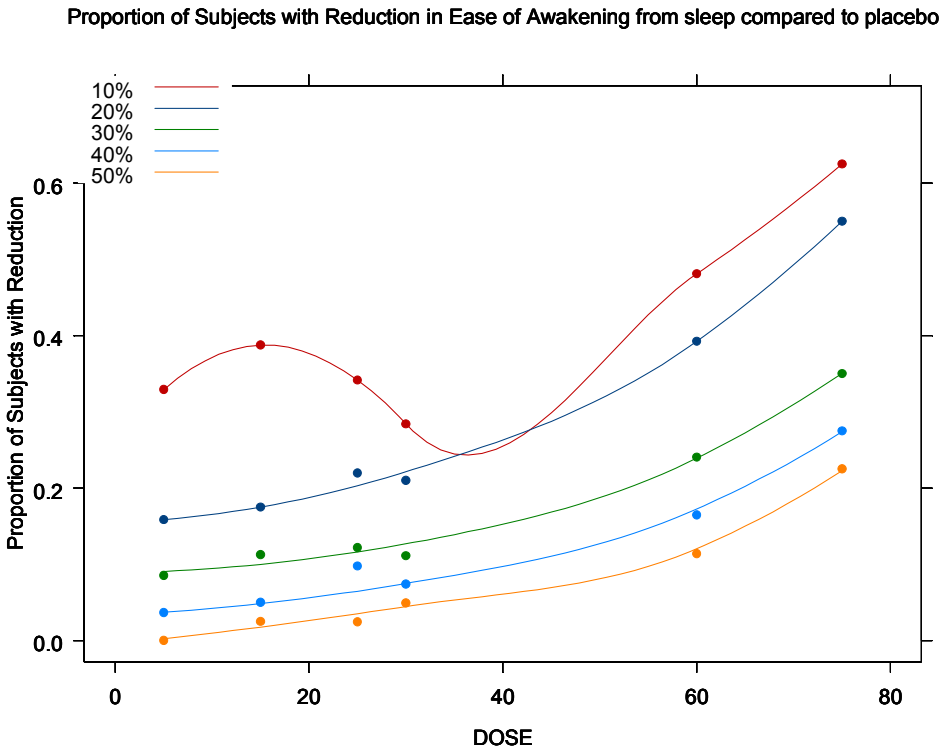

Figure 13 The result of categorizing the AFS scores in the pooled dataset obtained from two studies. Different lines represent a reduction in 10%-50% in AFS score from placebo. Top line is a 10% reduction to bottom line of 50% reduction. Fitted lines are loess through the data.

3. Dose and sample size selection for PD 0200390

For an insomnia drug to be viable in the market it should be able to induce sleep early, improve sleep consolidation, and have fewer residual effects in the morning following awakening. Finding a right dose is necessary; too high a dose may lead to residual effects and too low a dose may be less efficacious. Currently, benzodiazepines, zolpidem, eszopiclone, antidepressants and melatonin are the primary pharmacological interventions for the management of insomnia[9, 105, 106]. In the development of therapeutically similar, or "me too" drugs, the drug developer may also wish to ascertain whether or not the new molecular entity can replace the standard of care or provide a better benefit/risk profile over active comparators. Hence, it is desired that a future clinical trial show a joint outcome of substantial efficacy with minimal residual events in the trial population.

The purpose of this chapter is to apply the decision-based clinical utility described in Chapter 1.3 with the modeling results in Chapter 2 to dose finding and sample size selection for a future placebo-controlled trial for PD. The selection of dose of PD and sample size of a future trial is optimized to demonstrate efficacy and side effect profiles that are superior to an active comparator, zolpidem.

3.1. Decision-Theoretic Approach

The problem of selecting the right dose and finding the sample size can be defined in a Bayesian decision-theoretic framework. The decision-theoretic framework requires structuring of the decision problem, specification of possible alternatives/actions, the

consequence associated with each action and specification of preference among consequences.

Let, E be a set of alternatives/action space that contains all possible experiment designs, e, i.e., dose and sample size combinations and $\boldsymbol{\theta}$ define a *state of nature*, or more suitably, a vector of model parameters with all possible values of $\boldsymbol{\theta}$ contained in the set $\boldsymbol{\Theta}$. The uncertainty in the *states of nature* or in the model parameter vector $\boldsymbol{\theta}$ is represented by the posterior distribution, $\pi(\boldsymbol{\theta}|\mathrm{y})$, that was obtained through simultaneous modeling of WASO, NAASO, AFS and BFW, as described in Chapter 2. The posterior distribution of parameter, $\pi(\boldsymbol{\theta}|\mathrm{y})$, can be used to relate an experimental design, e, to the simulated outcome data y^*, as shown in equation (3.1).

$$p(\mathrm{y}^*|e.\mathrm{y}) = \int_{\Theta} p(\mathrm{y}^*|\boldsymbol{\theta}, e)p(\boldsymbol{\theta}|\mathrm{y})\ d\boldsymbol{\theta} \tag{3.1}$$

The simulated data vector y^* is the realization of future data that may be observed in a future trial. After observing the simulated data, y^*, conditioned on experimental design e and observed data y, a terminal decision d was chosen from the decision space D. The utility of making a terminal decision, d, after observing y^* following an experiment e is given as $u(d, \boldsymbol{\theta}, e, \mathrm{y}^*)$. In the present scenario, the decision space was limited to either describing the clinical trial results as success or failure. A binary utility function then evaluates this decision space. The successful trial is assigned a value of one, otherwise zero. The threshold responder rates of the efficacy and residual effects (adverse events) from an active comparator arm, zolpidem, were specified as the decision criterion in the utility function. The zero-one utility function is defined in equation (3.2),

$$u(e) = \begin{cases} 1 & P_{WASO} >= 38\% \bigcap P_{NAASO} >= 53\% \bigcap P_{AFS} <= 25\% \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

where, e is a simulated placebo controlled crossover trial in n subjects and a dose, $dose$, of PD. P_{WASO}, P_{NAASO}, P_{AFS} are the trial level responder rates for the *clinical significant events* of WASO, NAASO and AFS in the simulated trial population.

A *clinically significant event* is defined at the individual level, as a difference from placebo of <=40% for WASO, <=20% for NAASO and <=20% for AFS. The magnitude of the change from placebo that defines a clinical significant event for the endpoints of WASO, NAASO, AFS and BFW were obtained through communication with project team members at Pfizer Inc.

A placebo controlled crossover trial for PD was described as successful if after taking the drug >= 38% of subjects had >= 40% decline from their placebo value of WASO, and >= 53% of subjects had >= 20% reduction in the NAASO counts compared to placebo administration and <= 25% of subjects had an adverse event of >= 20% reduction in their ease of awakening compared to placebo. These responder rates thresholds were obtained from the data in study A4251003, where zolpidem was administered as an active control. Thus, the decision criterion evaluates the result of a trial for the joint outcomes of safety and efficacy. In this evaluation, it is assumed that a better benefit/risk profile of PD is demonstrated, if the responder rates in the trial are better than observed responder rates of zolpidem, i.e., higher responder rates for efficacy endpoints and lower responder rates for residual effects.

The optimum experimental trial design was solved as a two-part decision problem: select the best dose and secondly select the sample size. Consider that for a fixed sample size, n, the best experimental design is to choose a dose, $dose^{opt}$, that maximizes the expected value of the utility function,

$$U(dose^{opt}, n) = \max_{dose \in Dose} \int_{\mathbf{Y}^*} \max_{d \in D} \int_{\Theta} u(\mathbf{y}^*, e, \boldsymbol{\theta}, d) p(\boldsymbol{\theta}|\mathbf{y}^*) p(\mathbf{y}^*|\boldsymbol{\theta}, e) p(\boldsymbol{\theta}|\mathbf{y}) \, d\boldsymbol{\theta} \, d\mathbf{y}^*, \tag{3.3}$$

where, $Dose$, is a finite space which has been restricted to the range of doses in study A4251003 and A4251068, i.e., 0 to 75 mg. The expected utility, $U(dose^{opt}, n)$, is a monotonic function of sample size, n, and lies between 0 and 1. Increased sample size will lead to increased trial cost. The sample size in this trial design was restricted to a minimum number of subjects that gives the value of $u(\mathbf{y}^*, e, \boldsymbol{\theta}, d)$ close to 0.90. In other words, the proposed clinical trial should have a 90% success probability to demonstrate a joint event of safety and efficacy that is better than zolpidem. This was achieved by finding a minimal sample size, n^{opt}, that minimizes the absolute difference between the maximum expected utility for a particular dose, $U(dose^{opt}, n)$, and 0.90 as shown in equation (3.4).

$$U(dose^{opt}, n^{opt}) = \min_{n} |U(dose^{opt}, n) - 0.90| \tag{3.4}$$

The integral defined in equation (3.3) cannot be solved analytically; therefore a grid-based simulation was used to solve this integral.

3.2. Methodology

Software/Hardware

The simulation and post-processing of simulated data was carried out in R version 2.8.1[107]. A single Intel Xeon, Dual Processor CPU, 3GHz, 2GB RAM desktop computer was used for computation.

Simulation Plans

The grid-based search algorithm was carried out by simulating 200 single dose placebo-controlled crossover trials for a total of 160 scenarios. In each scenario, a dose was selected from 16 dose levels (0 to 75 mg at every 5 mg) and a sample size out of 10 sample sizes (50, 75, 100, 125, 150, 175, 200, 300, 400 and 500). The dose range (0-75) was limited to the range of doses that were utilized in building the dose-response models (Chapter 2). The sample sizes were arbitrarily chosen to obtain a 90% probability of trial success. Each crossover trial had two arms, a placebo arm and a dose arm, selected from the specified dose range. The placebo was also tested against placebo to monitor for false positive results.

The posterior distribution of the model parameters, obtained from the simultaneous-outcome-fitting approach (Chapter 2), was used to simulate data for the safety and efficacy variables. For a simulated subject in a placebo-controlled crossover trial, a correlated vector of outcomes was generated for the efficacy and safety variables. The efficacy variables were WASO and NAASO. The safety variable was AFS. The safety variable, BFW, was not used in the utility, since the two safety variables, AFS and BFW,

were highly correlated (mean correlation 0.81) and it was assumed that BFW provided little additional information after including AFS in the utility.

Each individual in the simulated trial was evaluated for the occurrence of a clinically significant difference from placebo for the efficacy variables. A reduction from placebo of 40% and 20% for WASO and NAASO, respectively, was considered as a clinically significant difference. If a subject had a clinically significant reduction, a value of 1 was assigned; otherwise a value of 0 was assigned. The safety variable, AFS, was modeled as a binary outcome with a value 1, if there was a clinically significant difference from placebo by >= 20%, otherwise 0. In simulation, a simulated value of 1 from the logistic model indicated that there was an adverse event.

The responder rates for the efficacy and safety variables were the proportion of successes (or ones) in each simulated trial. The responder rate in each simulated trial was evaluated by a utility function defined in equation (3.2). The simulation algorithm is given below.

Select a simulation scenario, e, by selecting the first dose level j from a set of dose levels $\{0, 5, 10, \ldots, 70, 75 \text{ mg}\}$ and a sample size n from a set of sample sizes $\{50, 75, 100, 125, 150, 175, 200, 300, 400, 500\}$.

Sample a parameter vector $\boldsymbol{\theta}^*$ from the joint posterior distribution of the parameters obtained in the simultaneous-outcome-fitting approach (Chapter 2).

Sample with replacement a covariate vector $\mathbf{z}_i$, from the original dataset, for the i^{th} individual in the simulated trial of n subjects

1. Simulate a correlated vector of observation,

   $\mathbf{y}_i^* = \{y_{kij}^*, \ldots, y_{3ij}^*, y_{kiplacebo}^*, \ldots, y_{3iplacebo}^*\}$, for WASO, NAASO and AFS, in

the i^{th} individual receiving placebo and the j^{th} dose (Step 1) using the parameter vector (step 2), the covariate vector (step 3) and the dose-response model for the respective endpoints (Chapter 2).

2. Calculate the change from placebo for each metric in each subject. The CSD for WASO and NAASO measures were a clinically significant reduction from placebo values by 40% and 20%, respectively. The variable AFS is modeled as a binary variable with occurrence of CSD simulated as 1. Define t^*_{kij} a value of 1 if there is a CSD, else 0. The CSD for WASO and NAASO is computed as follows:

let $\mathrm{t}^*_{kij} = \begin{cases} 1 & \left\{ \dfrac{y^*_{kij} - y^*_{ki(placebo)}}{y^*_{ki(placebo)}} \right\} \geq CSD_k \\ 0 & \text{otherwise} \end{cases}$, where $k = \{1, 2\}$, CSD_1 is 0.4

and CSD_2 is 0.2 for WASO and NAASO, respectively, $i = \{1, \ldots, n\}$ are the number of subjects in the simulated trial. The new variable t^*_{kij} defines the occurrence of a clinically significant event in the i^{th}individual after taking j^{th} dose of PD for the k^{th} response. For the safety variable, AFS, $t^*_{3ij/placebo}$ is the simulated data $y_{3ij/placebo}$.

Do step 5 for $i = \{1, \ldots, n\}$ subjects in the trial taking dose placebo and a dose level j.

Estimate trial-level responder rate, P_{WASO}, P_{NAASO} and P_{AFS} as

$$P_k = \frac{1}{n} \sum_{i=1}^{n} t^*_{kij}, \tag{3.5}$$

where P_k ($k = \{1, \ldots, 3\}$ is for WASO, NAASO and AFS, respectively) is the proportion of subjects in the trial with CSD for the k^{th} response.

Assign $u(e)$ the value of 1 or 0 using the following function

$$u(e) = \begin{cases} 1 & P_{WASO} >= 38\% \bigcap P_{NAASO} >= 53\% \bigcap P_{AFS} <= 25\% \\ 0 & \text{otherwise} \end{cases}$$

Repeat steps 2-8 for 200 simulations for each scenario of selected dose and the sample size n.

Compute $\frac{1}{200} \sum u(e)$ to get $U(dose, n)$, the probability of success for an experimental design, e, a placebo controlled crossover trial with n subjects and a dose, $dose$.

Select another experiment design e from the 160 possible combinations.

For a selected sample size, n, find the *optimal* dose that has the maximum value of the expected utility

$$U(dose^{opt}, n) = \max_{dose \in Dose} U(dose, n)$$

Find the *optimal* sample size that minimizes the absolute difference between 0.9 and $U(dose^{opt}, n)$.

$$U(dose^{opt}, n^{opt}) = \min_{n} |U(dose^{opt}, n) - 0.9|$$

The selected dose, $dose^{opt}$, and sample size, n^{opt}, are the solution the decision-theoretic design problem.

Sensitivity Analysis

A sensitivity analysis was done to assess the effect of changes in the decision attributes to the selection of dose and sample size. The responder-rate thresholds were varied relative to the observed responder rates (RR) in the zolpidem treated arm (ZR). The sensitivity analysis was performed by varying the responder rate thresholds, ZR + 10%, for each criterion, WASO, NAASO and AFS, while keeping the other fixed at the observed ZR

(Table 7). The changes in decision criteria will allow for identifying the relative effect of each component to changes in dose and sample size. In the algorithm described in the previous section, the sensitivity analysis relates to changing the attributes of the decision criteria in the utility function described in step 8. The values of decision attributes are given in Table 7.

Table 7 Observed responder rates for zolpidem (ZR), 90% and 110% of ZR computed from the data in study A4251003.

	WASO	NAASO	AFS
ZR	0.38	0.53	0.25
110% ZR	0.42	0.58	0.28
90% ZR	0.34	0.48	0.23

3.3. Results

The simulation results for the 160 scenarios are summarized as the proportion of the 200 simulated trials that were classified as successful, and are shown in Table 8. The utility curve is maximized at a dose of 20 mg for a sample size of 300, shown in Figure 14. In a placebo-controlled crossover trial, a 20 mg dose of PD with a sample size of 300 per arm has a 90% probability to have a joint outcome of safety and efficacy, where more than 38%, 53% and less than 25% of subjects will have a clinically significant event of

WASO, NAASO and AFS, respectively. A clinically significant event is defined as a 40%, 20% and 20% decrease from baseline for WASO, NAASO and AFS, respectively.

The decision criterion is a joint event of safety and efficacy. A drug is not successful at lower doses because of suboptimal efficacy and at higher doses the drug does not meet the criteria of safety. The false positive rate (placebo against placebo) for placebo group was 0 in 200 simulations (Table 8).

The changes in dose and sample size selection with varying responder rates for WASO, NAASO and AFS are shown in Figure 15, Figure 16 and Figure 17, respectively. The responder threshold for WASO was set at 0.34, 0.38 and 0.42 corresponding to 90%, 100% and 110% of zolpidem responder rate threshold (Table 7). A decrease in the efficacy threshold (90% ZR) does not change the dose or sample size to achieve 90% success (Figure 15). An increase in the efficacy threshold for WASO, 110% ZR, does not change the dose of 20 mg, however the sample size increases to 400 per arm (Figure 15).

The responder rate threshold for NAASO was set at 0.48, 0.53 and 0.58 corresponding to 90%, 100% and 110% of ZR (Table 7). A reduction in the efficacy threshold to 0.48 RR (90% ZR) did not decrease the dose (Figure 16), however sample size was reduced to 175 per arm to achieve a 90% success probability. An increase in the threshold to 0.58 RR (110% ZR) caused an increase in dose and sample size to 30 mg and 400 per arm (Figure 16).

AFS is a safety endpoint and limits the selection of higher doses. The responder rate threshold for AFS was set at 0.23, 0.25 and 0.28 corresponding to 90%, 100% and 110%

ZR (Table 7). As the safety threshold is made stringent, 90% ZR, the dose still remains the same (20 mg) however sample size increases to 500 per arm (Figure 17). On the contrary, when as much as 28% RR (110% ZR) can be tolerated in a trial, this relaxation of safety threshold leads to lowering of sample size to 200 per arm and increase in dose to 20-25 mg (Figure 17).

The impact of changes in the probability of trial success with respect to increasing sample size is shown in Figure 18. As shown in the figure, increasing or decreasing the daily dose, compared to a 20 mg dose, leads to reduction in the probability of success with the increase in sample size.

3.4. Discussion

A Bayesian decision-theoretic approach was used to select a dose and sample size for a future confirmatory trial of PD. A placebo controlled trial with a dose of 20 mg and a sample size of 300 per arm had 90% probability of demonstrating not just efficacy but a better benefit/risk profile compared to zolpidem.

The use of a sensitivity analysis was done to objectively evaluate changes in selection of dose sample size with changes in responder rates in the decision criteria. The joint utility restricts selection of lower doses because of the threshold responder rates for efficacy and limits selection of high doses because of an increase in the incidence of adverse events. Reducing the efficacy threshold of a single attribute (90%ZR WASO or 90%ZR NAASO) did not result in lowering of the dose, but did lead to the selection of a lower sample size. A lower dose was not selected because of the joint nature of the utility,

whereby the responder rate thresholds for WASO and NAASO must be met to yield a positive result.

The joint utility approach is different from a typical approach to sample size estimation, where a sample size is chosen to provide sufficient power with a usual type I error rate for a primary endpoint and may not necessarily power for other endpoints. The utility had only one residual measure, AFS, and it had the most impact on dose and sample size selection. Increasing the safety threshold for AFS responder rates from 0.25 to 0.28 decreased the sample size by 100 subjects and the utility curve shifter towards higher doses (Figure 17). For an insomnia compound it may not be desirable to develop a drug that has a higher residual event rate compared to an active comparator.

Several authors have used and described multi-attribute utility and decision theoretic concepts to dose/dosage regimen selection and drug development decisions[82-85]. Poland and Wada combined drug-disease, economic, pharmacokinetic and pharmacodynamic model of a HIV protease inhibitor to guide drug development decision of once daily versus twice daily dosing regimen for experienced or treatment naïve subjects[82]. Ouellet et al. combined safety and efficacy attribute as a clinical utility index to select between pharmacological similar compounds for insomnia[84]. Graham et al. combined adverse events and efficacy data in a Bayesian decision theoretic framework to choose between immediate versus controlled release formulation of oxybutynin for urinary urge incontinence[85]. Jonsson and Karlsson used a utility approach considering both beneficial and adverse effects of oxybutynin for individualizing a dosing regimen[86].

In most of these multi attribute utility functions, a utility is defined as a weighted sum of safety and efficacy attributes given as $U = wt_E \cdot U_E - wt_T \cdot U_T$, where wt_E and wt_T are the weights assigned to efficacy and toxicity endpoints and U_E and U_T are the utility functions for the efficacy and toxicity measures. A fundamental assumption of the weighted utility is that toxicity and efficacy outcomes are additive and the resulting sum represents the net therapeutic value or a clinical utility. Although weighted multi-attribute utility provides a framework for reducing the dimension of the problem, an individual or a regulatory agency may not see the net benefit as a sum of efficacy and toxicity. A regulatory agency evaluates whether the drug provides significant efficacy conditional on the incidence and severity of side effects and a consumer may choose between two drugs based on the odds of having efficacy to adverse outcomes. Both scenarios represent a joint event of safe and efficacious outcomes as defined in a decision-theoretic utility function. In the decision theoretic approach presented here, a drug must satisfy a joint event of both safe and efficacious outcome to be successful. This characteristic is even more desirable when the investigational compound is a "me-too" drug and must show a better efficacy and safety profile in order to capture market share.

An effective benefit-risk assessment should allow for the setting of preferences between endpoints. The occurrence of mortality or organ impairment is to be avoided at any cost whereas nausea or somnolence may be tolerated. Although it depends upon factors like disease state, individual preference and other factors, it is desirable to adjust preference for the endpoints. A responder rate of clinical events of an active comparator was used in this analysis to establish preference between endpoints. It is believed that responder rate

assessment can provide an objective assessment to preference. Responder rates from an active comparator can provide anchors to experts in making objective recommendations. Further, responder rates or probabilities of a clinical outcome can be easily converted to statistics like "number needed to treat/harm" or odds ratio that are familiar terminologies and can help in group decision making. In the decision theoretic approach, zolpidem responder rates were used as an anchor to provide objective weighting assessment. Zolpidem is widely prescribed in treatment of insomnia and provides a benchmark for compound selection and development for new chemical entities for insomnia. Using zolpidem RR decision criteria led to selecting a 20 mg dose. It is interesting to note that in a previously published clinical utility analysis of PD, the utility was maximized at a dose of 30 mg[84]. In this analysis, the weights were obtained by interviewing 581 physicians using a hybrid conjoint analysis[84]. It is important to note that Ouellet *et al*. used seven endpoints instead of three in this study, and the weights were expert's subjective assessment vs. responder rates for an active comparator as used in this methodology[84].

This analysis used two efficacy and one adverse event variable in the utility function. Clearly, they do not represent all the endpoints required to make a decision for drug's viability for an insomnia compound. However, the approach can be extended to any number of safety and efficacy endpoints. This analysis uses prior data obtained from two dose studies carried out for two nights per treatment period. A more relevant decision would be made if data were available from multiple dose studies which will help

incorporate tolerability and long-term adverse-effect measures into evaluating viability of PD.

Table 8 The proportion of 200 simulated trials that were classified as successful are shown as a function of dose (mg) and sample size (n). The utility function that is maximized is defined section 3.2. The grayed boxes are scenarios where the probability of a successful trial exceeds the required criterion of 0.90.

Dose / n	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75
50	0	0	0.06	0.275	0.545	0.500	0.510	0.345	0.225	0.070	0.045	0.045	0.010	0	0.005	0
75	0	0	0.045	0.315	0.585	0.610	0.455	0.335	0.175	0.065	0.030	0.015	0	0	0	0
100	0	0	0.025	0.315	0.725	0.725	0.515	0.300	0.165	0.055	0.005	0.005	0	0	0	0
125	0	0	0.020	0.350	0.755	0.735	0.460	0.215	0.100	0.030	0	0	0	0	0	0
150	0	0	0.010	0.325	0.775	0.750	0.510	0.250	0.090	0.035	0	0	0	0	0	0
175	0	0	0.015	0.350	0.810	0.700	0.435	0.180	0.040	0.005	0	0	0	0	0	0
200	0	0	0.010	0.405	0.865	0.765	0.550	0.225	0.060	0.010	0	0	0	0	0	0
300	0	0	0	0.390	0.915	0.805	0.490	0.170	0.020	0	0	0	0	0	0	0
400	0	0	0	0.420	0.945	0.900	0.480	0.105	0.010	0	0	0	0	0	0	0
500	0	0	0	0.420	0.975	0.855	0.460	0.130	0.020	0	0	0	0	0	0	0

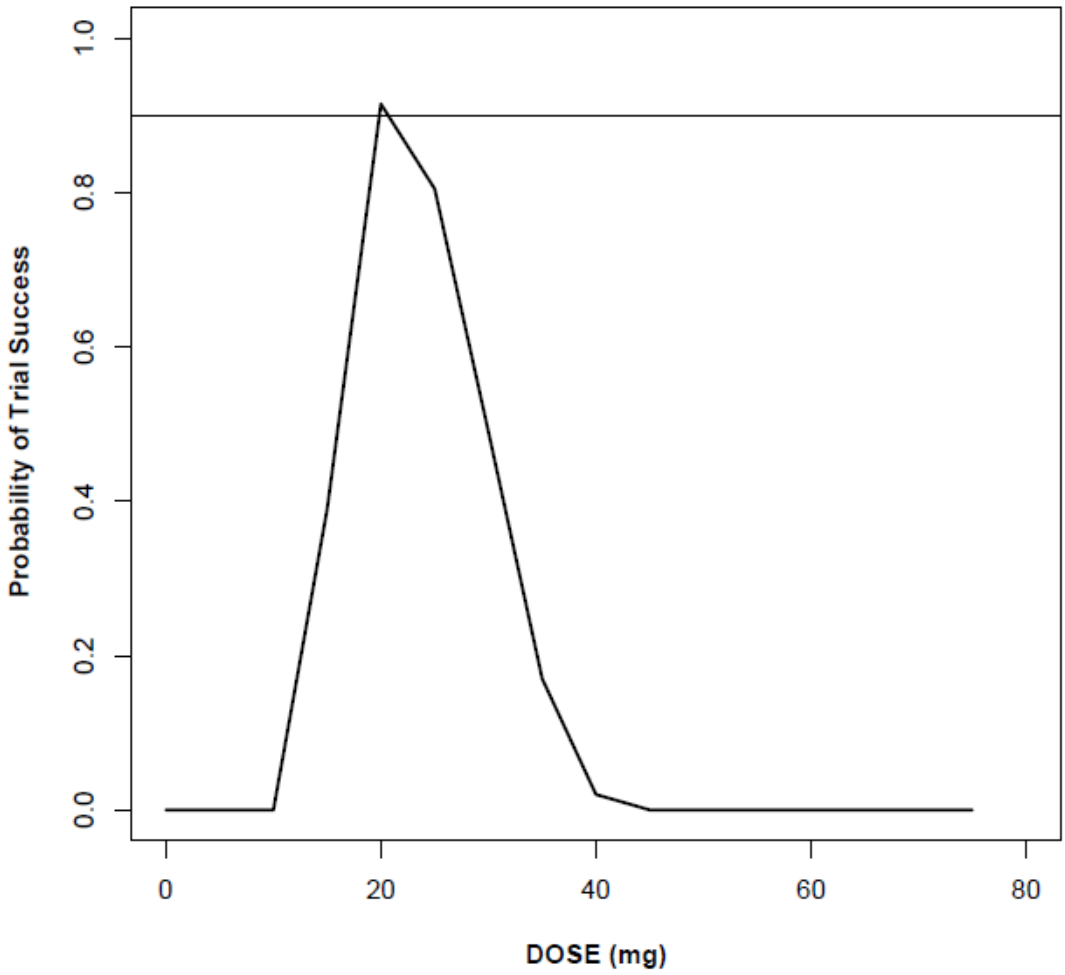

Figure 14 The plot of the probability of trial success vs. dose. The utility function reaches a maximum of 0.91 at a dose of 20 mg of PD with a sample size of 300.

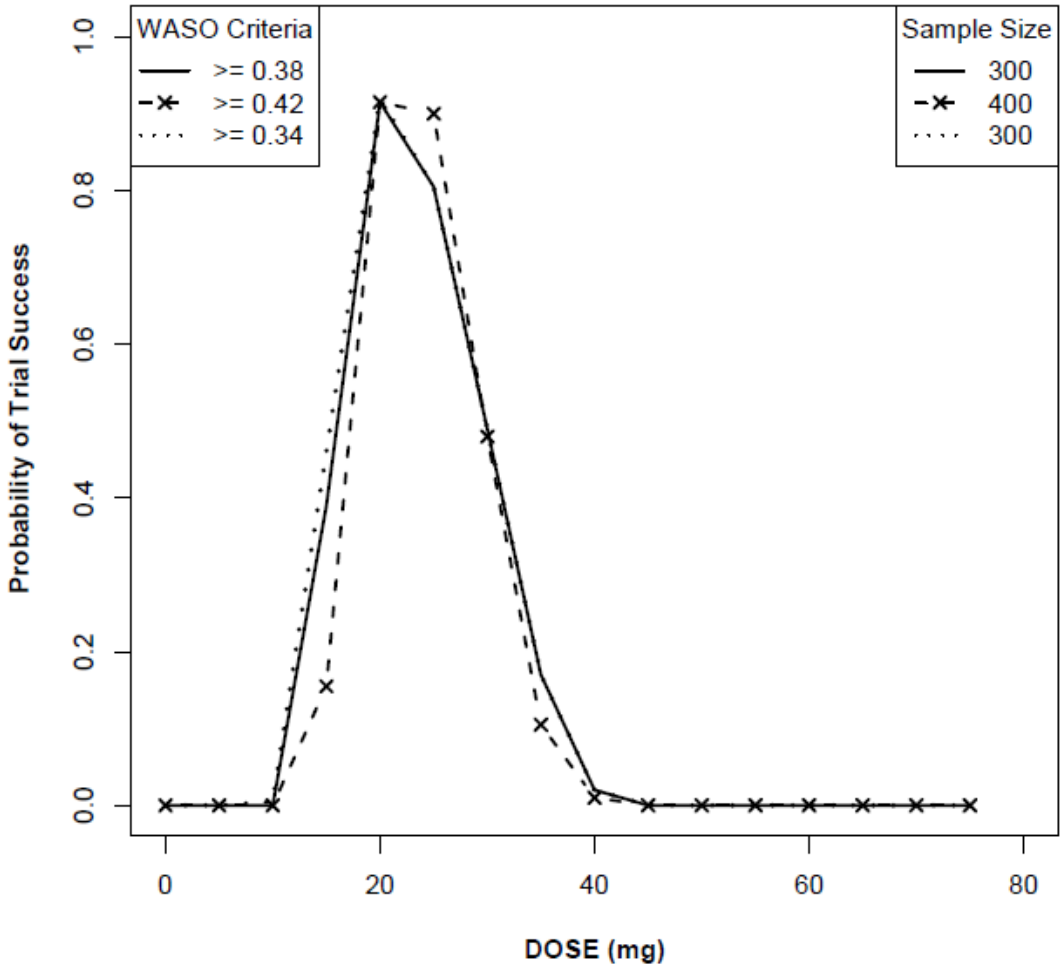

Figure 15 Sensitivity of probability of success with changes in responder rate thresholds for WASO. The decision-criteria thresholds of 0.38, 0.42 and 0.34 correspond to 100%, 110% and 90% of the zolpoidem's WASO responder rate thresholds, respectively. The corresponding sample size that provides a 90% success probability are 300, 400 and 300 for the decision threshold of 100%, 110% and 90% of the zolpoidem's WASO responder rates, respectively.

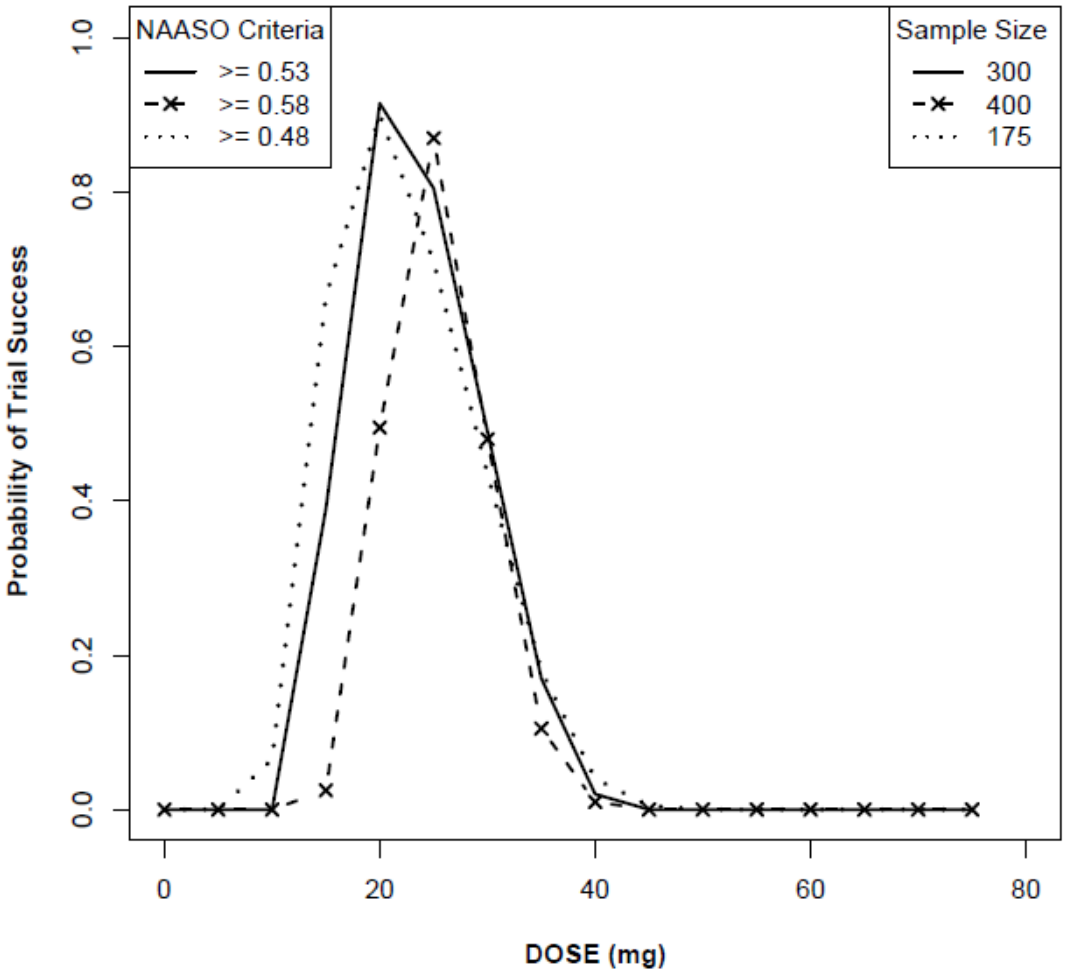

Figure 16 Sensitivity of probability of success with changes in responder rate thresholds for NAASO. The decision-criteria thresholds of 0.53, 0.58 and 0.48 correspond to 100%, 110% and 90% of the zolpoidem's NAASO responder rate thresholds, respectively. The corresponding sample size that provides a 90% success probability are 300, 400 and 175 for the decision threshold of 100%, 110% and 90% of the zolpoidem's NAASO responder rates, respectively.

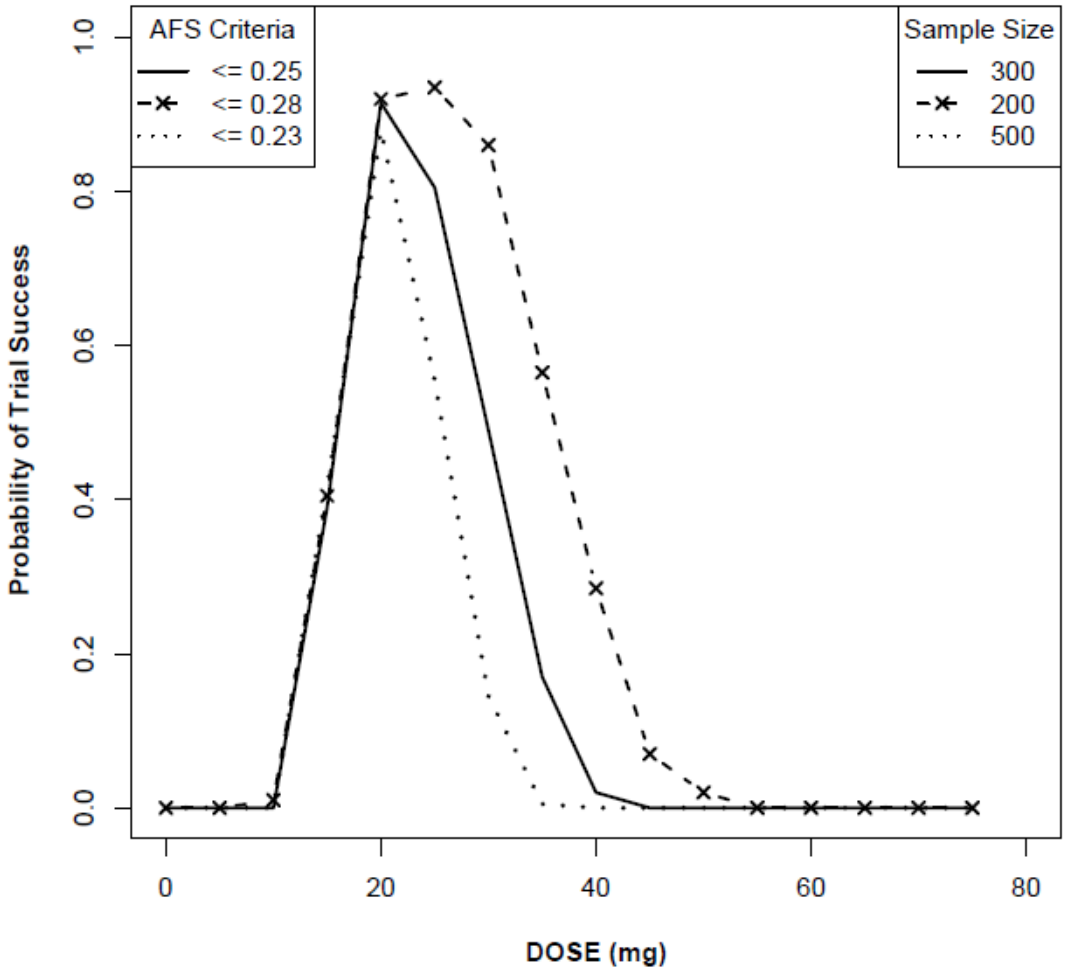

Figure 17 Sensitivity of probability of success with changes in responder rate thresholds for AFS. The decision-criteria thresholds of 0.25, 0.28 and 0.23 correspond to 100%, 110% and 90% of the zolpoidem's AFS responder rate thresholds, respectively. The corresponding sample size that provides a 90% success probability are 300, 200 and 500 for the decision threshold of 100%, 110% and 90% of the zolpoidem's AFS responder rates, respectively.

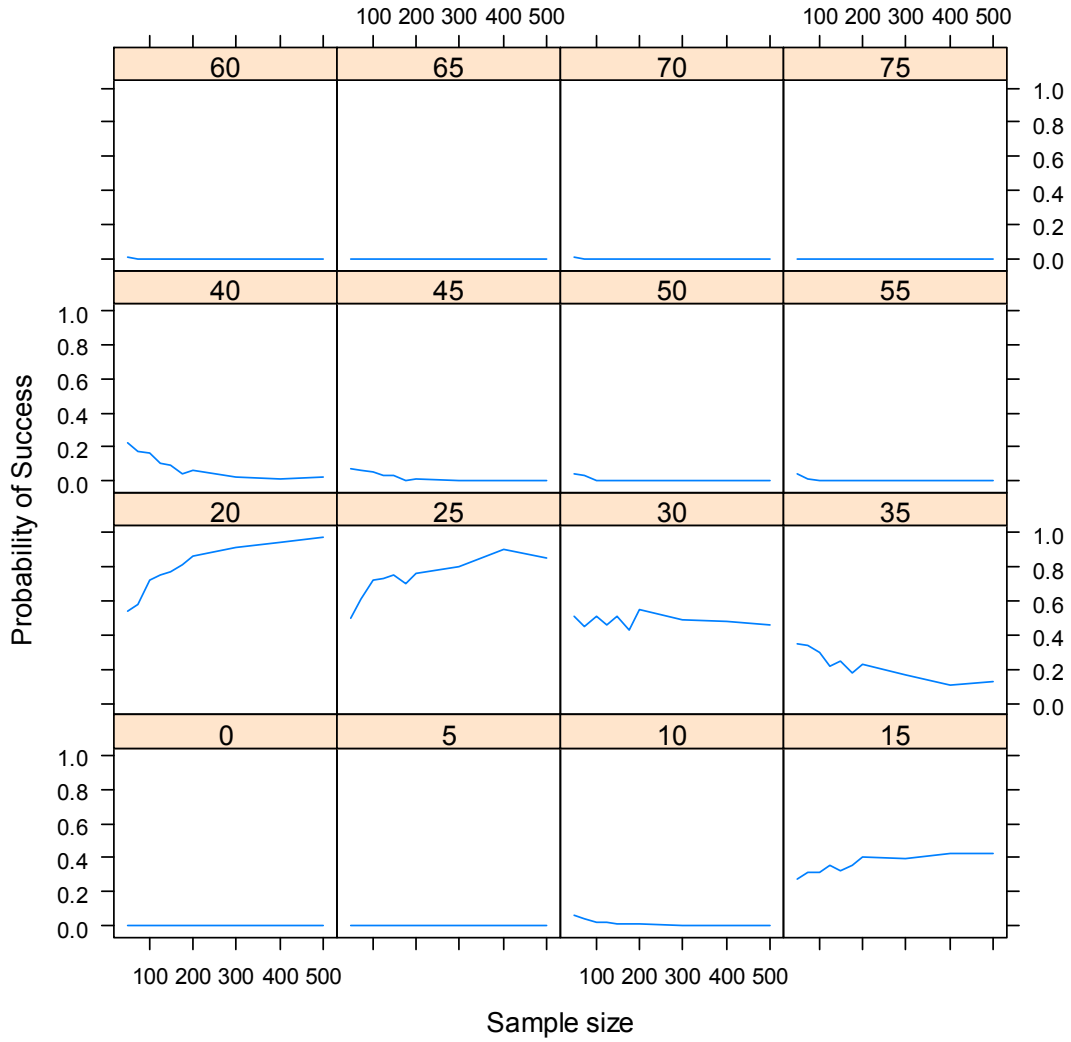

Figure 18 The plot of probability of trial success as sample size increases for different dose levels in the simulation study.

4. Future Directions

This thesis provides a conceptual framework for utilizing Bayesian modeling and decision theoretic approaches in drug development. These approaches are applied to dose-response modeling and clinical trial design for PD, an investigational compound for insomnia. Clearly, this scale is limited compared to a full scale drug development, where sequential decisions are made over multiple studies. Some key factors, that may improve decision making are highlighted below.

The pharmacokinetics of a drug is an important feature that links the administered dose to some surrogate of exposure at the site of action. The use of pharmacokinetic modeling to predict exposure may increase predictability and explain a portion of the variability in response. The historical knowledge about clinical pharmacokinetics of PD can be incorporated as prior information in the model to predict exposure.

The endpoints used in this modeling and decision analysis do not represent all relevant endpoints for treatment for insomnia. The subjective assessments of sleep are also of interest in drug approval for insomnia drugs. Incorporation of these endpoints will provide key information about the viability of PD as a treatment regimen. The incorporation of long-term dosing effects on the development of tolerance and side-effects is another other important consideration.

An important component of decision theory is to understand how people make decisions, and develop ways to improve decision making. In this analysis, an objective approach

was employed by using data from an active comparator arm in the trial. The active comparator data is believed to provide relevant starting values to the decision makers and help in group decision making. The use of experts to elicit utility functions and prior distributions has become increasingly important. The methods currently used for general and application-specific elicitation are reviewed by Kadane[108]. Although subjective utilities were not available and elicited in this analysis, they are an important next step in an iterative dialogue aimed at finding the optimum decision.

To find the optimal dose and sample size, a grid-based simulation approach over 160 arbitrarily selected design scenarios was used. Monte Carlo integration methods have been proposed to find the optimum decision at the maximum expected utility and use of this approach will likely allow the computation of utility to occur more rapidly[109].

A key assumption in the experimental design is that there is suitable information about the states of nature to obtain realization of future data through simulation. This assumption is made to design an experiment to make a decision on the viability of PD. The design alternatives, utility, and decision space are different if the objective of the experimentation were to gain knowledge about the states of the nature. The design of experiments using information-theory to gain knowledge about the states of nature is reviewed by Lindley[110].

Another important step in evaluating the viability of PD is to use cost-benefit analysis. Cost-benefit analysis takes into account the achieved health benefits and the cost of intervention. Multiple pharmacological interventions and generic and brand products for

insomnia are available in the market. The possibility that the cost of PD will be covered by insurance or other regulatory health plans may also depend upon the net monetary benefit, in direct and indirect savings, to the society and a patient and the cost acquired by switching to PD treatment.

The continued development of PD should consider the cost of development prior to approval and the financial gain of PD upon approval. The consideration of monetary factors in designing experiments and setting research priorities is covered under approaches of value of information analysis[111]. Resources are aligned so as to gain information about a set of parameters/experiments that are most important for the viability of PD.

All of the above aspects could be incorporated into the decision analytic approach as described in this thesis.

References

1. Shillingford CA, Vose CW. Effective decision-making: Progressing compounds through clinical development. *Drug Discovery Today*. 2001;6:941-946.

2. DiMasi JA, Hansen RW, Grabowski HG. The price of innovation: New estimates of drug development costs. *Journal of Health Economics*. 2003;22:151-185.

3. FDA. FDA's Critical Path Initiative. Available at: http://www.fda.gov/oc/initiatives/criticalpath/. Accessed 04/21, 2007.

4. Grasela TH, Fiedler-Kelly J, Walawander CA, et al. Challenges in the transition to model-based development. *AAPS* J. 2005;7:E488-95.

5. Sheiner LB. Learning versus confirming in clinical drug development.. *Clinical Pharmacology and Therapeutics*. 1997;61:275-291.

6. DiPiro JT, Talbert RL, Yee GC, Matzke GR, Wells BG, Posey LM. Pharmacotherapy: A pathophysiologic approach. In: McGraw-Hill Medical; 2008:1323-1333.

7. Wafford KA, Ebert B. Emerging anti-insomnia drugs: Tackling sleeplessness and the quality of wake time. *Nature Reviews Drug Discover*y. 2008;7:530-540.

8. National Sleep Foundation. How Much Sleep Do We Really Need? Available at: http://www.sleepfoundation.org/article/how-sleep-works/how-much-sleep-do-we-really-need. Accessed 01/03, 2010.

9. Schenck CH, Mahowald MW, Sack RL. Assessment and management of insomnia. *The Journal of the American Medical Association*. 2003;289:2475-2479.

10. Spiegel K, Leproult R, Van Cauter E. Impact of sleep debt on metabolic and endocrine function. *Lance*t. 1999;354:1435-1439.

11. Ayas NT, White DP, Manson JE, et al. A prospective study of sleep duration and coronary heart disease in women. *Archives of Internal Medicine*. 2003;163:205-209.

12. Meletis CD, Zabriskie N. Natural approaches for optimal sleep. *Alternative and Complementary Therapie*s. August 2008;14:181-188.

13. Bloch KE. Polysomnography: A systematic review. *Technology and health car*e. 1997;5:285-305.

14. Rechtschaffen A, Kales A. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects.* Washington, D.C.: Public Health Science, US Government Printing Office; 1968.

15. Attarian HP. *Clinical Handbook of Insomnia.* Humana Press; 2004.

16. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* Fourth ed. Arlington, VA: American Psychiatric Publishing Inc; 1994.

17. Doghramji K. The epidemiology and diagnosis of insomnia. *American Journal of Managed Car*e. 2006:S214-S220.

18. Stoller MK. Economic effects of insomnia. *Clinical Therapeutics*. 1994;16:873-897.

19. Passarella S, Duong MT. Diagnosis and treatment of insomnia. *American Journal of Health-System Pharmac*y. 2008;65:927-934.

20. Szabadi E. Drugs for sleep disorders: Mechanisms and therapeutic prospects. *British Journal of Clinical Pharmacology*. 2006;61:761-766.

21. Corrigan BW, Feltner DE, Ouellet D, Werth JL, Moton AE, Gibson G. Effect of renal impairment on the pharmacokinetics of PD 0200390, a novel ligand for the voltage-gated calcium channel alpha-2-delta subunit. *British Journal of Clinical Pharmacology*. 2009;68:174-180.

22. Donevon SD, Li Z, Serpa K, Meltzer L, Taylor CP, Thorpe A. PD0200390, a novel alpha-2delta-ligand enhances slow wave sleep in rats. *Biological Psychiatry.* 2008;63:301S.

23. Hindmarch I, Dawson J, Stanley N. A double-blind study in healthy volunteers to assess the effects on sleep of pregabalin compared with alprazolam and placebo. *Sleep*. 2005;28:187-193.

24. Dooley DJ, Taylor CP, Donevan S, Feltner D. Ca2 channel α2δ ligands: Novel modulators of neurotransmission. *Trends in Pharmacoogical Sciences*. 2007;28:75-82.

25. Corrigan BW, Fang J, Hurst S, et al. An open-label study to investigate the absorption, metabolism, and excretion of PD 0200390 in healthy volunteers. *Clinical Pharmacology and Therapeutics.* 2009;85:S41.

26. Corrigan BW, Werth J, Abel R, Feltner D, Ouellet D. The effect of food on the pharmacokinetics of PD 0200390 following single dose administration to healthy volunteers. *Clinical Pharmacology and Therapeutics.* 2008;83:S30.

27. Corrigan BW, Werth J, Moton A, Feltner D, Ouellet D. Pharmacokinetics of PD 0200390 following single dose administration to subjects with varying degrees of renal function. *Clinical Pharmacology and Therapeutics.* 2008;83:S29.

28. Corrigan BW, Werth J, Moton A, Alvey C, Feltner D, Ouellet D. Pharmacokinetics of PD 0200390 following multiple dose administration to healthy volunteers. *Clinical Pharmacology and Therapeutics.* 2008;83:S30.

29. Freedman D. Some issues in the foundation of statistics. *Foundations of Scienc*e. 1995;1:19-39.

30. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* Wiley; 2004.

31. Cowles K, Kass R, O'Hagan A. ISBA. Available at: http://www.bayesian.org/bayesexp/bayesexp.html. Accessed 07/31, 2009.

32. Efron B. Bayesians, frequentists, and scientists. *Journal of the American Statistical Associatio*n. 2005;100:1-5.

33. Fitelson B. Likelihoodism, bayesianism, and relational confirmation. *Synthes*e. 2007;156:473-489.

34. Howson C, Urbach P. Bayesian reasoning in science. *Natur*e. 1991;350:371-374.

35. Berry DA, Stangl DK. *Bayesian Biostatistics.* CRC; 1996.

36. Dufful SB, Friberg LE, Dansirikul C. Bayesian hierarchical modeling with markov chain monte carlo methods. In: Williams PJ, Ette EI, eds. *Pharmacometrics the Science of Quantitative Pharmacology.* Hoboken, NJ: Wiley-Interscience; 2007:137-164.

37. Berry DA, Mueller P, Grieve AP, et al. Adaptive bayesian designs for dose-ranging drug trials. *Case studies in Bayesian statistic*s. 2001;5:99–181.

38. Lunn DJ, Best N, Thomas A, Wakefield J, Spiegelhalter D. Bayesian analysis of population PK/PD models: General concepts and software. *Journal of Pharmacokinetics and Pharmacodynamics*. 2002;29:271-307.

39. Best NG, Tan KKC, Gilks WR, Spiegelhalter DJ. Estimation of population pharmacokinetics using the gibbs sampler. *Journal of Pharmacokinetics and Pharmacodynamics*. 1995;23:407-435.

40. Bayes T. An essay towards solving a problem in the doctrine of chances. 1763. *MD Computing*. 1991;8:157-171.

41. Carlin BP, Louis TA. *Bayesian Methods for Data Analysis.* Third ed. Chapman and Hall/CRC; 2009.

42. Gelman A, Rubin DB. Inference from iterative simulations using multiple sequences (with discussions). *Statistical Science*. 1992;7:457-511.

43. Gelfand AE, Smith AFM. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*. 1990;85:398-409.

44. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equations of state calculations by fast computing machines. *Journal of Chemical Physic*s. 1953;21:1087.

45. Hastings WK. Monte carlo sampling methods using markov chains and their applications. *Biometrika*. 1970;57:97-109.

46. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*. 2000;10:325-337.

47. Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS open. *R New*s. 2006;6:12-17.

48. Brooks SP, Gelman A. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*. 1998;7:434-455.

49. Kass RE, Raferty AE. Bayes factors. *Journal of the American Statistical Association*. 1995;90:773-795.

50. Dey DK, Chen MH, Chang H. Bayesian approach for nonlinear random effects models. *Biometrika*. 1997;53:1239-1252.

51. Geisser S, Eddy W. A predictive approach to model selection. *Journal of the American Statistical Association*. 1979;74:153-160.

52. Aitkin M. Posterior bayes factors. *Journal of the Royal Statistical Society, Series B*. 1991:111-142.

53. O'Hagan A. Fractional bayes factors for model comparison. *Journal of the Royal Statistical Society.Series B*. 1995:99-138.

54. Berger JO, Pericchi LR. The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association*. 1996;91:109-122.

55. Raftery AE. Hypothesis testing and model selection via posterior simulation. *Practical Markov Chain Monte Carl*o. 1996:163-188.

56. Dey DK, Kuo L, Sahu SK. A Bayesian predictive approach to determining the number of components in a mixture distribution. *Statistics and Computing*. 1995;5:297-305.

57. Gelfand AE, Dey DK, Chang H. Model determination using predictive distributions with implementation via sampling-based methods. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM. *Bayesian Statistics 4: Proceedings of the fourth Valencia International Meeting.* New York: Oxford University Press; 1992:147-167.

58. Akaike H. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*. 1969;21:243-247.

59. Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978;6:461-464.

60. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measure of model complexity and fit (with discussion). *Journal of Royal Statistical Society, Series B*. 2002;64:583-639.

61. Yano Y, Beal SL, Sheiner LB. Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *Journal of Pharmacokinetics and Pharmacodynamics*. 2001;28:171-192.

62. Chapman GB, Sonnenberg FA. Introduction. In: Chapman GB, Sonnenberg FA. *Decision Making in Health Care: Theory, Psychology, and Applications.* Cambridge, U.K.: Cambridge University Press; 2003:3-20.

63. Berger JO. *Statistical Decision Theory and Bayesian Analysis.* 2nd ed. New York: Springer-Verlang; 1985.

64. Chamberlain G. Econometrics and decision theory. *Journal of Econometrics*. 2000;95:255-283.

65. Sheiner LB. PK/PD approach to dose selection. *International Congress Series.* Elsevier; 2001;1220:67-78.

66. Keeney RL. Decision analysis: An overview. *Operations Research*. 1982:803-838.

67. Von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior.* Princeton: Princeton University Press; 1944.

68. Robert CP. *The Bayesian Choice: From Decision-Theoretic Motivations to Computational Implementation.* Springer Texts in Statistics. New York: Springer-Verlag; 2007.

69. Savage LJ. *The Foundations of Statistics.* New York: Dover Publications; 1972.

70. Pratt JW, Raiffa H, Schlaifer R. The foundations of decision under uncertainty: An elementary exposition. *Journal of the American Statistical Association*. 1964:353-375.

71. Fishburn PC. Reconsiderations in the foundations of decision under uncertainty. *The Economic Journal*. 1987:825-841.

72. Anscombe FJ, Aumann RJ. A definition of subjective probability. *The Annals of Mathematical Statistics*. 1963:199-205.

73. Wald A. Statistical decision functions. *The Annals of Mathematical Statistics*. 1949:165-205.

74. Von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior.* Princeton: Princeton University Press; 1944.

75. Raiffa H, Schlaifer R. *Applied Statistical Decision Theory.* Boston, MA: Harvard University Press; 1961.

76. DeGroot MH. *Optimal Statistical Decisions.* New York: McGraw Hill; 1970.

77. Parmigiani G. Decision theory: Bayesian. *International Encyclopedia Social and Behavioral Sciences*. 2001.

78. Clyde MA. Experimental design: A Bayesian perspective. *International Encyclopedia of Social and Behavioral Sciences*. 2001;8:5075-5081.

79. Lindley DV, Singpurwalla ND. On the evidence needed to reach agreed action between adversaries, with application to acceptance sampling. *Journal of the American Statistical Association*. 1991;86:933-937.

80. Chaloner K, Verdinelli I. Bayesian experimental design: A review. *Statistical Science*. 1995;10:273-304.

81. Lindley DV. Basic ideas in Bayesian statistics. In: *Bayesian Statistics, a Review.* Philadelphia, PA: Society for Industrial and Applied Mathematics; 1972:17-31.

82. Poland B, Wada R. Combining drug–disease and economic modelling to inform drug development decisions. *Drug Discovery Today*. 2001;6:1165-1170.

83. Poland B, Hodge FL, Khan A, Clemen, RT, Wagner JA, Dykstra K, Krishna R. The clinical utility index as a practical multi-attribute approach to drug development decisions. *Clinical Pharmacology and Therapeutics*. 2009;86:105-108.

84. Ouellet D, Werth J, Parekh N, Feltner D, McCarthy B, Lalonde RL. The use of a clinical utility index to compare insomnia compounds: A quantitative basis for Benefit–Risk assessment. *Clinical Pharmacology & Therapeutics*. 2009;85:277-282.

85. Graham G, Gupta S, Aarons L. Determination of an optimal dosage regimen using a Bayesian decision analysis of efficacy and adverse effect data. *Journal of Pharmacokinetics and Pharmacodynamics*. 2002;29:67-88.

86. Jönsson S, Karlsson MO. Estimation of dosing strategies aiming at maximizing utility or responder probability, using oxybutynin as an example drug. *European Journal of Pharmaceutical Sciences*. 2005;25:123-132.

87. Troche CJ, Paltiel AD, Makuch RW. Evaluation of therapeutic strategies: A new method for balancing risk and benefit. *Value in Health*. 2000;3:12-22.

88. Conaway MR, Petroni GR. Designs for phase II trials allowing for a trade-off between response and toxicity. *Biometrics*. 1996;52:1375-1386.

89. Yin G, Li Y, Ji Y. Bayesian dose-finding in phase I/II clinical trials using toxicity and efficacy odds ratios. *Biometrics*. 2006;62:777-787.

90. Stallard N, Thall PF, Whitehead J. Decision theoretic designs for phase II clinical trials with multiple outcomes. *Biometrics*. 1999;55:971-977.

91. Bazzoli C, Retout S, Mentre F. Design evaluation and optimisation in multiple response nonlinear mixed effect models: PFIM 3.0. *Computer Methods and Programs in Biomedicine*. 2009.

92. Foracchia M, Hooker A, Vicini P, Ruggeri A. POPED, a software for optimal experiment design in population kinetics. *Computer Methods and Programs in Biomedicine*. 2004;74:29-46.

93. Mentre F, Mallet A, Baccar D. Optimal design in random-effects regression models. *Biometrika*. 1997;84:429-442.

94. Atkinson AC, Chaloner K, Herzberg AM, Juritz J. Optimum experimental designs for properties of a compartmental model. *Biometrics*. 1993;49:325-337.

95. Stroud JR, Muller P, Rosner G. Optimal sampling times in population pharmacokinetic studies. *Journal of the Royal Statistical Society. Series C*. 2001;50:345-359.

96. Han C, Chaloner K. Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. *Biometrics*. 2004;60:25-33.

97. Palmer JL, Muller P. Bayesian optimal design in population models for haematologic data. *Statistics in Medicine*. 1998;17:1613-1622.

98. Wakefield J. An expected loss approach to the design of dosage regimens via sampling-based methods. *The Statistician*. 1994;43:13-29.

99. Rosner GL, Tang F, Madden T, Andersson BS. Dose individualization for high-dose anti-cancer chemotherapy. In: D'Argenio DZ. *Advanced Methods of Pharmacokinetic and Pharmacodynamic Systems Analysis.* Vol 3. Boston: Kluwer Academic Publishers; 2004:239-253.

100. Etzioni R, Kadane JB. Optimal experimental design for another's analysis. *Journal of the American Statistical Association*. 1993;88:1404-1411.

101. Müller P, Berry DA, Grieve AP, Smith M, Krams M. Simulation-based sequential Bayesian design. *Journal of Statistical Planning and Inference*. 2007;137:3140-3150.

102. Smith MK, Jones I, Morris MF, Grieve AP, Tan K. Implementation of a Bayesian adaptive design in a proof of concept study. *Pharmaceutical Statistics*. 2006;5:39-50.

103. Smith MK, Marshall S. A Bayesian design and analysis for dose-response using informative prior information. *Journal of Biopharmaceutical Statistics*. 2006;16:695-709.

104. Krams M, Lees KR, Hacke W, Grieve AP, Orgogozo JM, Ford GA. Acute stroke therapy by inhibition of neutrophils (ASTIN) an adaptive dose-response study of UK-279,276 in acute ischemic stroke. *Stroke*. 2003;34:2543-2548.

105. Gershell L. Insomnia market. *Nature Reviews Drug Discovery*. 2006;5:15-16.

106. Kupfer DJ, Reynolds CF. Management of insomnia. *New England Journal of Medicine*. 1997;336:341-346.

107. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria:2009.

108. Kadane JB, Wolfson LJ. Experiences in elicitation. *The Statistician*. 1998;47:3-19.

109. Bielza C, Müller P, Insua DR. Decision analysis by augmented probability simulation. *Management Science*. 1999;45:995-1007.

110. Lindley DV. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*. 1956;27:986-1005.

111. Willan AR. Optimal sample size determinations from an industry perspective based on the expected value of information. *Clinical Trials*. 2008;5:587-594.

Appendix A.

1. WinBUGS code for individual-outcome-fitting (relevant to Chapter 2)

```
model{

#WASO MODEL
# TYPE 1
for(i in 1:510){
log.y[i]<-log(y[i])
log.y[i] ~ dnorm(m[i],tau1)
pcb[i]<-exp(t[id[i],1])
ed1[i]<-exp(t[id[i],2])
e.m[i] <- pcb[i]*(1- (dose[i]/(ed1[i]+ dose[i])))
m[i]<-log(e.m[i])
y.pred[i]~dnorm(m[i],tau1)
}

# TYPE 2
for(i in 511:1020){
y[i] ~ dpois(m[i])
m[i] <- pcb[i]*(1-dose[i]/(ed2[i]+dose[i]))
ed2[i]<-exp(t[id[i],4])
pcb[i] <- exp(t[id[i],3])*pcb.mu[i]
pcb.mu[i] <-exp(b[1])*pow((age[i]/47),b[2])*pow(exp(b[3]),gen[i])*
pow(exp(b[4]),std[i])
y.pred[i]~dpois(m[i])
}

# TYPE 3
for(i in 1021:1412){
y[i] ~ dbern(m[i])
logit(m[i]) <- b[5]+b[6]*dose[i]+t[id[i],5]
y.pred[i]~dbern(m[i])
}

# TYPE 4
for(i in 1413:1804){
y[i] ~ dbern(m[i])
logit(m[i]) <- b[7]+b[8]*dose[i]+t[id[i],6]
y.pred[i]~dbern(m[i])
```

```
}

#Specification of non-informative priors for the individual-outcome fitting

for(i in 1:118){
  t[i,1] ~ dnorm(mu[1],omega[1])
  t[i,2] ~ dnorm(mu[2],omega[2])
  t[i,3] ~ dnorm(mu[3],omega[3])
  t[i,4] ~ dnorm(mu[4],omega[4])
  t[i,5] ~ dnorm(mu[5],omega[5])
  t[i,6] ~ dnorm(mu[6],omega[6])
}
for(i in 1:6){omega[i]~dgamma(0.001,0.001)}
for(i in 1:8){b[i]~dnorm(0,0.001)}
tau1~dgamma(0.001,0.001)

 mu[1]~dnorm(0,0.001)
 mu[2]~dnorm(0,0.001) #ed1
 mu[3]<-0
 mu[4]~dnorm(0,0.001) #ed2
 mu[5]<-0
 mu[6]<-0
}
#Initial Estimates
#init 1
list(omega=c(5,2,15,6,1.2,1.2),mu=c(3,3.4,NA,4,NA,NA),b=c(2,0.5,0.3,0.5,-
2,0.03,-3,0.04), tau1=7)

#init 2
list(omega=c(2.5,1,7.5,3,0.6,0.8), mu=c(2.1,2.4,NA,2.8,NA,NA),
b=c(2,0.5,0.3,0.5,-2,-0.03,-3,-0.04), tau1=7)
```

2. *WinBUGS code for simultaneous-outcome-fitting (relevant to Chapter 2)*

```
model{

#WASO MODEL
# TYPE 1
for(i in 1:510){
log.y[i]<-log(y[i])
log.y[i] ~ dnorm(m[i],tau1)
pcb[i]<-exp(t[id[i],1])
ed1[i]<-exp(t[id[i],2])
e.m[i] <- pcb[i]*(1- (dose[i]/(ed1[i]+ dose[i])))
m[i]<-log(e.m[i])
y.pred[i]~dnorm(m[i],tau1)
}

# TYPE 2
for(i in 511:1020){
y[i] ~ dpois(m[i])
m[i] <- pcb[i]*(1-dose[i]/(ed2[i]+dose[i]))
ed2[i]<-exp(t[id[i],4])
pcb[i] <- exp(t[id[i],3])*pcb.mu[i]
pcb.mu[i] <-exp(b[1])*pow((age[i]/47),b[2])*pow(exp(b[3]),gen[i])*
pow(exp(b[4]),std[i])
y.pred[i]~dpois(m[i])
}

# TYPE 3
for(i in 1021:1412){
y[i] ~ dbern(m[i])
logit(m[i]) <- b[5]+b[6]*dose[i]+t[id[i],5]
y.pred[i]~dbern(m[i])
}

# TYPE 4
for(i in 1413:1804){
y[i] ~ dbern(m[i])
logit(m[i]) <- b[7]+b[8]*dose[i]+t[id[i],6]
y.pred[i]~dbern(m[i])

}
```

```
#Specification of non-informative priors for the simultaneous-outcome fitting

for(i in 1:118){ t[i,1:6] ~ dmnorm(mu[1:6],omega[1:6,1:6]) }
omega[1:6,1:6] ~ dwish(R[1:6,1:6],6)#Wishart distribution
 tau1~dgamma(0.001,0.001)#Gamma distribution
 for(i in 1:8){b[i] ~ dnorm(0,0.001)}#Normal distribution
 mu[1]~dnorm(0,0.001) #Normal distribution
 mu[2]~dnorm(0,0.001) #Normal distribution
 mu[3]<-0
 mu[4]~dnorm(0,0.001) #Normal distribution
 mu[5]<-0
 mu[6]<-0

#Calculation for correlation between random effects from the precision matrix
 oinv[1:6,1:6] <- inverse(omega[1:6,1:6])
 for(i in 2:6){
 for(j in 1:i-1){
  corr[i,j] <- oinv[i,j]/sqrt(oinv[i,i]*oinv[j,j])}
}

}
#Initial Estimates
#init 1
list(omega=structure(.Data=c(5,0,0,0,0,0,
                             0,2,0,0,0,0,
                             0,0,15,0,0,0,
                             0,0,0,6,0,0,
                             0,0,0,0,1.2,0,
                             0,0,0,0,0,1.2), .Dim = c(6, 6)),
list(mu=c(3,3.4,NA,4,NA,NA),b=c(2,0.5,0.3,0.5,-2,0.03,-3,0.04), tau1=7)

#init 2
list(omega=structure(.Data=c(2.5,0,0,0,0,0,
                             0,1,0,0,0,0,
                             0,0,7.5,0,0,0,
                             0,0,0,3,0,0,
                             0,0,0,0,0.6,0,
                             0,0,0,0,0,0.8), .Dim = c(6, 6)),
list(mu=c(2.1,2.4,NA,2.8,NA,NA),b=c(2,0.5,0.3,0.5,-2,-0.03,-3,-0.04), tau1=7)

#Data
list(R = structure(.Data = c(1.3,0,0,0,0,0,
                             0,3.2,0,0,0,0,
                             0,0,0.36,0,0,0,
```

```
0,0,0,0.88,0,0,
0,0,0,0,5,0,
 0,0,0,0,0,3.75), .Dim = c(6, 6)))
```