View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by ScholarBank@NUS

BAYESIAN PROBABILITY ENCODING IN MEDICAL DECISION ANALYSIS

CHAN SIEW PANG

NATIONAL UNIVERSITY OF SINGAPORE
2007

BAYESIAN PROBABILITY ENCODING IN MEDICAL DECISION ANALYSIS

CHAN SIEW PANG

BSocSci(Hon) NUS
MSc(Management) NUS
MSc(Medical Statistics) UOL
CMathMIMA UK
CSci UK

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF INDUSTRIAL & SYSTEMS ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE
2007

Acknowledgements

I owe a great debt of gratitude to my supervisor Associate Professor Poh Kim Leng for introducing me to the exciting field of decision analysis. The subject is not a collection of isolated techniques, but is, instead, a coherent body of knowledge, which can be applied in daily life. As a student I'm very fortunately to have Prof Poh as my teacher. It is a pleasure to thank him for his patience, support and guidance.

I am also grateful to Mr. Ian White, former Senior Lecturer at the London School of Hygiene & Tropical Medicine, University of London, for sharing with me his insightful understanding of Bayesian analysis during my one-year stay in England.

Special thanks go to my collaborators and the anonymous patients who participated in the reported studies. Many helpful ideas and comments were provided by them when I was preparing this dissertation.

My wife Chew Cheng has been a constant source of encouragement throughout, especially during my rehabilitation from a career-threatening eye operation in April 2002. My heartfelt thanks go to A/P Lim Tock Han, Department of Ophthalmology, Tan Tock Seng Hospital Pte Ltd, for restoring my vision. Last but not least, I also wish to express my sincere thanks to Dr. Suresh Sahadevan, General Medicine, Tan Tock Seng Hospital Pte Ltd, for his continual spiritual support over the years. .

Views expressed in this project are solely the author's and should not be attributed to the relevant authorities. I am responsible for all mistakes made in the project. As a final note, I hope readers of this dissertation will quickly recognise that medical decision analysis is an extremely interesting field of study!

SP Chan

21 March 2007

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Summary

The primary objective of this dissertation is to develop two classes of Bayesian models for probability encoding in medical decision analysis. The models are developed from the original Bayes' Theorem and various fundamental concepts that underlie the development of contemporary statistics.

The models are developed with the nature of medical evidence in mind. This is because probability encoding hinges on the availability and features of evidence. Forming the basis of reasoning, evidence refers to any explicit warranted reference given in an appropriate and specific context for supporting or rejecting a hypothesis, claim or belief.

Specially designed for analysing subject-level evidences, the first class of models follows the framework of Generalised Linear Models (GLM). Unlike the conventional GLM approach, these models require the union of the observed evidences (likelihood) with a carefully chosen prior of the canonical parameter(s) that underlie the distribution of the outcome variable.

The second class of models may be referred to as meta-analytic methods as they are applied for synthesising aggregate-level evidences from reported sources. To reflect the large amount of heterogeneity among the studies to be combined, the models incorporate some random effects in the set-up. Inevitably, these models are hierarchical in nature and have to be estimated with the Gibbs sampler.

Although these techniques are complicated so that all salient features underlying the decision problems are adequately captured, they are also simple enough for routine use in clinical practice.

The recognition of the importance of Bayesian ideas in probability encoding will also bring considerable impact on how evidence-based medicine (EBM) is

practiced. One must be ready to embrace more sources of prior evidences which have hitherto being ignored in the current EBM practice. Through the Bayesian framework the synergism between subjective and objective evidences come into play, with the decision analyst and domain experts giving valid testimony and searching for relevant evidence useful for medical decision making.

The application of the proposed Bayesian models is a small step towards the fuflillment of EBM's objective of making use the most complete evidence available for treating patients. It is hoped that the practical aspect of the Bayesian models and their related concepts will appeal to clinicians and decision analysts engaged in routine decision making.

List of Tables

List of Figures

List of Symbols

θ	Canonical parameter/hyperparameter
μ	Mean parameter
$\boldsymbol{\beta}$	Coefficient of a relational model
$\hat{\boldsymbol{\beta}}$	Conventional estimator of $\boldsymbol{\beta}$
$\boldsymbol{\beta}^*$	Bayesian posterior estimator of $\boldsymbol{\beta}$
$\boldsymbol{\beta}_0$	Prior of $\boldsymbol{\beta}$
σ^2	Variance
$\hat{\sigma}^2$	Conventional estimator of σ^2
a, b, u and v	Prior parameters of a Bayesian model
[**y X**]	Data design of a relational model **y**: outcome **X**: covariates/predictors
B	Bayes factor
C	Constant
f(•)	Probability distribution function
F(•)	Cumulative distribution function
g(•)	Prior distribution
L(•)	Likelihood function
P(•)	Probability function
S(•)	Survival function
U(•)	Utility function
π(•)	full posterior distribution
π_c(•)	conditional posterior distribution
Γ(•)	Gamma function
γ(•), κ(•), ν(•), ψ(•)	Arbitrary functions
E[•]	Expectation of a random variable
V[•]	Variance of a random variable
COV[•]	Covariance of two random variables
$\boldsymbol{\Sigma}$	Covariance matrix
T[•]	Estimator
u	Residual of a relational model
ρ	Risk tolerance
η	Systematic component or function of a relational model
$\kappa(\eta)$	Link function
n	Sample size
p	Number of predictors in a relational model / Probability / Encoded probability
q	Shape parameter of a beta distribution
m	Number of algorithmic iterations
k	Number of studies selected for combining published evidences
i	*i*-th observation
j	*j*-th category
t	*t*-th iteration of a computational algorithm

CHAPTER 1

INTRODUCTION

1.1 Motivation

Due to the growing public awareness, medical practice is in the middle of a profound transition. Contemporary scientific medicine has entered upon a period of "paradigmatic instability"—that is, a period in which clinicians need to scrutinise their practice afresh. Advances in medical research and technology mean that clinicians know more about disease than ever. New medicines are constantly being developed, life support and intensive care improve all the time and patients can recover quickly after modern microsurgery. Yet, clinicians still do not have all the answers, and many disorders cannot be cured. Clinicians are also confronted with a wide range of decisions with ethical considerations, which their predecessors might not have encountered.

What is going on is that one of the most basic assumptions underlying medical practice is being challenged. The assumption is not about the validity of new medical discoveries, but concerns the intellectual foundation of medical care or simply put, whatever a clinician decides is sound and desirable for his patients. The implicit message of this transition in medical practice is that while many decisions are undoubtedly correct, some are not, and elaborate mechanisms are needed to sort out which are the desirable ones.

As such, this dissertation would like to point out that the burning issue is not whether there are variations in medical practice and the urgency to reconcile them, but rather how we ensure clinicians make good decisions. Undoubtedly, guidelines are

important in preventing malpractice, but one must bear in mind that medical practice is subject to change as scientific knowledge advances. Therefore, the more fundamental issue is to develop a reliable framework upon which clinicians could make sound decisions in view of the continual evolution of patterns of medical care. In fact, this is the desired attribute that forms the basis of all medical guidelines. We must reckon that the quality of medical care is determined mainly by the quality of clinical decisions that dictate what actions are taken.

With this in mind, the application of decision analysis is advocated. Decision analysis is a methodology based on a probabilistic framework that provides a logical and systematic structure for generating clear and consistent action for the decision-maker [1]. From the perspective of game theory [2], a decision problem is a triple (C, π, O), that consists of an option space (C) to be applied by the decision-maker, a set of outcomes (O) to be realised by the decision maker, and a mediation mechanism, or mapping function, π: C $\rightarrow$ O, that relates choices and outcomes [3]. The decision maker is an entity who is capable of making an autonomous choice from a set of options. He also has the authority and responsibility to implement the selected alternative.

While many clinicians may not appreciate the mathematical details involved, its framework does provide the structure and guidance for systematic thinking in difficult situations. The whole spectrum of activities concerning clinician-patient communication is also structured to help decision makers to identify choices under uncertainty. This is helpful for carrying out decision making related to their practice. Consciously or subconsciously, explicitly or implicitly, every decision maker might have applied some basic rules advocated by the discipline and it often proves useful in developing medical guidelines and for identifying the most desirable therapeutic

strategy for patients. Due to the hailstorm of uncertainties that surround medical care and therapeutic interventions, proper decision analysis is a reliable anchor in the sea of fuzziness.

1.2 Medical Decision Analysis

In its broad sense, medical decision analysis refers to a cluster of quantitative techniques useful for the modelling, measurement and evaluation of medical evidences, processes and outcomes. This notion is familiar to most clinical researchers who apply statistical methods to evaluate results generated from their studies. Several methodological issues of this nature are explored extensively in the dissertation and they serve to provide useful inputs for medical guideline development and decision making.

In addition, the narrower sense of decision analysis is also highlighted and implemented in various problems. More familiar to economists, industrial engineers, mathematicians and policy-makers, it refers to the modelling of a decision in the form of a tree or an influence diagram and the process of identifying the optimal course of actions that maximises the decision maker's satisfaction. It offers a structured, systematic and quantitative approach for evaluating decisions with alternatives, uncertain outcomes and competing objectives.

Decision tree [4] and influence diagram [5] are two different ways for presenting the decision problems. While the tree diagram may be a more conventional form of representation, influence diagram provides a more elegant and succinct representation when the size of the tree becomes ungainly large. However, a decision tree is preferred over an influence diagram should the problem on hand is less complicated, as it provides a more visual approach to decision problems. The comparison of decision trees and influence diagrams is documented in literature [6]. It is also worthwhile to note that both the decision trees and influence diagrams are isomorphic, that is, any property built on the latter can be converted into the former, and vice versa.

One of the many notable advantages for applying decision analysis is that it is able to generate a number of graphical tools for model evaluation. At each step of modelling a great deal of insights may be produced so that the analysis could be modified promptly and efficiently. The following sequence of steps is applied for developing a medical decision analysis [7]:

- Define the decision problem and its time horizon
- Identify a set of candidate decision alternatives
- List the possible clinical outcomes of each of the candidate alternatives
- Represent the sequence of events leading to the clinical outcomes
- Determine the probability of each chance event
- Assign a value to each clinical outcome

The term "decision alternative" denotes the decision maker's range of options. In a decision tree or an influence diagram, the decision and the chance outcomes are represented by nodes. The value of each outcome is often expressed in terms of the decision maker's utility. In the patient's context, the utility quantifies his differing attitudes to risk and his relative desirability of the outcome states. As a rational entity, he must be able to rank his preferences according to the outcomes of the various options. The probabilities on the chance nodes, on the other hand, quantify the pervading uncertainties, which always create clouds of discomfort to medical decision makers. A chance node is thus the point in a decision tree at which probability determines which outcome will occur. In medical decision analysis, possible outcomes of chance nodes include disease present/absent, survive/dead,

improvement/deterioration in health condition, remission/relapse following a surgical operation, and recovery/no recovery after treatment. A patient's utility and the probabilities on the chance nodes are determined independently.

The normative Expected Utility Theory (EUT) [2] states that the decision maker chooses between uncertain outcomes by comparing their expected utility value, which is the weighted sum obtained by adding the utilities of outcomes multiplied by their respective probabilities. The most desired decision is one that maximises the expected utility. The fundamental axioms of expected utility are documented in references [2, 6]. It is interesting to note that while these assumptions are reasonable under most circumstances; many decision theorists find some of the axioms controversial. These range from introspection regarding particular decision situations to formal psychological experiments in which human subjects make choices that are inconsistent with one or more of the axioms [8-12]. The behavioural paradoxes, however, do not necessarily invalidate the idea that one should still make decisions according to the EUT. The argument all along has been that people do not seem to make coherent decisions without some guidance. In constructive terms, the decision assessment process helps to mould the decision maker's preferences and his understanding about uncertainties. Individuals who do not think long and hard enough in developing their preferences and beliefs might have a tendency to make inconsistent judgements [6].

In terms of the above-mentioned set-up, there is no drastic difference between medical decision analysis and ordinary decision analysis frequently applied in business, economics, engineering, military operations and public policy evaluation. However, extra care must be taken in the formulation phase so that the chance and decision variables chosen should cohere with the medical domain and reflect the

current state of medical knowledge. This also helps to determine the types and number of alternatives and objectives for a specific decision problem. In addition, elicitation of patient utilities may also pose a serious challenge to the analyst as many patients may not know their preferences precisely.

The use of decision analysis in solving medical problems engages the patients in every single step of the process, as the primary goal is to maximise the patient's well-being. Hence, medical decision analysis should be duly recognised as an integral part of contemporary medical practice. It is also fast becoming an indispensable tool of evidence-based medicine (EBM), a particular branch of medical practice that is gaining world-wide attention in recent years. Emerged in the 1990s, EBM formalises the scientific principle of basing clinical practice on evidence. Advocating the conscientious, explicit and judicious use of current best evidence in health care [13], EBM allows research findings be critically appraised and interpreted, thus increasing the likelihood of making better informed decisions. .

To facilitate discussion, the terms used throughout the dissertation must be properly defined. A "clinician" is a qualified doctor who renders medical care to patients, either in the form of surgical operation or drug treatment or both. Next, "decision maker" is referred to both clinician and patient who are an integral part of the decision-making process. An "analyst" , who may be a decision analyst or statistician by profession, is one who provides expertise in solving specific technical problems at various stages of the process, including probability encoding and generation of patients' utility. An investigator is one who initiates and conducts the decision analysis. Last but not least, "domain experts" are those who provide specialised medical advice, upon invitation, for specific aspects of the decision.

1.3 Objective

Utility elicitation and probability encoding are crucial to the proper formulation and analysis of a medical decision problem. The objective of this dissertation is to focus on developing useful probability-encoding models for routine use in medical decision analysis.

Central to probability encoding and the analysis of medical decision problems is the collection and interpretation of evidence. However, evidence is always tentative and obscure in nature. This is because medical research bears a large degree of uncertainties, which may not be completely eradicated even by employing the most sophisticated study design and analytical method. In fact, all forms of inductive conclusions are provisional and are subject to change in light of new evidence. The major causes of uncertainty in medical decision analysis include the following:

- limited knowledge of the medical problem under study
- missing information for the complete understanding of a problem
- subjects enrolled for study are merely a sample of the larger population (sampling error)
- censored medical information
- errors due to both investigators' limited sensory power and sensitivity of the medical equipment
- varying conditions of related medical research findings
- inadequate or inconsistent conclusions from past medical studies

The public is often baffled with conflicting and uncertain medical evidence reported in news. For example, there are mixed published evidence regarding the

potential benefits for breast cancer screening on mortality [14]. Even in situations where there is consistent evidence, uncertainties pervade. While it is generally acknowledged that higher levels of physical activity are associated with decreased risk of coronary heart disease, hypertension, cancer and possibly longevity [15-16], there is a shortage of convincing evidence on what is the threshold level of desired physical activity. Contrary to the common belief that prolonged vigorous physical exercises might exert unnecessary burden on our body, there is evidence showing that professional athletes might enjoy better long-term life expectancy than the general public [17].

The persisting variable degree of uncertainty calls for the application of probabilistic thinking in medical decision analysis. Since uncertainty cannot be eliminated from decision problems, it has to be accommodated and modelled with relevant available evidences.

Relevant evidence is one that makes the fact requiring proof more or less probable. Therefore, the probabilities we assign to our conclusion(s) depend not only on how much evidence we have but also how we interpret the evidence and how confident we are with the interpretation. We must also revise our assigned probabilities when new evidence surfaces. These are then updated on the chance nodes of the decision model. Hence, the methodological issues involved in using evidence for medical decision making involves not only evidential collection, but also how we analyse the evidence and with what degree of assurance.

Probability provides decision analysts with the scientific theories, mathematical concepts and computational techniques for quantifying uncertainties. Under uncertainty, the decision maker knows the specific outcomes associated with each alternative, but he does not know the probabilities to be associated with the

states [18]. This is often the scenario of a typical medical decision problem and it leads one to recognise that medical decision making is "an art of probabilities" [19]. It is beyond the scope of this dissertation to provide a formal treatment of probability and its related concepts such as causality [20]. One may refer to the relevant references for a more rigorous treatise [21-24].

To sum up, this dissertation aims to develop a useful and versatile framework for probability encoding which may be routinely applied in solving medical decision problems. This calls for not only a proper understanding of probability but also the nature of medical evidence gathered and interpreted for decision making. A reliable probability-encoding framework is one that is able to reflect the very nature of medical evidence, which forms the main focus of the next section.

Considering the unique characteristics of clinical research and decision making, the Bayesian framework is advocated. With the help of several specific models developed under the framework, routine clinical decision making may be carried out with much ease. They are applied to shed light on a number of clinical and healthcare decision problems. However, the implications of the Bayesian framework are far more profound. Capable of transforming our current notion of evidence, probability and decision making, the Bayesian framework will enrich the practice of EBM which advocates the judicious use of best evidence in health care.

The Bayesian probability-encoding models advocated in this dissertation are sophisticated in nature but not beyond the scope of the less mathematically-inclined, especially the clinicians. To accede to their needs, this dissertation is prepared with medical professionals in mind. The specific Bayesian models advocated are designed and recommended for routine use in medical practice.

1.4 Medical Evidence

1.4.1 The Salient Nature

Since the reliability and accuracy of probability encoding hinges on the use of evidence, the nature of medical evidence must be closely examined. While the above discussion suggests that evidence is tentative and uncertain in nature, the following explains that it may be both objective and subjective. It is a common mistake that evidence can only exist in an objective state. This is partly caused by its confusion with other related terms such as facts, information and data.

What is taken as a fact depends upon the extent to which observations are corroborative. It is any thing capable of being received by the senses. We may gather evidence about some phenomenon, but if this evidence is to any degree inconclusive we are not entitled to conclude that it entails factual contents of the problem. Moreover, a fact is evidential only if it is applied in an appropriate context where inferences about the problem can be made. It is said to be "proved" or "disproved" when after considering all the evidence before it, the medical community believes it to exist, or considers its existence so probable that any prudent clinician ought to act upon the supposition that it exists. Similarly, while we might all agree that evidence generates information, we cannot equate the two terms. For instance, a document written in an obscure language may be recognised as relevant but non-informative for drawing inferences or decision-making. It becomes informative only when some explicit meanings are attached. Last but not least, data are quantified evidence intentionally gathered or established as references for verifying a hypothesis. These are typically clinical observations as seen, measured and recorded. Clinicians sometimes speak of "hard data". This refers to clinical or para-clinical data that can be precisely defined and measured, such as blood cell count, heart rate and glucose

level. By contrast, soft data are observations that are relatively difficult to define, measure and classify. Typical examples include sorrow, anxiety, general well-being and pain experienced by patients. The "hardening" of soft data refers to all means employed to improve the criteria, measurement and quantification of soft data in order to match that of hard data as closely as possible.

More importantly, our observations of any kind produce only abstraction or representation of the phenomenon in question. Observation is a subjective affair and subjects are known to differ widely in their sensory capacity and other observational characteristics. This implies that the concept of evidence should not be limited to references that are directly observable to the subject. Otherwise, a medical decision maker may have to discard a great deal of evidence that cannot be observed directly, such as patients' personal assessment of fear or depression.

Moreover, most clinicians are accustomed to believe that knowledge is only justified with empirical confirmation. According to the conventional scientific framework, the process of knowledge accumulation can be broken down into the following steps:

- propose a hypothesis concerning an observed phenomenon
- design a study to test the hypothesis
- acquire and analyse the data from the study
- test the results against the hypothesis
- draw conclusions given the results
- advance understanding of the phenomenon

However, this leaves open a number of metaphysical and ontological questions, including the source of inspiration for hypothetical development, the dependence of observation and analysis on the researchers' perceptions and the epistemological path to gaining insightful conclusions of the study. Clearly, scientific investigation is not a 100% objective affair.

Similarly, the warrantability of evidence may also be established through semantic clarification and logical reasoning. For example, a clinician does not need to conduct an experiment to prove that plunging from a high-rise building without any safety aid can cause death. Moreover, empirical warrantability stems from a confirmatory relation to specific conditions of first-person experience, which may be established outside the self in the real world (observation) or through personal experience, if honestly reported.

In a nutshell, it is erroneous to think that evidence can always be observed or measured objectively. To be useful for decision making, the relevance of evidence must be established. This requires a proper presentation of the qualitative and quantitative characterisation of phenomenon under study. Moreover, one must also ensure that the relevant evidence is collected and analysed within the appropriate context.

1.4.2 Expert Opinion

Taking into account its subjective nature, evidence may then be classified as tangible (real or documentary) or intangible, with testimony as the most common form of the latter. Simply put, evidence is the means by which the claimant tries to defend/prove his case and the opposition tries to cast doubt upon or disprove the

hypothesis. Therefore, medical evidence should also include testimonial assertions and authoritative opinions (direct evidence) that are admissible and relevant.

In this context, the so-called authorities or experts must be competent (based on verifiable collateral facts) and are able to elucidate their opinions. The challenges facing decision analysts are to assess the admissibility and relevance of these opinions and to quantify them so that they are evidential or informative for decision-making. It is a precondition for admissibility that evidence is relevant.

This dissertation asserts that testimony is a valid form of evidence, whereby a witness relates what he believes. In providing testimony for medical decision-making, the expert effectively acts as a "witness" and his evidence is often presented in the form of "opinion". To facilitate discussion it is important to distinguish the expert from the analyst, who elicits the evidence from the former in providing solutions to decision making.

Generally, opinion refers to ideas or beliefs provided by a subject while interpreting a particular phenomenon. It has been well-settled in the legal discipline that a view offered that is based on one's education, training and experience is an "expert opinion". Expertise, in its broadest sense, is the accurate application of knowledge, beliefs and experience to certain situations. Experts typically identify and understand the nature of a presented problem within their domain of knowledge and are able to establish its representation beyond the scope of novice. As such, expert opinions may only be offered by a suitably-qualified person widely acknowledged in his field of practise, and with a good credential and track record. Such evidence may be more appropriately termed as "opinion evidence", in accordance with the earlier discussion of evidence.

As such, the admissibility of expert opinion depends on two factors. First, the analyst who is responsible for eliciting the opinion evidence must be satisfied with the witness's status as an expert and this will, naturally, involve a consideration of his qualification and experience. The burden of proof in establishing expertise lies with the analyst seeking to call the witness. Second, an expert opinion must relate to an issue that goes beyond the competence of the analyst and must be necessary to aid the analyst in understanding the issue of reaching a decision of the presented evidence. The identified expert bears such evidential burden and he must be able to defend and justify his given opinions, including cross examination from his fellow specialists.

1.4.3 A Revised Definition and its Implications

Taking all these matters into consideration, "medical evidence" may mean any or all of the following: subjective assessment provided by patients (pain, depression, etc), directly observable/measurable evidence (state of emaciation, symptoms of disease, etc.), indirectly observable evidence (cancerous cells revealed in X-rays, heart murmur, etc), factual records of the patients (personal and family medical history, smoking and drinking habits, etc) and clinicians' expert knowledge acquired through individual training, practice and peer sharing.

Thus, the current definition that "evidence is a fact or datum which is used, or could be used, in making a decision or judgement in solving a problem" [25] is somewhat inadequate. As such, evidence should be more appropriately defined as "an explicit warranted reference given in an appropriate and specific context for supporting or rejecting a hypothesis, claim or belief" and it encompasses any facts, data or information, whether weak or solid, obtained through experience, published results and observational and experimental research. A reference qualifies as

evidence so long as it is relevant either to the understanding of the problem or to the clinical decisions made about the case.

What is the implication of this revised definition of evidence? It suggests that all medical evidence must be organised, analysed and interpreted with the Bayesian framework. With this in mind, the Bayesian probability-encoding models are advocated in this dissertation. It is capable of coping with the unique nature of medical evidence, including a priori beliefs and expert opinions, and thus, should be recognised as the most appropriate and versatile framework for medical decision analysis and EBM practice as a whole. Through fulfilling the objective depicted earlier, this dissertation sets off to prove that the incorporation of Bayesian thinking into medical decision analysis is never an expensive or painful endeavour. Hopefully this is a welcome addition to the literature of contemporary medical practice, including that of EBM.

1.5 Contributions

The specific Bayesian models proposed in the dissertation are developed from either the original Bayes' Theorem [26] or from the various fundamental concepts that underlie the development of contemporary statistics. Considering the nature of evidences often encountered in medical decision analysis, two classes of probability-encoding models are developed. The first deals with subject-level evidence, while the second accommodates aggregate-level evidences reported in medical literature. Both are designed for routine use in medical practice.

The models developed for synthesising aggregate-level evidences may have profound implications on medical decision analysis. Clinicians spend a large proportion of their time reviewing the medical literature in search for evidential support of their actions. The published evidence or existing data from secondary sources effectively form the basis for medical decision making. These may be the quickest available "objective evidence" at hand as it is often beyond the scope of the clinicians to conduct a new observational or experimental study to justify his hypothesis or claim. Thus, the proposed random-effect hierarchical models designed for handling aggregate-level evidences is deemed to be an indispensable tool for achieving this aim. They are also capable of combining evidences from different published sources. On the other hand, the relational models that utilises patient-level evidences are also extremely helpful in situations where prior information of all the model coefficients are not available or obtainable. Instead of fitting non-informative priors to the coefficients, these models only require the most critical priors be specified in analysis. This is certainly a very attractive feature for routine probability encoding.

Next, the beta distribution is duly credited for its versatility in evidential analysis. Unlike the conventional Bayesian approach, beta is applied in this dissertation as both a prior distribution for quantifying previous/expert evidences and as a likelihood function for summarising collected data. Beginning to gain popularity among mainstream statisticians in recent years, this dissertation hopes to popularise its use in applied medical research.

On a broader perspective, the discussion of the nature of medical evidence has also helped to shape a more complete definition of evidence, the cornerstone of medical decision analysis. Conceptually, evidence refers to observational, experimental and inferential information forming part of the grounds for upholding or rejecting claims or beliefs relevant to medical decision making. Forming the basis of reasoning, evidence is thus referred to any explicit warranted reference given in an appropriate and specific context for supporting or rejecting a hypothesis, claim or belief.

The new notion of evidence could bring enormous contributions to EBM. The protagonists of EBM place case reports near the bottom of the medical evidence pyramid alongside editorials and opinions [27], even though they may be the primary source of information one can apply in some decision problems. In view of the profound implications of the Bayesian framework, the current definition of EBM [25, 28-29] must be revised and this will help EBM practitioners to recognise the practical importance of such evidence that has hitherto deemed to be falling short of the "scientific standards of proof" [27]. The proposed Bayesian probability-encoding models are able to accommodate these evidences and synthesise with those generated from randomised controlled trials, analytical observational studies and uncontrolled experiments. Such practice is desirable in view of the broader scope of evidence.

This may in turn help to shed light on some of the unresolved issues of EBM [30] and consequently, lead to a paradigm change in its practice.

Subjective medical evidence—so often intertwined with medical dogma, which is derived from untested hypotheses and uncritical assessment of research findings—bears a poor reputation and this in turn shapes the traditional scientific thinking, with empirical investigation universally recognised as the only undisputable source of evidential organisation and interpretation. However, one ought to think twice before discounting all subjective evidence in scientific investigations. In view of the earlier discussion, it must be reckoned that effective decision making draws upon a broad spectrum of clinicians' capabilities that include their shrewd application of fellow scientists' testimony. In fact, clinical instincts and independent thinking—developed through personal experience and communication with experts—are essential attributes of a competent clinician. Nothing, not even the best form of education, can replace the role of experience. It is an asset that all clinicians earnestly strive for. With experience, clinicians are able to approach problems confidently and identify feasible solutions quickly.

Summarising the views put forth above, this dissertation asserts that scientific medicine is a decision-oriented discipline about evidentiary interpretation. Clinicians are ardent users, organisers and interpreters of medical evidence. Thus, they must pay special attention to the way their decisions are formulated. This may in turn transform the way medicine is practiced in future.

Inevitably, the supreme authority of clinicians in decision making is challenged and eroded with the application of decision analysis. Although scientific medicine has always maintained that patients are fresh and blood and should be treated as such, many clinicians are often more interested in the diseases than in the

patients who suffer from the diseases. Clinicians have always had power and exclusive, if not elusive, knowledge about health issues. They possess specialised knowledge about diseases, drugs, remedies and treatments not accessible to the public at large. They have let it be thought that they know exactly what they are doing even they may not necessary be so and this may undermine patients' autonomy. Unfortunately, this is detrimental to medical care as it fails to recognise patients' preferences. Clinicians must begin to realise that their interests are intertwined with that of the patients. Moreover, patients have the basic need to explain their concerns, hopes, fears, desires and misfortunes. While clinicians are experts in healthcare matters, patients are owners of their health. They also have the right to understand every single detail about the decisions made on them. Through medical decision analysis this dissertation hopes to correct the dogmatic attitude of contemporary clinical practice, which has become more and more depersonalised in recent years.

On the technical aspect of medical decision analysis, there is a wrong perception that clinicians will not comprehend the beauty of complicated quantitative analysis and mathematically-trained professionals will not understand the profound medical practice. As such, this dissertation is prepared to enable clinicians to appreciate decision science, especially Bayesian probability encoding. Hopefully, this dissertation provides some useful ideas to meet the growing demand for the highly technical and yet easy-to-follow procedures of Bayesian analysis. Likewise, the choice of case studies featured in this dissertation should also benefit well-informed non-medical professionals who want to know more about contemporary medical science, i.e., aetiology of diseases, their signs and symptoms, and possible diagnoses and treatments.

1.6 Outline

This chapter begins with a burning issue facing the current medical practice, that is, how to ensure clinicians make good decisions. Following the recommended routine use of structured decision analysis in solving medical problems, the objective of the dissertation is explicitly defined. Taking into account the persisting nature of uncertainties, this dissertation aims to develop a versatile framework for probability encoding useful for routine applications in the clinical context. The Bayesian framework is judged to be the most appropriate framework for quantifying the uncertainties underlying all medical decision problems, in view of the multi-faceted and profuse nature of medical evidence. A revised definition of medical evidence is also given in an attempt to accommodate a broader evidential scope, and this in turn lends support to the application of Bayesian models in decision analysis.

A systematic review of the proposed Bayesian modelling framework and all related philosophical and technical issues are given in the next chapter. The general aspects of Bayesian analysis is reviewed in the first two sections, followed by the specific modelling strategies related to the Bayesian probability-encoding models to be developed and applied in the dissertation. These include the generalised linear model, survival model, hierarchical model and meta-analysis. An overview of the computational issues often encountered in Bayesian analysis is given. It also provides some clarification to the controversy of the Bayesian framework in scientific research.

Then, the specific Bayesian probability-encoding models are developed in Chapter 3. They are designed for different types of evidence collected for decision analysis. As described before, there are two such classes of models. The first is designed for analysing subject-level evidences while the second helps to synthesise aggregate-level or published evidences. The reason for not considering the empirical

Bayes technique for handling aggregate-level evidences is presented. .In addition, issues concerning Bayesian model evaluation are discussed.

In Chapter 4, the models are illustrated with 10 clinical applications involving patient-level as well as aggregate-level evidences. The studies cover several common diseases and medical conditions in Singapore and these include depression, osteoporosis, colon cancer, dengue fever, intracerebral haemorrhage (stroke), obesity, ischaemic heart disease, asthma, end-stage renal failure and breast cancer. Some of these illnesses are regarded as the major causes of death among Singaporeans. In terms of medical disciplines, the case studies cover psychiatry, public health, oncology, infectious disease, ophthalmology, respiratory medicine, surgery, nephrology, emergency medicine and cardiology.

The final chapter is devoted to the discussion of the nature of scientific medicine and the future practice of EBM. Several related philosophical questions, such as the nature of medical truth and the correspondence between knowledge and truth, are surfaced and discussed based on the proposed probability-encoding framework. A number of future methodological research topics are also presented.

Readers may realise that all views are expressed and addressed in the context of EBM. This stance is shaped by the following reasons. First, EBM explicitly highlights the importance of medical evidence, which is viewed as the cornerstone for medical practice and decision making. As such, all discussion concerning the use of medical evidence must make reference with EBM. Second, EBM is fast becoming an encompassing field that integrates clinical practice with decision analysis and public health. As a budding field in the medical discipline, EBM will serve as a good testing ground for new developments in decision analysis, especially in the area of probability encoding.

CHAPTER 2

LITERATURE REVIEW

2.1 The Bayesian Framework

Contemporary medicine is perceived as a probabilistic activity [26]. Probability encoding in medical decision analysis clings on the availability, collection, organisation and interpretation of relevant medical evidence. Uncertain, truncated and obscure in nature, medical evidence seldom exist in isolation. Medical-evidence seekers must consciously embark on an intriguing investigative process to unlock the latent relatedness among bits and pieces of elusive clues that are often inadvertently tampered, under-utilised or suppressed. One needs to emancipate evidence from all forms of confinement before its hidden meaning becomes interpretable, albeit a provisional or incomplete one.

To discover or unearth its meaning, one must follow the rules of systematic inquiry which may be loosely described as scientific methodology. Offering a systematic framework in which collected evidences are organised, the Bayesian methodology seeks to interpret the obscure evidential meanings based on the union of two distinct sources, which adequately reflect the data-capturing process and the salient nature of medical evidences. The details are given below.

In applying evidence to make medical decisions, one effectively conducts investigations on some unknown parameter, say θ. Statistically speaking, a parameter is an unknown quantity that characterises the features of a population where evidences are drawn. An example is the extent of transmission of foot-and-mouth disease among school children within a city over a period of one month. In the context of

clinical trials, the parameter could be the difference in survival rate between two groups of patients who are randomised to receive different therapeutic treatments.

Note that θ may be a vector with multiple component parameters investigated simultaneously. Following the celebrated Bayes' Theorem [26], the proposed framework may be formulated as:

$$\begin{aligned} P[\theta \mid \text{evidence}] &\propto P[\theta] \times L[\text{evidence} \mid \theta] \\ \text{posterior} &\propto \text{prior} \times \text{likelihood} \end{aligned} \tag{2.1}$$

The prior distribution, $P[\theta]$, summarises what is believed, aware or known of θ before observing the collected evidence. The likelihood, $L[\text{evidence} \mid \theta]$, contains evidence provided by observations, given a probability model with θ as the parameter. The posterior distribution, $P[\theta \mid \text{evidence}]$, gives the final analysis and interpretation of θ after observing the evidence. The parameter θ is considered as a random variable since one is not certain about its "true" value.

Intuitively, the Bayesian approach suggests that the prior evidence support fuses with the data support (likelihood) to produce the posterior evidence support. With more evidence built into the analysis, one expects the Bayesian framework to be more appropriate and useful than the conventional framework, which considers the likelihood of collected evidence as the only basis for analysis. There is a rich volume of well-cited theoretical and methodological literature on the conventional framework [31-34]. Statisticians often refer to the conventional framework as the frequentist or classical approach.

The following summarises how the Bayesian approach is implemented in evidential analysis:

- select the most relevant and appropriate probability model for the problem
- specify the joint probability distribution for all quantities (observed and unknown) in the problem
- use prior evidence explicitly as part of that specification
- condition on the observed evidences, compute the conditional probability of the unknown quantities of interest
- evaluate the model

These are the premises upon which Bayesian evidentiary organisation, investigation, analysis and interpretation are based. Collectively, the steps serve as the conceptual framework for building advanced statistical models for analysing the association between variables, which is the crux of probability encoding in most medical decision problems.

Another way of dealing with an uncertain event is to form its odds. The odds of an event (A) is defined as the probability of A happening divided by the probability of A not happening, i.e., $\text{odds}(A) = \frac{P(A)}{1-P(A)}$. It is easy to prove that $P(A) = \frac{\text{Odds}(A)}{1+\text{Odds}(A)}$, thus illustrating the one-to-one correspondence between odds and probability (p). However, while $p \in [0,1]$, $\text{odds} = \frac{p}{1-p} \in [0, \infty)$. In the medial context, it is sometimes more helpful to inform the patients what are the odds of suffering from a complication should they decide to receive a particular medication.

According to the logic of Bayes' Theorem [26],

posterior odds = prior odds × likelihood ratio

(2.2)

The likelihood ratio is often referred to as the Bayes factor (B). It contains the evidence relevant to the question about the occurrence of event A. As readily seen, B=posterior odds / prior odds, or the amount of evidence that changes the prior odds to the posterior odds. If B>1, then the evidence has made us believe that event A is more probably to happen than we first thought. On the other hand, if B<1, then the evidence has given us more reasons to believe that event A is less probable to occur than we originally perceive.

In most clinical studies, the aim is to ascertain if there is an association between exposure to a factor E (say, following a medication plan) and the prognosis (R). The subjects face 4 possible scenarios:

- exposed (E) and recovered (R)
- exposed (E) and not recovered (~R)
- unexposed (~E) and recovered (R)
- unexposed (~E) and not recovered (~R)

One is then able to compute two odds, namely odds(R | E) and odds(R | ~E). The ratio of these odds is known as odds ratio (OR). If OR is unity, one reports that the exposure (E) is not associated with prognosis (R). If OR>1, then one claims that the patients benefit from the exposure (E). On the other hand, the exposure brings negative impact to the patients if OR<1. As such, OR allows direct comparison of the odds of recovery (R) between the exposed and the non-exposed groups. One may also compute the OR for ascertaining the relationship between exposure to a harmful

agent and the onset of disease or the association between a treatment and the status of mortality.

If the evidences about an OR are available, then one is able to form the likelihood of OR with a suitably chosen distribution. Suppose also that some prior of the OR is obtained. The following can be formulated according to the Bayes' Theorem [26]:

$$\text{posterior OR} = \text{prior OR} \times \text{likelihood OR} \tag{2.3}$$

It is useful to clarify here that it is sometimes difficult to encode probabilities directly from statistical models supported by patient-level or aggregate-level evidences, so it may be more relevant in some clinical contexts to present the odds ratio (OR) instead. In the example presented, the odds ratio is the ratio of odds for two different events that differ only in one variable (E). In advanced statistical modelling it is possible to include multiple variables for analysis.

2.2 Some Insights

Although the Bayes' Theorem [26] has been applied in statistical inference for more than two centuries, the Bayesian interpretation of probability is a fairly recent endeavour. The following discussion provides a thorough inspection of the nature and features of the proposed Bayesian framework. Supported by literature review and some generalisations from current philosophical thinking, the discussion focuses on the framework's conceptual set-up, evidential-updating property, nature of inferences, probabilistic interpretations and the benefits in evidential analysis. The literature review and its following discussion provide the impetus for advocating the application of Bayesian analysis in medical decision making. However, the discussion is not entirely one-sided. Some of the common problems and criticisms concerning the application of Bayesian models are also highlighted. Inevitably, the discussion also draws some comparison with the conventional framework of evidential analysis.

First and foremost, the Bayesian framework's evidential-updating property reflects how knowledge is accumulated and is very much in agreement with the hermeneutic circle [35-36]. Hermeneutics is a philosophical concept of interpretation and understanding of phenomena. One always forms an incomplete picture of the phenomenon, when observed, with his subjective horizon of understanding (prior). Through observed evidence, the subject develops a revised understanding of the phenomenon and the final interpretation is achieved with the fusing of the subjective and objective horizons. The procedure allows one to change his probability assessment after observing or obtaining new evidence. The final understanding (posterior) incorporates the subject's pre-understanding and his revised understanding of the phenomenon. This Bayesian property requires the new evidence be

incorporated by a process of "refute and rescale" [37]. By allowing prior evidence be integrated with observed evidence, it offers great merit in medical decision making.

The incorporation of a prior in evidential investigation immediately suggests that Bayesian analysis is at odds with the more established conventional framework, where scientific objectivity is given a paramount status and should be preserved at all costs. In fact, the objectivity of the conventional scientific approach is achieved by disregarding all forms of prior knowledge about the phenomenon under investigation. In practice, there are usually some reliable priors, say based on expert opinions, that can be quantified. This is common in legal investigation where eyewitness testimony is often the primary source of information that the court must consider in order to reach a verdict [27].

Inevitably, evidential interpretation based on the Bayesian framework is subjective in nature, as it depends on the subject who initiates the investigation. While this may seem at odds with our conventional understanding about mathematics and science, the subjective nature of probability is not new to theorists. In fact, it is now widely accepted as the modern view [38]. This view was contributed by several forefront mathematicians and statisticians [39-41]. The treatise of de Finetti (1930/1974) [39] begins with the provocative statement that "probability does not exist". This means that probability never exists in an objective sense. Rather, probability exists only subjectively within the minds of individuals. The view is also shared by Ramsey (1950) [40], Savage (1954) [41] and Anscombe and Aumann (1963) [42].

As such, the interpretation of probability as a long-run relative frequency is only one of the interpretations. Based on degrees of belief, the Bayesian decision analysts interpret probability as a quantified judgement of an individual. This notion

has some profound impact on our interpretations of evidential investigation and statistical modelling.

Despite the works cited above [39-42], the conceptual framework of Bayesian is not universally accepted in statistical science and the debates between the Bayesians and their critics, notably the conventional statisticians, shape the history of development of the subject. While conventional statisticians agree with the implication of the Bayes' Theorem [26], they generally do not accept subjective evidence, other than the likelihood function, as a source of information for inference. As a result, the conventional approach makes no room for the use of subjective evidence and is severely limited in the context of decision analysis. On the other hand, the Bayesian framework makes use of all available information and leaves no room for data omission in analysis. The prior evidence, if available, reflect the available knowledge about the phenomenon under investigation before the collected evidence is obtained. There is a large volume of literature on the elicitation of prior distributions. See references [43-46] for details. However, it is worthwhile to note that the issue of probability elicitation is not free from controversy [47-48] and the assignment of priors is viewed as a "critical issue" in all inductive inference [49]. On a practical ground, it is true that the prior distribution is difficult to specify reliably, despite the fact that complete ignorance or absence of prior information may not exist [50].

Some Bayesians argue that evidential interpretation with prior assignment may not be as private as it seems. The priors may be formulated in unambiguous mathematical terms and communicated to others. Subjective priors may be obtained from a team of experts with guidance from trained analysts. In addition, the details of prior elicitation may be documented and reported alongside the main analysis. It is crucial to note that the Bayesian framework is also not as deterministic or dogmatic as

it is perceived because the prior is described in probabilistic terms. It reflects the variable degree of uncertainty involved in using prior evidence.

Moreover, the prior distribution may also be generated from relevant evidence obtained from past studies. In this sense, Bayesian analysis is "objectified" with the use of an objective prior [51]. One may also perform Bayesian analysis using a constant prior distribution for the unknown parameters, as suggested by Reverend Thomas Bayes (1702-1761) and Marquis Pierre-Simon Laplace (1749-1827) [26, 52], the earliest Bayesians. A recent article shows that objective Bayesianism does allow learning to be facilitated from experience [53], a long-lasting criticism from the devoted subjective Bayesians who interpret probability as "the degree of belief" or "quantified judgement of individuals". In actual fact, it is difficult to follow the subjective school strictly as most subjectivists do make at least some use of objective Bayesian methods in practice [54]. This prompts some researchers to believe that the objective Bayesian methods offer the most promising route to unify the Bayesian and the frequentist frameworks [55]. The chief exponents of the objective school are Jeffreys (1965) [56], Jaynes (2003) [57] and Berger (2006) [54]. A full discussion of the subjective and objective Bayesianism can be found in references [58-67].

Despite its reliance on prior evidence and distributions, the Bayesian framework also offers a solution to evidential analysis in situations where there is no prior evidence. As mentioned before, the analyst may choose a distribution that has little or no influence on the likelihood [68]. Such priors are known as non-informative. Not surprisingly, the result is identical to that produced by the likelihood alone. This is because with no prior knowledge, the posterior is solely based on evidence summarised by the likelihood. That means, Bayesian analysis is applicable

in situations where there is no available prior knowledge about the phenomenon. Thus, the Bayesian framework offers a very versatile approach for decision analysis.

The Bayesian framework also generates more intuitive and meaningful concepts for inductive inference. As mentioned before, population parameters (θ) are specific to the decision problem and are not generally subject to random variability. According to the conventional framework, population parameters are uncertain only because of lack of knowledge. Therefore, they are not recognised as random and all probabilistic statements about them are deemed to be meaningless. However, the Bayesian framework asserts that it is perfectly legitimate to make probabilistic statements about the parameters simply on the ground that they are unknown [69]. Thus, a quantity is regarded as a random variable even when its uncertainty is not due to randomness but to imperfect knowledge. In the context of hypothesis testing, the probability-value generated from the conventional approach does not say how likely the null hypothesis is based on the collected evidence. On the other hand, the Bayesian framework is able to attach a valid probabilistic statement and hence a more direct interpretation about how plausible the hypothesis might have been in light of the evidence. Obviously, the Bayesian approach offers a more intuitive interpretation of unknown quantities.

According to the conventional methodology, the only source of uncertainty admitted to analysis is sampling uncertainty. A fundamental advantage for applying the Bayesian framework to decision analysis is that both prior and posterior estimates are described in probabilistic terms and therefore offers a more realistic procedure for dealing with the myriad sources of uncertainty faced by decision makers in real-life applications. As a result, the Bayesian framework is well-suited for decision making. What makes decision hard is uncertainty. The Bayesian framework can quantify

these uncertainties using subjective or personal probabilities. This quantification of uncertainties may be seen as a crucial component of rational, evidence-based decision making [69].

The Bayesian theory provides a solution to the famous Hempel's paradox [70], which discovers a serious challenge to inductive logic. Implying the arbitrariness of all human knowledge, the paradox highlights that inductive logic violates intuition and is bound to result in various absurdities. According to the paradox, the evidence (E) that an object is a non-black non-raven confirms the hypothesis (H) that every raven is black. The standard Bayesian resolution suggests it is to a minute degree that E confirms H [71]. However, the argument is based on an assumption that the probability of H should not be affected by evidence that an object is non-black. A recent resolution shows that this assumption is not plausible, but the Bayesian concept is still able to cope with the paradox [72].

On the philosophical ground, Bayesian estimators are found to be "theoretical simple" according to the Minimum Message Length (MML) [73], thus confirming the Bayesian practitioners' persistent claim that Bayesian analysis is elegant [69]. The Bayesians are now equipped with a wide range of established methods for handling diversified issues in data analyses, and these include sample size determination [74-77], point estimation [78], probability computation [79], hypothesis testing [80-84], clinical trial design and monitoring [85], model evaluation [86-87] and statistical modelling [88-90].

Applications of Bayesian methods in biomedical research can be found in well-cited journals like Bayesian Analysis, Biometrics, Journal of the American Statistical Association, Statistics in Medicine and the various Journals of the Royal Statistical Society.

2.3 An Overview of Bayesian Models

The different Bayesian models to be reviewed may be classified in terms of their set-up and the nature of data analysed. This will in turn determine what research questions can be answered by the models. Loosely speaking, a relational model (more commonly known as regression in the statistical literature) is applied to quantify the association between an outcome variable Y with a vector of covariates or predictors (**X**). The following discussion adheres to the convention that a random variable be denoted by a capital Latin letter and its value by a corresponding lower case letter. All Greek alphabets represent unknown parameters or coefficients to be estimated.

If properly modelled, these relational models enable one to generate reliable predictions about the outcome in probabilistic terms. The outcome concerned may be continuous (symmetrical as well as non-symmetrical), categorical (e.g., counts and nominal) and censored (e.g., time to event). In situations where evidences are clustered or collected repeatedly over time, one of the most critical modelling assumptions called i.i.d. is violated, and some kind of hierarchical or multi-level relational models must be applied. The assumption of i.i.d. refers to the situation where the observations are independently and identically distributed. Last but not least, if evidences from various sources are to be synthesised to address a research question or hypothesis, a meta-analysis is conducted.

2.3.1 Generalised Linear Model

In the context of Generalised Linear Model (GLM) [91], which underlies most of the statistical analyses in biomedical research, the observed outcome variable Y is assumed to be generated from a probability distribution belonging to the exponential

family that includes binomial, Poisson, normal, gamma and inverse-Gaussian, etc. It is applicable to a wide class of qualitative and quantitative outcome variables (counts, rates and continuous outcomes). GLM aims to unify a large family of statistical models that are applied for relational analysis.

Consider a sample of n independently and identically distributed (i.i.d.) observations or measurements. The GLM model requires the expected value of the underlying distribution, μ_i. be dependent on covariates $\mathbf{x_i}$ (i=1, 2.,…, n) through $E[Y_i]=\mu_i=\kappa(\mathbf{x'}_i\boldsymbol{\beta})$, where κ is known as the link function and $\boldsymbol{\beta}$ is the coefficient vector that quantifies the association between $\mathbf{x}_i$ and y_i. The idea is to estimate $\boldsymbol{\beta}$ so that the magnitude and direction of the association between $\mathbf{x}_i$ and y_i can be deciphered. The variance of Y_i is a function of the mean such that $V[Y_i]=V[\mu_i]=V[\kappa(\mathbf{x'}_i\boldsymbol{\beta})]$.

In summary, the GLM has three components:

- a distribution from the exponential family
- a linear component $\eta_i=\mathbf{x'}_i\boldsymbol{\beta}$
- a link function such that $E[Y_i]=\mu_i=\kappa(\eta_i)$

The link function provides the relationship between the linear component (η_i) and the mean of the distribution function. There are many commonly-used link functions, and the choice should reflect the nature of the outcome variable and the desired interpretation of $\boldsymbol{\beta}$.

Under the conventional modelling framework, $\boldsymbol{\beta}$ is estimated with the maximum likelihood (MLE) method. As the name implies, the iterative procedure seeks to identify the solution that maximises the likelihood function. Most statistical

software implements the Newton-Raphson algorithm for the root-finding problem. Theoretically, the MLE of β is asymptocally normally distrbuted.

On the other hand, an appropriate prior distribution for β must be fitted should the model be analysed with the Bayesian framework. Data are then collected to update the prior specifications and the resultant posterior serves as the basis for inference. One advantage of the Bayesian framework over its conventional counterpart is that knowledge from previous sources or derived from theoretical considerations may be incorporated into the model. For example, one might want to restrict the signs of certain covariates in the model. Such knowledge, if incorporated as "informative priors", may help to improve the precision of the estimates of β. Detailed discussions on model specification, estimation, hypothesis testing, model selection and diagnostics for various parameters of interest from a Bayesian point of view is now available in references [92-93].

While theoretically appealing, finding reliable priors for fitting Bayesian GLM is a daunting task. Very often in practical situations, non-informative priors are fixed and the resultant posterior distribution is essentially dominated by the likelihood. As such, the Bayesian estimates are numerically identical to the conventional estimates. Nevertheless, more complex model specifications and state-of-the-art computational methods are explored in recent years to meet the needs of biomedical and industrial research. For example, flexible semi-parametric approaches are now available to model the link functions [92]. Splines are used for handling nonlinear covariates [94]. In addition, some robust procedures can also be incorporated to provide reliable estimates in the face of outlying or influential observations [94].

There are a number of other excellent Bayesian references that address the modelling issues (such as the popular linear regression model) in a more accessible manner [80, 95-96]. While their presentation is not necessary adhering to the GLM approach, the discussion does provide readers with some glimpses of the power of the related models.

2.3.2 Survival Model

The next relational model for analysing i.i.d. subject-level data deals with time to event data. Commonly known as survival analysis in biostatistics [97-98], such models have one salient feature: the time to event is skewed and censored. Generally, the term "censoring" refers to an individual's time is partially observed and not followed through to its completion [94]. Such situations include patients' premature withdrawal from the study, or simple because the outcome is not observed before the study ends.

Suppose an individual's time to event, y_i, follows $f(y_i, \theta)$. The cumulative distribution function (cdf) is $F(Y<y_i)$. There are two scenarios at a particular time, say y_i. First, there are patients with an event reported (death, readmission, etc.) and they form the risk set. In addition, there are patients who do not encounter an event by time y_i and they are represented by the survivor function $S(y_i)=P[Y\geq y_i]=1-F(Y<y_i)$, which reports the probability of surviving beyond time y_i. These subjects are "censored" because no event is observed in them. To examine the effects of the subjects' multiple covariates on their survival the hazard function is derived. Analogous to the death rate in discrete time, the hazard rate for an event at time y_i is defined as $h(y_i)= \lim_{\Delta y_i \to 0} \frac{F(y_i + \Delta y_i) - F(y_i)}{y_i S(y_1)} = \frac{-S'(y_i)}{S(y_i)}$ [98]. The hazard rate is at the

heart of modern survival analysis [99]. One models this hazard rate as a function of the baseline hazard at time y_i and the effects of the identified covariates, which may be fixed throughout the observation period or time varying [94]. The hazard rate may also be interpreted as the instantaneous event rate or conditional event rate [99].

Thus, the likelihood is made up of two components, namely the non-censored $f(y_j)$ and the censored $S(y_j)$. The censored component is in turn made of three possible portions: right-censored, left-censored and interval-censored. Right censoring is most common in biomedical research. Putting these components and requirements together, the general likelihood function may be written as $L(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\beta}) =$

$$\prod_{i \in UC} P[Y = y_i \mid x_i, \boldsymbol{\beta}) \prod_{j \in RC} P[Y > y_j \mid x_j, \boldsymbol{\beta}) \prod_{k \in LC} P[Y < y_k \mid x_k, \boldsymbol{\beta}) \prod_{m \in IC} P[y_{m,0} < Y < y_{m,1} \mid x_m, \boldsymbol{\beta}) =$$

$$\prod_{i \in UC} h(y_i \mid x_i, \boldsymbol{\beta}) S(y_j \mid x_j, \boldsymbol{\beta}) \prod_{j \in RC} S(y_j \mid x_j, \boldsymbol{\beta}) \prod_{k \in LC} 1 - S(y_k \mid x_k, \boldsymbol{\beta}) \prod_{m \in IC} S(y_{m,0} \mid x_m, \boldsymbol{\beta}) - S(y_{m,1} \mid x_m, \boldsymbol{\beta})$$

where UC, RC, LC and IC refer to the subsets of uncensored, right-censored, left-censored and interval-censored observations, respectively. Interpreted as hazards ratios, the effects $\boldsymbol{\beta}$ are estimated by maximising the above-mentioned likelihood function. In practice, the likelihood function is less complicated because the study designs may only allow at most one type of censoring.

A particular sub-class of the survival or time-to-event models is the proportional hazards model [100], under which the mean function of the covariates is independent of the time function. This is a simplifying assumption applicable to most survival densities [94]. Under the proportional hazards assumption, the hazard ratio is constant over time, provided that the covariates do not change over time [101]. For example, if taking drug A halves a patient's hazard at time 1, it will also halve his hazard at time 2. Sir David Cox (1924—) shows that if the assumption of

proportional hazards holds, then the effects (β) can be estimated without considering the hazard function [100].

The common time-to-event distributions for Bayesian analysis are exponential, Weibull, gamma, Gompertz, log-normal, log-logistic and extreme-value, which are compatible with the proportional hazards assumption. Bayesian survival analysis begins with the specification of priors on the parameters of the chosen distribution. The model specification is usually very complicated, in view of the nature of the problems encountered in survival analysis and the choice of distributions.

As in the case of GLM, Bayesian survival analysis is becoming popular in recent years and a large number of research activities is conducted. For example, it is well-known that the selection of a particular model may be subjected to errors, so a Bayesian averaging process is proposed [102]. Model based on other lifetime distributions such as Pareto may also be considered [103]. Attempts to produce more flexible specification with piecewise hazard models are presented in reference [104]. Following the proportional hazards assumption, these models utilise several types of nonparametric/semi-parametric prior processes [101, 104]. Thanks to the latest development in computational algorithms, these models can be implemented efficiently in practice.

It is legitimate to assert that Bayesian survival models may be presented under the wider GLM framework. However, survival analysis has now become common and well accepted in practice that it has cast in a language all its own [99]. Bayesian survival analysis has evolved into an independent area of theoretical research in its own right.

2.3.3 Multi-Level Model

Some medical evidences are measured over time or clustered. In the former case, the outcome values are no longer independently and identically distributed (i.i.d.) as they are contributed by the same subject at different time points (e.g., follow-ups) [105]. Such studies are usually known as longitudinal design with repeated measurements. With two or more levels of observations in the clustered scenario (multi-centre trials, studies involving paired observations, etc.), the outcome values are likely only to be independent conditional on the clusters.

The immediate implication is that the issue of dependence within clusters must be modelled appropriately with a suitable likelihood. That is, care must be taken for constructing the hierarchical data structure, with cluster-specific parameters and some covariance matrices incorporated into analysis. These cluster-specific parameters are usually assumed to be random effects either drawn from independent distributional functions or from some multivariate distributional functions.

Consider outcome y_{ij} related to predictors $\mathbf{x}_{ij}$ for observations i=1, 2, …, n_j within clusters j=1, 2, …, k, such that $\sum_{j=1}^{k} n_j = n$. While the clusters are likely to be independent, the evidences within each cluster are not. In clinical studies involving paired observations, say vision and kidney failure, the first and second level data refer to the paired observations and individual patients, respectively. In the case of multi-centre clinical trials, the first and second levels are individual patients and their belonging centres, respectively. While there is no theoretical restriction to the number of levels that can be specified in the hierarchy, the practical restriction is that in specifications that have greater than three to four levels, the interpretation of the estimated coefficients can be challenging [96]. Frequently, there is no good reason to go beyond a two-level model [96].

A surge in biomedical research activities involving such hierarchical or multi-level models with the frequentist perspective is observed in recent years [105-110]. The Bayesians are quick to respond and a number of excellent references are now available [92, 94, 110-111]. These hierarchical models may be perceived as an extension of the more established GLM framework [92, 112]. Many survival models may also be carried out with the hierarchical modelling strategy [113].

In reality, many Bayesian models exhibit a hierarchical structure in specification. This is because the underlying likelihood function may contain a number of parameters, thus requiring a series of priors to be fitted. Collectively, the likelihood and the priors form a hierarchical structure in the model [80].

2.3.4 Meta-Analysis

The conventional approach of systematic literature review in EBM involves discussion of results from a number of published studies that have investigated a common question. Such review considers the evidences from individual studies one at a time. There are some obvious drawbacks to this approach. Evidence from individual isolated studies may be inconclusive because they lack power. Moreover, the studies may differ in quality in terms of their sample sizes and rigor of analysis. This may be resolved by pooling the evidences with a suitably-chosen weighing scheme that quantifies the quality of each selected study.

In statistics, a meta-analysis combines the results of several studies that address a set of related hypotheses [114-115]. The combined effect could be a risk ratio, odds ratio, continuous outcome or probability concerning some uncertain events (death, recovery, relapse or development of a complication). Ideally, the studies considered in a meta-analysis should be similar in terms of the population of

subjects/patients, end-point outcomes, study designs (case-control, cross-sectional, cohort, randomised clinical trial), nature of treatments and statistical analyses. However, it is very rare for two published studies to be identical in all aspects. Consequently, certain inclusion criteria must be defined so that the selected studies are "sufficiently similar" and the pooled evidence are meaningful and generalisable. This may help to minimise the effects of clinical heterogeneity [116-117].

The biggest challenge for all meta-analysis concerns the availability of unbiased published evidences. The most potential source of bias concerns the publication process. Studies which report a relatively dramatic result are more likely to be published in journals and cited in other relevant publication [118]. The dire consequence is obvious. One may be deprived of the less dramatic but accurate results and the combined evidence may then give a distorted picture concerning the significant effects of a therapeutic treatment.

Nevertheless, meta-analysis remains a realistic approach for quantifying systematic review of biomedical studies. If properly conducted, it may provide EBM practitioners a very useful and least expensive solution to many medical decision problems. Conventionally, there are two approaches to conducting meta-analysis. The first assumes no obvious heterogeneity in the selected studies. In contrast, the random-effect methods consider the effects to vary randomly about a population. Usually, this approach is more realistic in view of the large degree of heterogeneity in reported studies. Moreover, the common statistical method for detecting heterogeneity often lacks power and may report no significant heterogeneity even if the selected studies differ in many aspects. The conventional approach to meta-analysis is found in reference [119]. It is also possible to cast the meta-analysis in a regression approach [120].

While meta-analysis is an indispensable tool for systematic reviews, the Bayesian paradigm is not fully recognised in EBM practice. At the point when the dissertation is prepared, there is no published textbook on Bayesian meta-analysis. While the Bayesian updating formula is often cited and applied in encoding probability for decision analysis [6], the approach is often too simplistic that it fails to take into consideration the nature of underlying heterogeneity in reported medical evidences. Such desired models are multi-level in nature as the various sources of heterogeneity are captured and described as random effects, thus exhibiting a hierarchical structure in model specification [14, 121].

2.4 Conjugacy and Monte Carlo Markov Chain

Given a model with the prior and likelihood determined, the computational phase of the Bayesian inference requires a practical method for summarising the posterior distribution. In simple cases, the estimators for unknown posterior parameters may be obtained analytically after some tedious mathematical manipulations based on integration. However, this is generally not the case for most Bayesian analyses. Usually, such posterior is mathematically non-tractable. Thus, analysts must rely on advanced simulation techniques for providing the solutions.

It follows that the choice of distributional forms for priors and likelihood is a critical feature of Bayesian analysis. It is well known that the posterior distribution might not have an analytically tractable form if the priors are freely chosen. A way to guarantee that the posterior has a calculable form is to specify a conjugate prior. When the posterior has the same distributional family as the prior, one says that the prior and the likelihood distributions are conjugate. While adequately elegant and computationally simple for expressing an analyst's opinion, it may be worthwhile to point out that the use of conjugate priors has no real theoretical advantage. Before 1990s, conjugacy was crucial to the ability to apply Bayesian methods because non-conjugate priors usually lead to posteriors that are not analytically tractable. With the advent of the Markov chain Monte Carlo (MCMC) techniques [122-128], this limitation becomes greatly reduced.

It is worthwhile to provide a brief review of MCMC here, as EBM practitioners will find it extremely useful for conducting Bayesian analyses. The MCMC technique provides an approach for simulating the posterior distribution in complex multi-parameter scenario without resorting to integration techniques or a search for close-form solutions. In short, MCMC is used to estimate integrals in high

dimensions. Instead of making direct simulation from the posterior distribution, MCMC simulates values from a stationary Markov chain, which describes an idealised pattern of movement or transitions through a set of states. As the name implies, the process moves from a state v at time t to state w at time t+1 depending only on state v. The method is called "Markov chain" because each generated parameter value is used to generate the next. It is "Monte Carlo" because it repeatedly simulates parameter values from the posterior distribution. As such, it offers a less painful approach for finding solutions as the posterior may be extremely complex. Stochastic in nature, MCMC runs the chain until convergence is achieved. If the chain is run for a long time, simulated values of the chain can be used as a basis for summarising features of the joint posterior (or conditional posterior) of interest. In contrast to classical simulation, MCMC generates samples where successive observations are non-independent of the previous observation.

A specific MCMC technique known as Gibbs sampler [123] is advocated in this dissertation. As the most widely used MCMC technique [96], the Gibbs sampler generates a Markov chain by cycling through the conditional posterior distributions instead of the full posterior. In many situations, it is possible to define and derive the conditional posterior of the unknown parameters, thus making the implementation of the Gibbs sampler fairly straight-forward. With the true posterior distribution of parameters as its limiting distribution [96], the Gibbs sampler converges under very mild conditions [127]. If the Gibbs sampler possesses the property of irreducibility (i.e., the existence of a path between any two points in the space), convergence is assured of the n-step-ahead distribution to the invariant distribution for almost all starting points [124]. This justifies the practical use of the Gibbs sampler to start from any arbitrary chosen initial condition and the sampler averages to approximate

integrals of the posterior. In fact, the Gibbs sampler converges at a geometric rate: the total variation distance between an arbitrary time and the point of convergence decreases at a geometric rate in time [96]. The application of Gibbs sampler will be discussed in details with the introduction of the Bayesian hierarchical meta-analysis models useful for synthesising aggregate-level evidences for probability encoding (Chapter 3).

An alternative MCMC algorithm, known as the Metropolis-Hastings, is based on the generalised rejection sampling scheme. Values are drawn from arbitrary distributions and "corrected" so that they "behave" as random observations from the target distribution asymptotically [122]. A fairly easy-to-follow discussion of MCMC and some of its practical implementation issues, such as the number of iterations, starting point determination and graphical modelling is found in reference [128].

However, conjugacy continues to be highly appreciated by analysts as it offers a systematic framework for finding priors with little pain and effort. Consequently, conjugate priors are still extremely useful in practice and they are an excellent expository tool [51]. Convenience is a powerful argument for justification of the use of conjugate priors, which serve as user-friendly representations of prior evidence. As discussed in the next chapter, it is more an art than a science in fitting conjugate priors for Bayesian statistical modelling. In practice, both conjugacy and MCMC are applied as complimentary tools. The selection is based on practicability.

Statistical theory shows that likelihoods in the exponential family of distributions always possess conjugate priors [129]. Comprising a number of widely-cited distributions like binomial, poisson, normal, beta, exponential and gamma, the exponential family's unique mathematical form means that combining an exponential family likelihood and prior will always result in an exponential family posterior that is

promised to be more concentrated (less diffuse) than either the likelihood or the prior. This is a very appealing feature of the exponential family, and it further supports the use of conjugate priors in data analysis involving GLM. One may refer to references [80, 84, 87, 94-95] for further information on how conjugate priors may be chosen with selected likelihoods belonging to the exponential family.

2.5 The Unfounded Controversy

The idea of Bayesian thinking is reasonably straight-forward. When properly presented, it is an uncontroversial probabilistic concept. It is not difficult to understand that no other approaches can provide a more unified treatment of inference and decision, while accounting for parameter and model certainty. Unfortunately, Bayesian methodology has not been universally accepted in the statistical discipline in spite of the compelling logic behind its approach. In fact, specific uses of the Bayesian concept have been the subject of continued controversy for several centuries, giving rise to a steady stream of polemical arguments in methodological science. So why do such controversies persist? Why do conventional likelihood-based statistics dominate Bayesian usage in data analysis?

There are several reasons to explain the above-mentioned controversies. First, several prominent figures in the development of modern statistics, notably Professor Ronald Fisher (1890-1962), had strong prejudices against the Bayesian ideas. See referenced for his criticism of inverse probability—the old name for Bayesian inference [130-131]. This was largely caused by the misunderstanding of the nature of Bayesian's post-data interpretations. Although the definition of probability is well accepted by almost every statistician, its interpretation or the sense attributed to it varies considerably. Despite its overwhelming popularity, the interpretation of probability as a long-run relative frequency is only one of the interpretations. Bayesian theory offers a more realistic approach in which personalised beliefs can be incorporated into the context of uncertainty, with the aim of developing rules and procedures for consistent and convenient decision making. The probabilistic statements of Bayesian analysis is interpreted as a degree of belief or a quantified judgement of the individual. It emphasises the subjective basis for analysis and a

post-data basis of inference. By allowing personalised beliefs be incorporated into the contexts of uncertainty, the Bayesian theory should be seen as the most appropriate framework for medical decision making.

A related reason for the under-appreciation of Bayesian methodology lies with the use of priors in analysis. In fact, many present-day applied statisticians and EBM practitioners are reluctant to use Bayesian methodology because of the requirement of a prior distribution. From a practical point of view, the prior is an extremely difficult requirement [69] that the investigator must meet and this imposes some costs on the use of Bayesian methods. As described earlier, others are unwilling to utilise prior evidence based on concerns of violation of "objectivity". In the case where the prior evidence reflects the personal opinions of individual investigators conducting the research, or possibly those of an expert who has immense knowledge of the subject matter, Bayesian statistics is subjective in nature. However, when the prior information is a direct result of previous studies, or when prior information reflects no knowledge about the problem at hand (non-informative), the Bayesian analysis becomes objective [54]. As such, it is not fair to allege that Bayesian methodology is purely subjective. The prior is also not as deterministic as it is perceived, given that it is presented in probabilistic terms. Moreover, EBM practitioners often find themselves devoid of usable evidence for decision making. The use of priors, based on elicited expert opinion evidence, offers a practical approach for solving problems. Following a similar line of argument, the process of prior elicitation helps EBM practitioners to collect, organise and document the thoughts of the consulted experts. In view of this, the use of priors in decision analysis is not really a cost but is actually a benefit [80].

Another important reason for the dominance of conventional likelihood-based statistics lies with the complexity of Bayesian analysis. The development of a Bayesian model requires complicated modelling specifications and tedious computations. In the case of linear regression with conjugate priors, the estimators for regression coefficients can be derived analytically after some tedious mathematical manipulations (see Chapter 3). However, this is generally not the case for most Bayesian analyses. The resultant posterior, which is the heart of all Bayesian analyses, may be mathematically intractable and extremely complicated. Thus, analysts may have to reply on advanced simulation and numerical techniques for finding the solutions. This imposes another cost on the use of Bayesian models.

Such impracticability stems from our computational deficiency. As such, there is no accident that the recent rejuvenation of Bayesian statistics coincides with the development of computer-intensive techniques. With the wide applicability of high-speed computers, the new ideas offered by Bayesian statistics have captured the imagination of researchers. These methods have a wide range of potential applications, especially in EBM, as a result of the increasing complexity of problems and data structures. Their analysis and refinement will be a formidable prospect for the EBM community in the coming years. It must be emphasised that while computers can never be as wise as people, they can explore a forest of possibilities faster than we can comprehend. To this end, Bayesian analysts utilises the latest computational breakthroughs to the fullest and this immediately makes Bayesian models extremely attractive in real-life applications. Thus, the cost of computational disadvantage is significantly reduced, thanks to the current technological advancement.

The new millennium has witnessed a burst of research activities in applying Bayesian methods to solving medical problems [14, 102-103, 113]. The appealing features offered by Bayesian analysis unleashes a revolution in data analysis and triggers powerful impulse to continue to apply such approach to problems hitherto considered forbidden and unthinkable.

CHAPTER 3

BAYESIAN PROBABILITY-ENCODING MODELS

3.1 Prelude

Adhering to the earlier discussion about medical decision analysis (Chapter 1) and the Bayesian framework (Chapter 2), a useful model for probability encoding must not only enable the analyst to organise observed evidence, but also to quantify individual judgement and opinions about the uncertain quantities. With such requirement in mind, this dissertation proposes to develop various Bayesian models for probability encoding in medical decision analysis. With subjective probabilities attached, it is believed that one may obtain more useful results and insights than if a pure empirical approach is adopted. In fact, the Bayesian approach may be recognised as the underlying or unifying philosophical theory of decision analysis.

From the gathering of data to the cross-examination of expert opinions, the Bayesian framework offers a practical methodology upon which evidence is organised and presented with meanings elucidated from the traces of clues. An acceptable statistical or methodological process should help to probe the details, discern relevant facts from baseless information, exclude the impossible from the possible and crack the useful meaning(s) of evidence.

A good probability-encoding model should also reflect the evidence-capturing process and all salient features of evidence under question. Furthermore, such model must also be simple enough for routine use in clinical decision analysis. This is, unfortunately, a view which conventional methodologists, in particular mainstream

statisticians, fail to appreciate and understand. It is also erroneous to think that a good model must be sophisticated. With this in mind, the specific Bayesian models developed in this dissertation are catered for routine use in different aspects of medical practice.

First, three relational models within the Bayesian GLM framework are developed for handling subject-level or patient-level evidence. These relational models seek to encode the probabilities for medical decision analysis based on the estimated relationship between an outcome variable of interest and a set of identified covariates or predictors. Second, if the observed evidences are presented in aggregate form, say results based on a number of independent secondary sources, the full Bayesian random-effect hierarchical models are recommended for performing the necessary meta-analysis. The observed evidences are mainly published results in journal articles. With some ideas about these results, a prior is organised and presented to combine with the published results to generate a posterior for probability encoding. The second category of Bayesian models is extremely useful in situations where EBM practitioners or medical decision makers do not have raw data at hand. The two classes of models—Bayesian GLM and meta-analysis—are developed based on the first principles and several related theorems [26, 129].

Note that nothing is said about prior elicitation. The main focus of this dissertation is to develop models for probability encoding relevant to medical decision analysis. Thus, it is beyond the scope of this dissertation to give a more detailed treatment of prior elicitation. For further information one may refer to the relevant literature [43-46].

Probability encoding, according to the protocol developed by the Decision Analysis Group, Stanford Research Institute, Stanford University [1], is one of the

steps of the full spectrum of probability assessment. Its accuracy depends largely on the communication among the analyst, domain expert and decision-maker(s) (motivating phase), unambiguous structuring of the uncertain quantity to be elicited (structuring phase) and unbiased judgement provided by the encoders (conditioning phase). In encoding probabilities, the analyst must make sure the domain expert utilises his knowledge to the fullest, anchor his assessment on the right basis, able to provide an assessment representative of the event in question and state all assumptions made. The experts should also try their best not to assign probabilities to an event based on the ease with which they can fabricate a plausible scenario that would lead to the occurrence of the event.

While the main emphasis of this dissertation is to encode probabilities from the recommended Bayesian models, subjective assessment of probability may also be sought from identified experts if there are no available data for model building. This include discrete approximation should the underlying probability distribution is continuous in nature. Upon completion of probability encoding, the decision analyst must verify that the model(s) employed are accurate, reliable, and interpretable. The results may be plotted as a cumulative distribution function and probability density function. If necessary, the entire process outlined above—motivation, structuring, conditioning and encoding—may be repeated.

To facilitate discussion, the following symbols and notations are used. $U(\bullet)$, $L(\bullet)$, $f(\bullet)$, $F(\bullet)$, $g(\bullet)$, $\pi(\bullet)$ and $\pi_c(\bullet)$ refer to the utility function, likelihood function, pdf, cumulative distribution function (cdf), prior distribution, full posterior distribution and conditional posterior, respectively. The other common functions may be written as $\kappa(\bullet)$, $\nu(\bullet)$, $\psi(\bullet)$ or $\gamma(\bullet)$. However, $\Gamma(\bullet)$ is exclusively used as a gamma function. Likewise, $S(\bullet)$ represents survival function. Greek alphabets are used for

representing parameters, with θ reserved specifically for the parameter of primary interest. $\boldsymbol{\beta}$ refers to the vector of parameters concerning a relational model; its estimator $\hat{\boldsymbol{\beta}}$ is based on conventional statistical procedure (e.g., maximum likelihood or ordinary least squares); $\boldsymbol{\beta}^*$ its Bayesian counterparts and $\boldsymbol{\beta}_0$ the specified parameter of the relevant prior. The lowercase alphabets, say a, b, u and v are used for the specified parameters of a prior distribution. The symbol p means proportion or probability. The full data design is [**y** **X**], with **X** the matrix of covariates or predictors and **y** the vector of outcome variable. Letters in bold prints refer to quantities either in vector or matrix forms. η is the systematic component or function of a relational model which connects **x** with $\boldsymbol{\beta}$. The letters n and m are sample size and number of iterations for an algorithm, respectively. To make a distinction, k refers to the number of studies selected for combining published evidences. The subscripts i and j refer to the *i*-th observation and *j*-th category, respectively. Last but not least, the postscript t refers to the *t*-th iteration of a computational algorithm.

All analyses and computations are carried out with Microsoft Excel 2000 (Microsoft Corporation, U.S.A.), R (R Project; http://cran.r-project.org/), Stata 9.0 (Stata Corporation, Texas, U.S.A.; http://www.stata.com), WinBUGS (Medical Research Centre, University of Cambridge, U.K.; http://www.mrc-bsu.cam.ac.uk/bugs/), LOTUS 2.3 (National University of Singapore and University of Wisconsin-Madison) and DPL™ 4.0 (Applied Decision Analysis LLC, U.S.A.).

3.2 Relational Modelling for Subject-Level Evidence

3.2.1 The Modelling Approach

The following discussion is devoted to illustrate how a useful conjugate prior can be found for relational models. A pdf of observed data y_i, is said to belong to an exponential family characterised by a set of multiple canonical parameters θ_i if it is presented as follows:

$$f(y_i; \theta_i) = \exp\{[\theta_i y_i - \psi(\theta_i)] + \gamma(y_i)\} \qquad i=1, 2,\ldots, n \tag{3.1}$$

where $\psi(\bullet)$ and $\gamma(\bullet)$ are known functions. The form of conjugate priors can be easily determined. The parameters θ_i are related to the model's coefficients by the link function $\theta_i = \kappa(\eta_i)$, with $\eta_i = \mathbf{x}'_i\boldsymbol{\beta}$ (systematic linear component). Here, $\mathbf{x}'_i$ is a $1\times p$ vector denoting the *i*-th row of predictors and $\boldsymbol{\beta}$ is a vector of p coefficients. These coefficients quantify the association (direction and magnitude) between y_i and $\mathbf{x}'_i$. It follows that the conjugate prior distribution for θ_i is

$$g(\theta_i) = C \exp\{uv\ \theta_i - u\psi(\theta_i)\} \tag{3.2}$$

where the normalising constant C=C(u, v) is selected such that $g(\theta_i)$ is a proper distributional function characterised by two natural parameters u and v. This in turn formulates the posterior distribution of θ_i as follows:

$$\pi(\theta_i \mid y_i) = C(v+1, \frac{y_i + uv}{1+u})\exp\{(y_i + uv)\theta_i - (1+u)\psi(\theta_i)\} \tag{3.3}$$

where the link function may now be considered as a monotonic transformation of θ_i. Thus, it follows that the Jacobian transformation from θ_i to η_i is $J(\frac{\theta_i}{\eta_i}) = \kappa'(\eta_i) \neq 0$, where $\kappa(\bullet)$ is a twice differentiable monotonic function. Consequently, the posterior for η_i is given as:

$$\pi(\eta_i \mid y_i) = C(v+1, \frac{y_i + uv}{1+u}) \exp\{(y_i + uv)\kappa(\eta_i) - (1+u)\psi[\kappa(\eta_i)]\}\kappa'(\eta_i) \tag{3.4}$$

Next, performing logarithmic transformation on both sides of (3.4) one obtains

$$\log\pi(\eta_i \mid y_i) = \log C' + \{(y_i + uv)\kappa(\eta_i) - (1+u)\psi(\kappa(\eta_i))\} + \log[\kappa'(\eta_i)] \tag{3.5}$$

where C' is a constant. Setting $\frac{\partial \log \pi(\eta_i \mid y_i)}{\partial \eta_i} = 0$, one shows that

$$\begin{aligned} &(y_i + uv)\,\kappa'(\eta_i) - (1+u)\psi'[\kappa(\eta_i)]\kappa'(\eta_i) + \frac{\kappa''(\eta_i)}{\kappa'(\eta_i)} = 0 \\ &\Rightarrow \psi'[\kappa(\eta_i)] = \frac{1}{1+u}\{(y_i + uv + \frac{\kappa''(\eta_i)}{[\kappa'(\eta_i)]^2}\} \end{aligned} \tag{3.6}$$

Hence, the mode for the posterior of η_i is derived. That means, one can generate the solution for the desired posterior based on estimates of the relational model.

The above discussion suggests that the Bayesian models for relational problem may be derived with Jacobian transformation and some simple algebraic manipulations. Unfortunately, this may not be the case for situations where the priors are specified differently or when the data structure becomes more complicated. But

such limitations should not be amplified. The specific details of Bayesian modelling are discussed in the remaining sections.

Under the generalised linear model set-up (subsection 2.3.1) [91], a generic relational model involving the exponential family designed for analsysing subject-level data is:

$$f(y_i; \theta_i, \tau) = \exp\{\nu^{-1}(\tau)[\theta_i y_i - \psi(\theta_i)] + \gamma(y_i, \tau)\} \qquad i=1, 2,\ldots, n \tag{3.7}$$

where y_i, θ_i, $\psi(\bullet)$ and $\gamma(\bullet)$ have their usual meanings, τ is the scale parameter and $\nu^{-1}(\bullet)$ is a known function. As in (3.1), the canonical parameters θ_i are related to the model's coefficients $\boldsymbol{\beta}$ by the link function $\theta_i = \kappa(\eta)$, with $\eta_i = \mathbf{x}'_i\boldsymbol{\beta}$, where $\mathbf{x}'_i$ is a $1\times p$ vector denoting the i-th row of the covariate matrix $\mathbf{X}$. Essentially, the data are connected with the coefficients $\boldsymbol{\beta}$ through η_i, the systematic linear component.

Bear in mind the primary objective for modelling $\mathbf{y}$ is to estimate $\boldsymbol{\beta}$, which quantifies the relationship between $\mathbf{X}$ and $\mathbf{y}$. By inserting η_i into (3.4) and (3.7) the likelihood function and posterior for $\boldsymbol{\beta}$ are as follows:

$$\begin{aligned}
&\text{Likelihood} \quad L(y_i \mid \boldsymbol{\beta}, \tau) \propto \exp\{\nu^{-1}(\tau)[\kappa(\mathbf{x}'_i\boldsymbol{\beta})y_i - \psi[\kappa(\mathbf{x}'_i\boldsymbol{\beta})] + \gamma(y_i, \tau)]\} \\
&\text{Posterior} \quad \pi(\boldsymbol{\beta} \mid y_i, \tau) \propto L(y_i \mid \boldsymbol{\beta}, \tau) \times g(\boldsymbol{\beta}) \\
&\qquad\qquad \propto \exp\{\nu^{-1}(\tau)[\kappa(\mathbf{x}'_i\boldsymbol{\beta})y_i - \psi[\kappa(\mathbf{x}'_i\boldsymbol{\beta})] - \frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta_0})'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta_0})]\}
\end{aligned} \tag{3.8}$$

with $\boldsymbol{\beta}\sim\text{Normal}[\boldsymbol{\beta_0}, \boldsymbol{\Sigma_0}]$ as the chosen prior. The posterior is analytically non-tractable. As such, MCMC techniques must be employed for estimating $\boldsymbol{\beta}$. With $\boldsymbol{\beta}$ estimated

and $\mathbf{x'}_i$ given, $\eta_i = \mathbf{x'}_i \hat{\boldsymbol{\beta}}$ is formulated for encoding the required probabilities for medical decision analysis. Model (3.8) may be used for handling a number of different pdfs such as binomial, normal and poisson.

In situations where $\mathbf{y}$ is categorical, say $P(y_i=j \mid \mathbf{x}_i, \boldsymbol{\beta}) = \frac{\exp(x'_{ij})}{\sum_{j=1}^{J} \exp(x'_{ij})}$, j=1, 2,.., J,

the likelihood and posterior may be formulated as:

$$L(\boldsymbol{\beta} \mid \mathbf{x}_i, \mathbf{y}) = \frac{\exp(x'_{ij})}{\sum_{j=1}^{J} \exp(x'_{ij})}$$

$$\pi(\boldsymbol{\beta} \mid \mathbf{x}_i, \mathbf{y}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\{\frac{1}{2}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\} \quad (3.9)$$

where $\boldsymbol{\Sigma}$ is the covariance matrix and $\hat{\boldsymbol{\beta}}$ may be the maximum likelihood estimator (MLE) for $\boldsymbol{\beta}$, the posterior mode. $\boldsymbol{\Sigma}$ could be the Hessian of the likelihood evaluated at $\hat{\boldsymbol{\beta}}$. It is also possible to model outcomes which are ordinal or polychotomous in nature [132].

3.2.2 Binary Counts

A special case based on the above set-up (3.9) is known as logistic regression or logit [133], where $y_i \sim \text{Binomial}[n_i, p_i]$, $p_i \in [0,1]$. Logit is arguably one of the most frequently used and widely-reported statistical techniques in medical research. With $y_i \in \{0, 1\}$, logit is suitable for predicting the probability of binary or dichotomous outcome (e.g., alive/dead, recover/relapse, improve/deteriorate, successful/fail). The link function is always chosen as $\theta_i = \log\frac{p_i}{1-p_i}$ (logit) where it has a very meaningful

interpretation—log odds. See section 2.1 for a discussion of odds. The systematic linear component is $\kappa(\eta_i)=\eta_i=\mathbf{x}'_i\boldsymbol{\beta}$, such that $\kappa'(\eta_i)=1$ and $\kappa''(\eta_i)=0$.

In the case where $p_i \in [0, 1]$ is the unknown parameter of interest and where its prior is available, the likelihood based on binomial pdf may be reformulated as:

$$
\begin{aligned}
&f(y_i \mid p_i) \propto p_i^{y_i}(1-p_i)^{n_i-y_i} \\
&= \left(\frac{p_i}{1-p_i}\right)^{y_i}(1-p_i)^{n_i} \\
&= \left(\frac{p_i}{1-p_i}\right)^{y_i}\left(\frac{1}{1-p_i}\right)^{-n_i} \\
&= \left(\frac{p_i}{1-p_i}\right)^{y_i}\left(1+\frac{p_i}{1-p_i}\right)^{-n_i} \\
&= \exp\left[y_i \log\frac{p_i}{1-p_i} - n_i \log\left(1+\frac{p_i}{1-p_i}\right)\right] \\
&= \exp\left[y_i \log\frac{p_i}{1-p_i} - n_i \log\left(1+e^{\log\frac{p_i}{1-p_i}}\right)\right] \\
&= \exp\left[y_i\theta_i - n_i \log\left(1+e^{\theta_i}\right)\right]
\end{aligned}
\tag{3.10}
$$

where $\theta_i = \log\frac{p_i}{1-p_i}$ is the logit link. As before, the systematic linear component is $\eta_i = \mathbf{x}'_i\boldsymbol{\beta} = \theta_i$. In fact, $\kappa(\eta_i) = \eta_i$, so $\kappa'(\eta_i) = 1$ and $\kappa''(\eta_i) = 0$.

Since $p_i \in [0,1]$ is a bounded quantity, the most suitable conjugate prior for the set-up is a standard Beta[a, b], where a>0 and b>0 are the shape parameters:

$$
\begin{aligned}
&g(p_i) \propto p_i^{a}(1-p_i)^{b} \\
&= \left(\frac{p_i}{1-p_i}\right)^{a}(1-p_i)^{a}(1-p_i)^{b} \\
&= \left(\frac{p_i}{1-p_i}\right)^{a}(1-p_i)^{\frac{a+b}{n_i}n_i} \\
&= \left(\frac{p_i}{1-p_i}\right)^{a}\left(\frac{1}{1-p_i}\right)^{-\frac{a+b}{n_i}n_i}
\end{aligned}
$$

$$= \left(\frac{p_i}{1-p_i}\right)^a \left(1+\frac{p_i}{1-p_i}\right)^{-\frac{a+b}{n_i}n_i}$$
$$= \exp\left[\, a\log\frac{p_i}{1-p_i} - \frac{a+b}{n_i}n_i \log\left(1+\frac{p_i}{1-p_i}\right)\right]$$
$$= \exp\left[\, a\log\frac{p_i}{1-p_i} - \frac{a+b}{n_i}n_i \log\left(1+e^{\log\frac{p_i}{1-p_i}}\right)\right]$$
$$= \exp\left[\, a\theta_i - \frac{a+b}{n_i}n_i \log\left(1+e^{\theta_i}\right)\right] \qquad (3.11)$$

Thus, the conjugate prior for θ_i is $g(\theta_i) = \exp\left[\, a\theta_i - \frac{a+b}{n_i}n_i \log\left(1+e^{\theta_i}\right)\right]$ where C=1, $a=uv$, $u=(a+b)/n_i$ and $\psi(\theta_i) = n_i \log\left(1+e^{\theta_i}\right)$, according to (3.2).

According to (3.6), the posterior mode of η_i is given by

$$\psi'(\kappa(\eta_i)) = \frac{1}{1+u}(y_i + uv)$$
$$= \left(1+\frac{a+b}{n_i}\right)(y_i + a)$$
$$= \frac{(y_i + a)n_i}{n_i + a + b} \qquad (3.12)$$

after some algebraic manipulations. Also, $\psi'(\kappa(\eta_i)) = n_i \frac{e^{\eta_i}}{1+e^{\eta_i}}$, so

$$n_i \frac{e^{\eta_i}}{1+e^{\eta_i}} = \frac{(y_i + a)n_i}{n_i + a + b}$$
$$\text{or } e^{\eta_i} = \frac{(y_i + a)}{n_i + a + b} + e^{\eta_i}\frac{(y_i + a)}{n_i + a + b}$$
$$\Rightarrow e^{\eta_i}\left[1 - \frac{(y_i + a)}{n_i + a + b}\right] = \frac{(y_i + a)}{n_i + a + b}$$
$$e^{\eta_i}[n_i + b - y_i] = y_i + a$$
$$e^{\eta_i} = \frac{(y_i + a)}{n_i + b - y_i}$$

(3.13)

Taking logarithmic transformation on both sides yields

$$\hat{\eta}_i = \log \frac{(y_i + a)}{n_i + b - y_i} \tag{3.14}$$

Hence,

$$\begin{aligned} \boldsymbol{\beta}^* &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\hat{\eta}_i \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\log\frac{(y + a)}{n_i + b - y} \end{aligned} \tag{3.15}$$

where $\boldsymbol{\beta}^*$ is the posterior estimator. It is not difficult to show that if $y_i \sim$ Bernoulli[1, p_i], a special case of the binomial distribution, then

$$\boldsymbol{\beta}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\log\frac{(y + a)}{1 + b - y} \tag{3.16}$$

where $\mathbf{X}=[\mathbf{x}_1\ \mathbf{x}_2\ \ldots\mathbf{x}_p]'$ and $\mathbf{y}=(y_1, y_2,\ldots y_n)'$, $p<n$, are the data matrices. As readily seen, the posterior estimate $\boldsymbol{\beta}^*$ is dependent on the choice of Beta[a, b], where a and b are the shape parameters. Unlike (3.9), such model produces closed-form solution for $\boldsymbol{\beta}$.

The celebrated Bernstein-von Mises Theorem [134] provides the clue to the construction of the credible or probability interval (P.I.) for $\boldsymbol{\beta}^*$. It guarantees that $\boldsymbol{\beta}^*$ are consistent and asymptotically normal. Since the posterior distribution of $\sqrt{n}(\eta$-

$\hat{\eta}_n) \rightarrow \text{Normal}[\mathbf{0}, \mathbf{I}(\eta_0)^{-1}]$ when $n \rightarrow \infty$, where $\mathbf{I}(\eta_0)$ is the Fisher's information [135] with η_0 as the true value, then the posterior distribution of

$$\sqrt{n}\,(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_n) \rightarrow \text{Normal}[\mathbf{0}, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \tag{3.17}$$

as $n \rightarrow \infty$, where $\boldsymbol{\Sigma}$ is a diagonal matrix with elements $\sigma_i^2 = n\mathbf{I}(\eta_0)^{-1}$. As such, one is able to construct the $100(1-\alpha)\%$ P.I. for $\boldsymbol{\beta}$ when n is sufficiently large. The issue of Bayesian estimators' rate of convergence is documented in literature [136]. Incidentally, Bayesian modelling with the above-mentioned strategy was also explored in reference [137] when this dissertation was prepared.

Such intervals are conceptually different from the conventional confidence intervals (C.I.). In the conventional paradigm, the interpretation is that $100(1-\alpha)\%$ of the intervals contains the "true" parameter value. Moreover, the computation is based on the sampling distribution of the underlying distribution of the estimator, or how it varies over all possible samples. It does not depend on the particular sample where computation is made. However, the Bayesian P.I. has a more intuitive interpretation. It makes a probabilistic statement about the parameter from the computed interval. Furthermore, it summarises one's beliefs about the parameter value that could possibly or credibly believed given the observed evidence. This gives a post-data interpretation as opposed to the conventional C.I.

3.2.3 Rates

Poisson regression [138], on the other hand, deals with rates, an important entity for quantifying risk in medical research. The most appropriate distribution for

rates, say y_{ij}, is Poisson[μ_i], i=1, 2, …n, j=1, 2, …, k. Here, $\mu_i>0$ is the unknown mean parameter. The appropriate link is $\log(\eta_i) = \mathbf{x}'_i\boldsymbol{\beta}$.

Given that $R_i = \sum_{j=1}^{k} Y_{ij}$ is a sufficient statistic for μ_i, such that $R_i \sim$ Poisson[$k\mu_i$], the likelihood of R_i is:

$$f(r_i; \mu_i) \propto \exp(r_i \log k\mu_i - k\mu_i) \tag{3.18}$$

which is clearly a member of the exponential family. It follows that the appropriate conjugate prior for μ_i is Gamma[a, b], where a>0 and b>0:

$$g(\mu_i) \propto \exp[a \log(\frac{k\mu_i}{k}) - bk\frac{\mu_i}{k}] \tag{3.19}$$

With the log link function, i.e., $\theta_i=\log(k\mu_i)$, the prior is $g(\theta_i)\propto\exp(a\theta_i - \frac{b}{k}e^{\theta_i})$ where a=uv, u=b/k and $\psi(\theta_i)=\exp(\theta_i)$. Take $\eta_i = \theta_i$, then

$$\psi'[\kappa(\eta_i)] = \frac{1}{1+\frac{b}{k}}(r_i + a)$$

$$= \frac{(r_i + a)k}{b + k} \tag{3.20}$$

Also, $\psi'[\kappa(\eta_i)]=\psi'(\eta_i)=\exp(\eta_i)$, so

$$\exp(\eta_i) = \frac{(r_i + a)k}{b + k} \tag{3.21}$$

Taking logarithmic transformation, $\hat{\eta}_i = \log\frac{(r_i + a)k}{b + k}$ is the posterior mode. Finally,

$$\begin{aligned}\boldsymbol{\beta}^* &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\eta}_i \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\log\frac{(r_i + a)k}{b + k}\end{aligned} \tag{3.22}$$

The use of Poisson regression requires one crucial assumption, that is, the mean and variance are equal. In most cases, this assumption may not be valid and it leads to a problem commonly known as over-dispersion. When such problem arises, it is more appropriate to use negative binomial [139] as the underlying distribution.

Suppose y_i~Negative binomial[r, p_i], then

$$f(y_i; p_i) = C_{y_i}^{r+y_i-1} p_i^r (1 - p_i)^{y_i} \tag{3.23}$$

The canonical link function is taken as $\eta_i = \log\frac{r(1 - p_i)}{p_i}$. With prior p_i~Beta[a, b], the posterior is Beta[a+r, y_i +b]. Therefore, the posterior mode is

$$\hat{\eta}_i = \log\frac{r(y_i + b)}{a + r} \tag{3.24}$$

According to (3.6), the posterior estimator for $\boldsymbol{\beta}$ becomes

$$\boldsymbol{\beta}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\eta}_i$$

$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \log \frac{r(y+b)}{a+r} \tag{3.25}$$

The $100(1-\alpha)\%$ P.I. for $\boldsymbol{\beta}$ of both the Poisson and Negative binomial regression models are based on (3.17).

The above-mentioned Bayesian estimators (3.16, 3.22 and 3.25) are applied for analysing binary counts and rates with subject-level evidences.

3.3 Models for Combining Evidences from Published Sources

3.3.1 The Generic Approach

The following Bayesian models are developed for combining the published results of k>1 studies while allowing for each study to have its own effect. Known as meta-analysis in biostatistics (subsection 2.3.4), these models are used for encoding the probabilities for medical decision analysis based on relevant reported evidences. They may also serve to produce the priors for the above-mentioned relational models.

Such models are also "hierarchical" because, loosely speaking, more than one level of likelihood and/or prior is specified. In this case, a particular observed quantity depends on an unknown parameter, which in turn follows a second-stage prior. This sequence of priors and parameters constitute a model with an extended or hierarchical structure.

When combining evidences from relevant reported studies which are conducted at different locations and in different times, it is desirable to consider the following features:

- different subject characteristics (age, racial mix, disease severity)
- different study designs (retrospective vs. prospective; experimental vs. non-experimental; randomised clinical trial vs. observational cohort; single-centre vs. multi-centres)
- different sampling schemes
- different inclusion and exclusion criteria (demographics, disease progression)
- different safety and quality considerations
- different study periods
- different end-point outcomes (30-day mortality vs. 90-day mortality)

- different reported statistical analyses (crude vs. adjusted results; univariate vs. multivariate analyses)

The idea is to combine the reported evidences (y_i) to estimate the overall effect θ—a hyperparameter. The combined evidence θ in turn facilitates probability encoding for medical decision making. It is the combined overall effect that is of primary interest in such exercises.

Unlike the relational models discussed earlier, the hierarchical model designed for handling aggregate-level evidences is not concerned with the estimated association between **X** and **y**. It is primarily concerned with the combination of all relevant y_is, which are reported in academic literature. In each reported study, a specific y_i may be generated from a relational model or some simpler analysis. But in the case where they are combined, the concern is to generate θ for probability encoding.

The proposed method for combining reported study effects, commonly known as meta-analysis in the biostatistical literature, requires the observed study effects (y_i) to vary around some unobserved/latent study-specific effect (φ_i), which in turn belong to a distribution characterised by the overall or combined effect (θ). As such, this is also a random-effect model, where the study effects vary randomly around their respective study-specific effects. Each of the study-specific effects, with its specific parameters, describes the populations where the reported study effect is generated. In a similar fashion, they are also allowed to vary randomly around a parent distribution characterised by θ. Consequently, the model consists some hierarchical structure where the observed study effects precede the unobserved study-specific effects, which in turn precede the unobserved overall effect.

Taking into account the various sources of heterogeneity underlying the reported evidences, the hierarchical structure is the most critical and desired feature of the proposed Bayesian meta-analysis model. It is theoretically wrong to ignore the intermediate study-specific effects (φ_i). The inclusion of the study-specific effects serves to capture the salient features of the underlying sources of heterogeneity described above. The model is severely misspecified should the study-specific effects are suppressed. The nature of the proposed model is depicted in Figure 3.1.

To cast the above-mentioned features in the form of a Bayesian model, the following is proposed:

$$\pi(\theta, \boldsymbol{\varphi} \mid \mathbf{y}) \propto \prod_{i=1}^{k} f(y_i \mid \varphi_i) \times \prod_{i=1}^{k} f(\varphi_i \mid \theta) \times g(\theta) \tag{3.26}$$

where $\mathbf{y}=(y_1,y_2,\ldots y_k)'$, $\boldsymbol{\varphi}=(\varphi_1,\varphi_2,\ldots,\varphi_k)'$ and $g(\theta)$ are the vector of observed effects, study-specific effects and prior of the overall combined effect, respectively. characterised by θ, (3.26) is the joint posterior distribution of interest. Once the joint posterior is derived the combined effect θ emerges simultaneously. In most instances, there may be more than one prior involved, depending on what distribution is used for representing the study-specific effects. If it is a multi-parameter distribution one may need to specify more than one prior.

The idea of specifying a random-effect model is not new in statistics. The conventional analysis of variance (ANOVA) model [34], multi-level model (subsection 2.3.3) and meta-analysis (subsection 2.3.4) may incorporate some random effects for dealing with latent variables in data analysis.

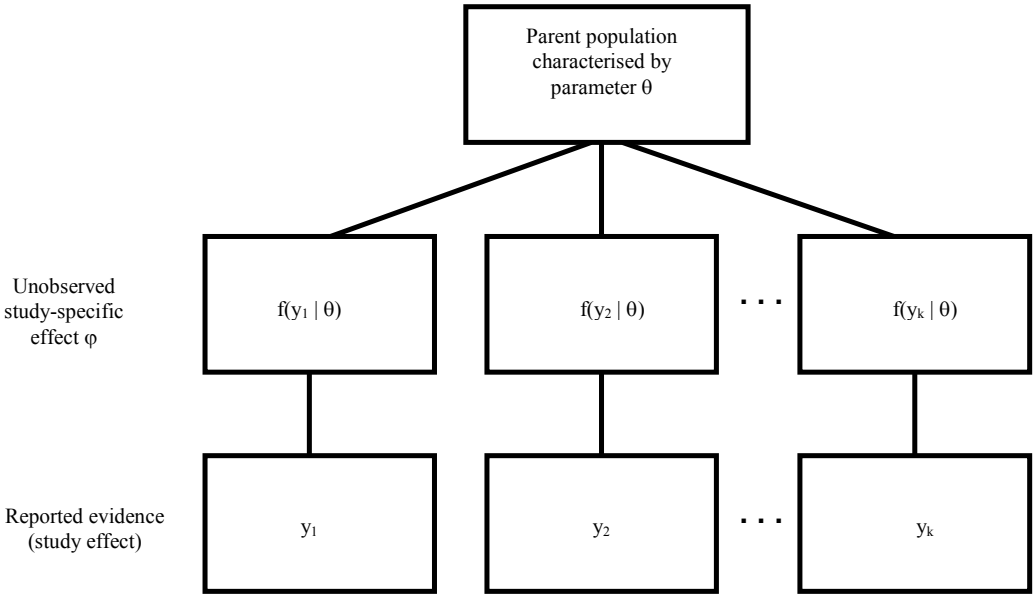


Figure 3.1: Proposed modelling framework for combining study effects

In practice, the joint posterior distribution is extremely complicated and it is sufficient to consider the conditional prior distribution of θ, given the observed study effects but does not depend on the unobserved study-specific effects **φ**. A conditional posterior distribution is the posterior for one parameter given the values of the other parameters, and is obtained from the joint posterior by treating the other parameters fixed.

Theoretically, this is $\pi_c(\theta \mid \mathbf{y}) = \frac{\int_{-\infty}^{\infty} f(y \mid \varphi) f(\varphi \mid \theta) g(\theta) d\varphi}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y \mid \varphi) f(\varphi \mid \theta) g(\theta) d\varphi d\theta}$. Following probability theory, the Bayesian estimator for T(θ) is defined as:

$$E[T(\theta)|\mathbf{y}] = \frac{\int_{-\infty}^{\infty} T(\theta) f(y \mid \varphi) f(\varphi \mid \theta) g(\theta) d\varphi}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y \mid \varphi) f(\varphi \mid \theta) g(\theta) d\varphi d\theta}.$$ This is unlikely to be mathematically tractable and to estimate T(θ), MCMC (section 2.4) must be executed.

As mentioned before, the most popular MCMC technique is Gibbs sampler [123]. For t=1, 2, …, m, where m is specified before-hand, the *t*-th step of the algorithm is

Generate θ^t given **y** and φ^{t-1} from $\pi_c(\theta \mid y, \varphi^{t-1})$
Generate φ^t given **y** and θ^t from $\pi_c(\varphi \mid y, \theta^t)$

(3.27)

Essentially, the technique generates a sequence of (θ^1, φ^1), (θ^2, φ^2), …, (θ^m, φ^m). The value of φ^{t-1} is updated for generating θ^t which in turn generates φ^t. Under general conditions, the distribution of the chain stabilises or reaches the equilibrium (limiting

distribution) as the length of the chain increases. Furthermore, the arithmetic average $\frac{\sum_{j=1}^{m} T(\theta)}{m}$ converges to $E[T(\theta)|\mathbf{y}]$ as $m \to \infty$.

The strength of the proposed hierarchical model is that it leverages on relevant studies conducted in the past. It is often possible to pool estimates from relevant studies to increase precision, accuracy and generalisability. EBM practitioners and clinicians do not need to conduct new studies for encoding the required probabilities for decision making. Moreover, the random-effect hierarchical model produces a more sensible estimate than one which ignores the underlying random effects. It is also more superior than the MLE which treats each y_i as an isolated entity. Note that the idea is not exclusively restricted to models designed for combining published evidence. It is also applicable for subject-level relational models as described earlier.

The posterior result may be an end in itself, or serves as the prior for relational modelling involving patient-level evidences. Three specific models are described below.

3.3.2 Continuous Combined Effect

The following model is useful for combining or summarising effects which are continuous in nature. To facilitate discussion, let

y_i = observed effect in study i

φ_i = study-specific effect in study i

θ = overall combined effect

s_i^2 = within-study variance of y_i

ϕ_i = within-study precision of y_i $(=1/s_i^2)$

σ^2 = between-studies variance

τ = between-studies precision (=$1/\sigma^2$)

k= total number of studies under consideration

Only two quantities, namely y_i and s_i^2, are observed (ϕ_i is derived), while φ_i, θ and σ^2 are unobserved and unknown. Since the studies are conducted at different locations and times, it is reasonable to assume $y_i \sim iid[\varphi_i, s_i^2]$ and $\varphi_i \sim iid[\theta, \sigma^2]$. These observed samples in turn form the likelihood function for generating the posterior distribution, which is the ultimate source of information required for any Bayesian analysis. In a random-effect setting, priors must be established for the unobserved quantities, i.e., θ and $\tau=1/\sigma^2$. Following (3.26) and collecting terms with reference to the sequential nature of Bayes' Theorem [26], one yields the following joint posterior distribution:

$$\pi(\theta, \boldsymbol{\varphi}, \tau, \boldsymbol{\phi} | \mathbf{y}) \propto \prod_{i=1}^{k} f(y_i; \phi_i | \varphi_i) \times \prod_{i=1}^{k} f(\varphi_i | \theta) \times g(\theta) \times g(\tau) \tag{3.28}$$

where $\mathbf{y}=(y_1,y_2,\ldots y_k)'$ is the vector of observed study effects, $\boldsymbol{\phi}=(\phi_1,\phi_2,\ldots,\phi_k)'$ the within-study precision, $\boldsymbol{\varphi}=(\varphi_1,\varphi_2,\ldots,\varphi_k)'$, the vector of individual study-specific effects belonging to a distribution characterised by θ, $g(\theta)$ the prior for θ, $g(\tau)$ the prior for τ, $f(y_i; \phi_i | \varphi_i)$ the individual sample distribution for y_i and $f(\varphi_i | \theta)$ the distribution for study-specific effects. The first two terms on the R.H.S. of (3.28) are likelihood functions. The priors may be based on initial subjective assessment before observing the published evidences.

Usually, normal distribution is appropriate for y_i and the priors are chosen within the related conjugate family such that θ~Normal[a, 1/b] and τ~Gamma[c, d]. Assuming normal distribution for y_i is valid for a wide class of applications. As a result, the joint posterior distribution may be written as:

$$\pi(\theta, \boldsymbol{\varphi}, \tau, \boldsymbol{\phi} \mid \mathbf{y}) \propto \prod_{i=1}^{k} \sqrt{\frac{\phi_i}{2\pi}} \exp[\frac{-\phi_i}{2}(y_i - \varphi_i)^2] \times \prod_{i=1}^{k} \sqrt{\frac{\tau}{2\pi}} \exp[\frac{-\tau}{2}(\varphi_i - \theta)^2] \times \sqrt{\frac{b}{2\pi}} \exp[\frac{-b}{2}(\theta - a)^2] \times \frac{c^d}{\Gamma(d)} \tau^{d-1} \exp[-c\tau] \tag{3.29}$$

where $\Gamma(\bullet)$ is a gamma function. Theoretically, inferences about θ should be made from this joint posterior by integrating out the other unknown parameters. Unfortunately, its complicated form makes computation extremely difficult and it is more efficient to work on the conditional posteriors. Such treatment will become clear when the Gibbs sampler is presented.

In a complex hierarchical model such as (3.29) the conditional posteriors are simpler in structure. For example, the posterior of major concern

$$\pi_c(\theta \mid \boldsymbol{\varphi}, \tau, \boldsymbol{\phi}, \mathbf{y}) \propto \prod_{i=1}^{k} \sqrt{\frac{\tau}{2\pi}} \exp[\frac{-\tau}{2}(\varphi_i - \theta)^2] \times \sqrt{\frac{b}{2\pi}} \exp[\frac{-b}{2}(\theta - a)^2] \tag{3.30}$$

is a normal distribution. Similarly, φ_i occurs in only two terms and its posterior conditional on other parameters is:

$$\pi_c(\varphi_i \mid \boldsymbol{\varphi}_{-i}, \theta, \tau, \boldsymbol{\phi}, \mathbf{y}) \propto \prod_{i=1}^{k} \sqrt{\frac{\phi_i}{2\pi}} \exp[\frac{-\phi_i}{2}(y_i - \varphi_i)^2] \times \sqrt{\frac{\tau}{2\pi}} \exp[\frac{-\tau}{2}(\varphi_i - \theta)^2] \tag{3.31}$$

where $\boldsymbol{\varphi}_{-i}$ represents the vector of all other study-specific effects in studies other than i. This is a product of two normal distributions. Last but not least, the conditional posterior for the between-studies precision, τ, is:

$$\pi_c(\tau \mid \theta, \boldsymbol{\varphi}, \boldsymbol{\phi}, \mathbf{y}) \propto \prod_{i=1}^{k} \sqrt{\frac{\tau}{2\pi}} \exp[\frac{-\tau}{2}(\varphi_i - \theta)^2] \times \frac{c^d}{\Gamma(d)} \tau^{d-1} \exp[-c\tau] \tag{3.32}$$

which can be identified as a gamma distribution. Consequently, the posterior of each individual parameter conditional on the values of other parameters is a good starting point and this helps to avoid working on the original joint posterior.

As mentioned before, the Bayesian approach attempts to encode prior knowledge of the parameter through subjective probability based on a prior distribution. In the above-mentioned case, the decision-maker and analyst must specify quantities a, b, c and d for the prior distributions.

While there are different ways to construct a Markov chain that converges to the posterior distribution, the most popular scheme is Gibbs sampler [123]. The algorithm works by sampling from the conditional posteriors of the parameters. Several practical issues must be observed in order to achieve quick convergence from performing Gibbs sampling. First, one may start the chain by setting the initial parameter values equal to the conventional maximum likelihood estimates (MLE), such as the sample means. Next, it is advantageous to select conjugate priors as in some Bayesian analyses. This is because it is more efficient to run the Gibbs sampler with standard distributions. Third, calculation of sample features from a Markov chain should not commence immediately as each chain needs a burn-in period to reach

equilibrium. This is because early results from the Gibbs sampler depend on the initial values and are not representative of the posterior distribution. While one may examine the autoregressive structure of the iterative results, the simplest way to determine the number of burn-ins is to set a very large number. In addition, a large number of updates after burn-in should also be specified. As the number of samples becomes large, the later elements of the sequence will come close to having the stationary distribution. To ascertain if stationary is achieved, one may plot the history of the chain. Analysts are reminded that it is a good practice to fix the number of burn-ins and updates before kicking off the computation.

The following discussion deals with the Gibbs sampler for working with the conditional posteriors outlined earlier. Once the joint posterior and the conditional posteriors are derived and with the initial values specified (usually based on MLE), the MCMC Gibbs sampler algorithm works in the following manner:

i. with starting values $\varphi_i^0 = y_i$, $\theta^0 = \sum_{i=1}^{k} \varphi_i^0 / k$ and $\tau^0 = k / \sum_{i=1}^{k} (\varphi_i^0 - \theta^0)^2$

ii. draw each φ_i randomly using its conditional posterior and the current values of θ and τ

iii. draw θ randomly using its conditional posterior and the current values of $\boldsymbol{\varphi}$ and τ

iv. draw τ randomly using its conditional posterior and the current values of $\boldsymbol{\varphi}$ and θ

v. record the current values of $\boldsymbol{\varphi}$, θ and τ

vi. repeat steps ii. to v. for a sufficiently large number of times, say m=100-1000

vii. summarise θ from the generated sample of posteriors by computing its mean, median, mode and variance (with P.I.s)

Properties of Gibbs sampling can be found in references [140-142]. As mentioned before, the advantage of this elegant yet computationally-demanding approach is that it enables medical decision analysts to encode probability without resorting to the complicated joint posterior (3.29). Even if it is possible to generate directly from the joint posterior, the Gibbs sampler produces a more efficient manner in producing the required results.

The above model is also suitable for summarising any reported evidences analysed with maximum likelihood. This is because MLEs are asymptotically normally distributed if the underlying density satisfies certain regularity conditions [143]. This important result suggests that the above-mentioned model is versatile.

The proposed model also differs from the highly popular empirical Bayes method [144] in several ways. Although the latter is hierarchical in nature, the element of random-effect is not adequately captured. As mentioned before, the random-effect is a critical issue while combining evidences from different published sources. Instead of attempting to model the parameter with a pdf as in the proposed random-effect Bayesian model, the empirical Bayes technique works as follows:

$$
\begin{aligned}
f(y, \varphi \mid \theta) &= \frac{f(y,\varphi,\theta)}{h(\theta)} \\
&= \frac{f(y \mid \varphi) g(\varphi \mid \theta) h(\theta)}{h(\theta)} \\
&= f(y \mid \varphi) g(\varphi \mid \theta)
\end{aligned}
\qquad (3.33)
$$

The likelihood function $\int_{-\infty}^{\infty} f(y \mid \varphi) g(\varphi \mid \theta) d\varphi$ is then established, with ML technique applied for generating the required estimates. Though attractive in many ways, the empirical Bayes technique is found to be inappropriate for combining evidences from published sources.

3.3.3 Combined Effect as Proportions

In situations where the observed study effect is a proportion (y_i=counts/sample size; $0 \le y_i \le 1$), the most appropriate underlying pdf is Beta[q_{1i}, q_{2i}], where both q_{1i}, q_{2i} are the shape parameters. With y_i bounded, the beta distribution is a flexible distribution that takes various shapes (see Figure 3.2). It may be tempting to use model (3.29) for combining reported proportions as they are continuous and distributed as normal asymptotically. However, one must bear in mind that proportions are bounded and this property makes beta distribution a more appropriate underlying distribution.

The development of the model begins with the re-parameterisation of the beta pdf:

$$f(y_i; q_{1i}, q_{2i}) = \frac{(y_i)^{q_{1i}-1}(1-y_i)^{q_{2i}-1}}{B(q_{1i}, q_{2i})}, \qquad 0 \le y_i \le 1;\ q_{1i} > 0;\ q_{2i} > 0 \tag{3.34}$$

where $B(q_{1i}, q_{2i}) = \Gamma(q_{1i}+q_{2i})/\Gamma(q_{1i})\Gamma(q_{2i})$ is the beta function. State without proof, the mean and variance of the distribution are $E[Y_i] = q_{1i}/(q_{1i} + q_{2i})$ and $V[Y_i] = q_{i1}q_{2i}/[(q_{1i} + q_{2i})^2(q_{i1}+ q_{2i}+1)] = (E[Y_i])(1-E[Y_i])/(q_{1i}+q_{2i}+1)$, respectively.

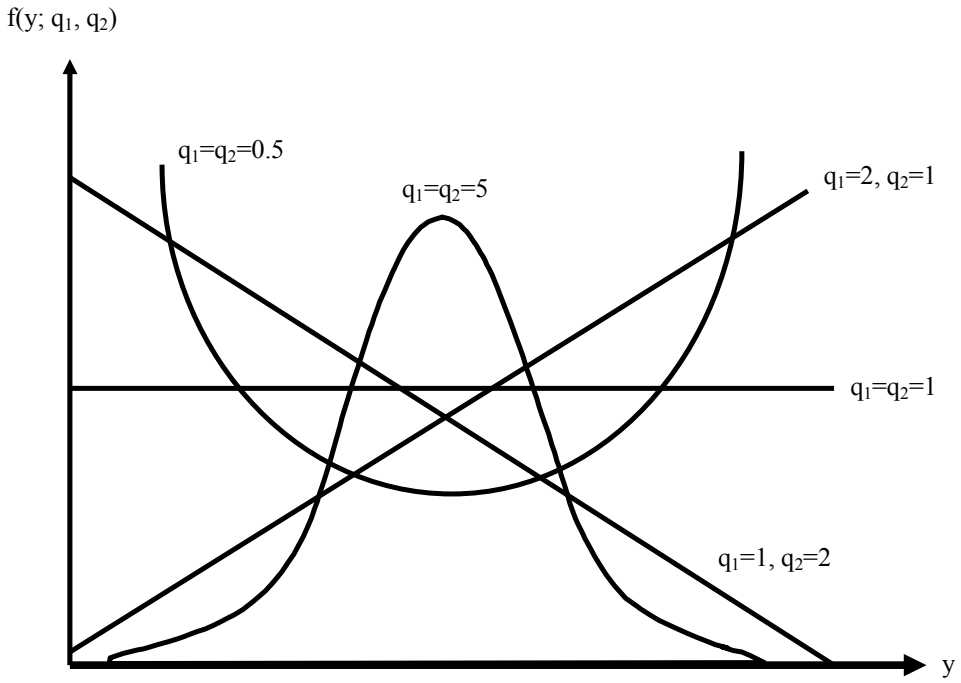


Figure 3.2: Beta distribution

Note that the variance is a function of the mean and a dispersion parameter ($q_{1i}+q_{2i}+1$). Denoting ϕ_i as ($q_{1i}+q_{2i}$) and φ_i as the mean, the beta pdf (3.34) may be re-written as:

$$f(y_i;\, q_{1i},\, q_{2i}) \propto y_i^{\varphi_i\phi_i-1}(1-y_i)^{(1-\varphi_i)\phi_i-1}, \quad 0\le y_i\le 1;\ 0\le\varphi_i\le 1;\ \phi_i>0 \tag{3.35}$$

The required model for combining proportions is then constructed as:

$$\begin{aligned}\pi(\theta,\, \boldsymbol{\varphi},\, \tau,\, \boldsymbol{\phi} \mid \mathbf{y}) \;\propto\; & \prod_{i=1}^{k} y_i^{\varphi_i\phi_i-1}(1-y_i)^{(1-\varphi_i)\phi_i-1} \times \prod_{i=1}^{k} \varphi_i^{\theta\tau}(1-\varphi_i)^{(1-\theta)\tau} \\ & \times \theta^{a-1}(1-\theta)^{b-1} \times \exp[-\tau]\end{aligned} \tag{3.36}$$

where θ, $\boldsymbol{\varphi}$, τ, $\boldsymbol{\phi}$ and $\mathbf{y}$ have their usual meanings. The prior for θ is Beta[a, b]. Without loss of generality, the prior of τ is taken as Exponential[1]. The conditional posteriors are as follows:

$$\begin{aligned}\pi_c(\theta \mid \boldsymbol{\varphi},\, \tau,\, \boldsymbol{\phi},\, \mathbf{y}) &\propto \prod_{i=1}^{k} \varphi_i^{\theta\tau}(1-\varphi_i)^{(1-\theta)\tau} \times \theta^{a-1}(1-\theta)^{b-1} \\ \pi_c(\varphi_i \mid \boldsymbol{\varphi}_{-i},\, \theta,\, \tau,\, \boldsymbol{\phi},\, \mathbf{y}) &\propto \prod_{i=1}^{k} y_i^{\varphi_i\phi_i-1}(1-y_i)^{(1-\varphi_i)\phi_i-1} \times \prod_{i=1}^{k} \varphi_i^{\theta\tau}(1-\varphi_i)^{(1-\theta)\tau} \\ \pi_c(\tau \mid \theta,\, \boldsymbol{\varphi},\, \boldsymbol{\phi},\, \mathbf{y}) &\propto \prod_{i=1}^{k} \varphi_i^{\theta\tau}(1-\varphi_i)^{(1-\theta)\tau} \times \exp[-\tau] \\ &= \prod_{i=1}^{k} \exp[\theta\tau\ln\varphi_i + (1-\theta)\tau\ln(1-\varphi_i)]\exp[-\tau] \\ &= \prod_{i=1}^{k} e^{\tau}[\theta\ln\varphi_i + (1-\theta)\ln(1-\varphi_i) - 1]\end{aligned} \tag{3.37}$$

The Gibbs sampler algorithm may be executed in the following manner:

i. with starting values $\varphi_i^0 = y_i$, $\theta^0 = \sum_{i=1}^{k} \varphi_i^0 / k$ and $\tau^0 = k / \sum_{i=1}^{k} (\varphi_i^0 - \theta^0)^2$

ii. draw each φ_i randomly using its conditional posterior and the current values of θ and τ

iii. draw θ randomly using its conditional posterior and the current values of $\boldsymbol{\varphi}$ and τ

iv. draw τ randomly using its conditional posterior and the current values of $\boldsymbol{\varphi}$ and θ

v. record the current values of $\boldsymbol{\varphi}$, θ and τ

vi. repeat steps ii. to v. for a sufficiently large number of times, say m=100-1000

vii. summarise θ from the generated sample of posteriors by computing its mean, median, mode and variance (with P.I.s)

3.3.4 Combined Effect as Rates

The combination of effects of rates (y_i) works in a similar fashion. It is a well-known fact that if a rate y_i follows Poisson(μ), then the conjugate prior of μ is gamma(a, b), where a, b>0. In the case of the full Bayesian random-effect hierarchical model, the observed effect, y_i, follows Poisson(μ_i), and μ_i follows Gamma(1, θ). The prior distribution for θ—the parameter of interest—is chosen to be $g(\theta) = \frac{\exp(-1/a\theta)}{a\theta^2}$, a>0, θ>0.

The joint posterior for μ and θ given **y** is:

$$\pi(\theta, \boldsymbol{\mu} \mid \mathbf{y}) \propto \prod_{i=1}^{k} \exp(\mu_i) \mu_i^{y_i} \times \prod_{i=1}^{k} \frac{\exp(-\mu_i/\theta)}{\theta} \times \frac{\exp(-1/a\theta)}{a\theta^2} \tag{3.38}$$

The conditional distribution for $\boldsymbol{\mu}$ is:

$$\pi_c(\boldsymbol{\mu} \mid \theta, \mathbf{y}) \propto \prod_{i=1}^{k} \exp(\mu_i)\mu_i^{y_i} \times \prod_{i=1}^{k} \frac{\exp(-\mu_i/\theta)}{\theta}$$
$$\propto \prod_{i=1}^{k} \mu_i^{y_i+1} \times \exp[-\sum_{i=1}^{k} \mu_i(1+\frac{1}{\theta})] \tag{3.39}$$

which is a gamma function. The conditional for θ is more complicated, but it can be simplified as follows:

$$\pi_c(\theta \mid \boldsymbol{\mu}, \mathbf{y}) \propto \prod_{i=1}^{k} \frac{\exp(-\mu_i/\theta)}{\theta} \times \frac{\exp(-1/a\theta)}{\theta^2}$$
$$= \frac{\exp[-\frac{1}{\theta}(\sum_{i=1}^{k} \mu_i + \frac{1}{a})]}{\theta^{k+2}}$$
$$= \lambda^{(k+1)-1} \exp[-\lambda(\sum_{i=1}^{k} \mu_1 + \frac{1}{a})] \tag{3.40}$$

which is also a gamma distribution, by making use of the transformation $\lambda=1/\theta$. The Jacobian is $\frac{d\theta}{d\lambda} = -\lambda^{-2}$.

The Gibbs sampler is executed as follows:

i. with starting values $\mu_i^0 = y_i$ and $\theta^0 = \sum_{i=1}^{k} \mu_i^0 / k$

ii. draw each μ_i randomly from its gamma conditional posterior and the current values of λ

iii. draw λ randomly from its gamma conditional posterior and the current values of $\boldsymbol{\mu}$

iv. record the current values of $\boldsymbol{\mu}$ and λ

v. iterate the process ii. to iv. for a specified m times

vi. summarise θ based on the back-transformation of λ by computing its mean, median, mode and variance (with P.I.s)

3.4 Other Relational Models

Three additional classes of Bayesian relational models are presented in this section. They are useful for analysing continuous outcomes, time to event data and clustered and hierarchical observations. These models are chosen based on a detailed literature review given in Chapter 2 and their usability in the case studies to be presented in Chapter 4.

3.4.1 Continuous Outcome

The linear regression model is most familiar to medical data analysts. Based on a linear link function, it is suitable for modelling continuous y_i like body weight, temperature, size of tumour and glucose level, etc. It is legitimate to apply (3.8) for modelling y_i in this case, but a more straight-forward approach makes use of the distributional property of the residuals (u_i), e.g., distance between the actual observation and its predicted value.

A very attractive feature of such formulation is that it generates closed-form solutions for $\boldsymbol{\beta}$ [51, 88, 96]. Formulate the linear model as:

$$y_i = \mathbf{x}'_i\boldsymbol{\beta} + e_i \tag{3.41}$$

where residual $u_i \sim$Normal[0, σ^2]. With conjugate priors $\boldsymbol{\beta} \sim$Multivariate Normal[$\boldsymbol{\beta_0}$, $\boldsymbol{\Sigma_0}$] and $\sigma^2 \sim$Inverse Gamma[a, b], the joint posterior distribution of $\boldsymbol{\beta}$ and σ^2 becomes

$$\pi(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{X}, \mathbf{y}) \propto L(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{X}, \mathbf{y}) \times g(\boldsymbol{\beta}, \sigma^2)$$
$$= (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]$$
$$\times (2\pi\sigma^2)^{-(p+1)/2}|\Sigma_0|^{-1}\exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\beta_0})'\Sigma_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta_0})\right] \times \sigma^{-(a-p-1)}\exp[-b/\sigma^2]$$

$$\propto \sigma^{-n} e\{-\frac{1}{2\sigma^2}[\hat{\sigma}^2(n-k-1) + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})]\}$$
$$\times (2\pi\sigma^2)^{-(p+1)/2} |\Sigma_0|^{-1} e[-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\Sigma_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)] \times \sigma^{-(a-p-1)}\exp[-b/\sigma^2] \quad (3.42)$$

by making use of the fact that $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\hat{\sigma}^2 = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})/n-p-1$, where $\mathbf{X}=[\mathbf{x}_1\ \mathbf{x}_2\ \ldots\mathbf{x}_p]'$ and $\mathbf{y}=(y_1,y_2,\ldots y_n)'$, p<n, are the data matrices. $\hat{\boldsymbol{\beta}}$ is in fact a MLE. As readily seen, the posterior (3.42) is extremely complicated despite its simple set-up (3.41). Fortunately, it is mathematically tractable with suitable substitution and transformation. Re-express the joint posterior as:

$$\pi(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{X}, \mathbf{y}) \propto \sigma^{-a-n} \exp\{-\frac{1}{2\sigma^2}[\sigma^* + (\boldsymbol{\beta}-\boldsymbol{\beta}_0)'(\Sigma_0^{-1} + \mathbf{X}'\mathbf{X})(\boldsymbol{\beta}-\boldsymbol{\beta}_0)]\} \quad (3.43)$$

by making use of he following quantities:

$$\boldsymbol{\beta}^* = (\Sigma_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\Sigma_0^{-1}\boldsymbol{\beta}_0 + \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}})$$
$$\sigma^* = 2b + \hat{\sigma}^2(n-p-1) + (\boldsymbol{\beta}_0-\boldsymbol{\beta}^*)'\Sigma_0^{-1}\boldsymbol{\beta}_0 + (\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \quad (3.44)$$

The joint posterior is now recognised as a normal-inverse gamma distribution. The quantity $\boldsymbol{\beta}^*$ may now be used as the estimator for $\boldsymbol{\beta}$. This Bayesian estimator is also known as an "shrinkage estimator" because the MLE $\hat{\boldsymbol{\beta}}$ is "shrunk" towards the prior $\boldsymbol{\beta}_0$. With suitable transformation the marginal posterior of $\boldsymbol{\beta}$ is derived:

$$\pi(\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{y}) \propto [\sigma^* + (\boldsymbol{\beta}-\boldsymbol{\beta}^*)'(\Sigma_0^{-1} + \mathbf{X}'\mathbf{X})(\boldsymbol{\beta}-\boldsymbol{\beta}^*)]^{-(n+a)/2} \quad (3.45)$$

which is easily recognised as a multivariate-t distribution. Therefore, the mean and covariance of the posterior coefficient estimates are given by:

$$\begin{aligned} &E[\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{y}] = \boldsymbol{\beta}^* \\ &COV[\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{y}] = \sigma^*(\Sigma_0^{-1} + \mathbf{X}'\mathbf{X})^{-1} / [n+a-(p+1) -3] \end{aligned} \tag{3.46}$$

Using the same line of argument, one can deduce that the joint posterior of $\boldsymbol{\beta}$ and σ^2 with non-informative priors, say c and $1/\sigma$, respectively:

$$\begin{aligned} &\pi(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{X}, \mathbf{y}) \propto \sigma^{-n} \exp\{-\frac{1}{2\sigma^2}[\hat{\sigma}^2 (n-p-1) + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})]\} \times c \times 1/\sigma \\ &\propto \sigma^{-n-1}\exp\{-\frac{1}{2\sigma^2}[\hat{\sigma}^2 (n-p-1) + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})]\} \end{aligned} \tag{3.47}$$

with the following marginal posterior of $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{y}) \propto [(n-p-1) + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'(\hat{\sigma}^{-2}\mathbf{X}'\mathbf{X})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})]^{-n/2} \tag{3.48}$$

which can again be easily recognised as a multivariate t-distribution. Thus, $E[\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{Y}] = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $COV[\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{y}] = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$ where $\hat{\sigma}^2$ is usually taken to be $(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})/n-p-1$. The $100(1-\alpha)\%$ P.I. for $\boldsymbol{\beta}$ can thus be constructed accordingly. The above result is not surprising and it illustrates an important feature of all Bayesian models, that is, $\boldsymbol{\beta}^*$ reduces to the conventional estimator ($\hat{\boldsymbol{\beta}}$) if non-informative priors are fitted.

3.4.2 Time to Event

The next relational model is designed for analysing subject-level time to event data (y_i). Recall from subsection 2.3.2, if the underlying skewed distribution for y_i is $f(y_i)$, then the cumulative distribution function (cdf) is $F(Y<y_i)$ and the survivor function is $S(y_i)=P(Y\geq y_i)=1-F(Y<y_i)$. These subjects are "censored" because no event is observed in them. Thus, the likelihood for right-censored observations becomes $\prod_{i=1}^{n} f(y_i)^{\delta_i} S(y_i)^{1-\delta_i}$ where δ_i takes the value 0 if the subject is censored and 1 if otherwise.

Bayesian analysis requires a suitable distribution, say Weibull[α, λ], to fit the data. Like the beta distribution, Weibull is extremely flexible. In this case, α and λ are the scale and shape parameters, respectively. The required functions are as follows [101]:

$$
\begin{aligned}
&f(y_i \mid \alpha, \lambda) = \alpha\, y_i^{\alpha-1} \exp[-\lambda - \exp(\lambda)\, y_i^{\alpha}] \qquad\qquad y_i,\ \alpha, \lambda > 0,\\
&S(y_i \mid \alpha, \lambda) = \exp[-\exp(\lambda)\, y_i^{\alpha}]\\
&L(\alpha, \lambda \mid y_i, \delta_i, \mathbf{x}_i) = \prod_{i=1}^{n} f(y_i \mid \alpha,\lambda)^{\delta_i} S(y_i \mid \alpha,\lambda)^{1-\delta_i}\\
&\qquad \propto \alpha^{\sum_{i=1}^{n}\delta_i} \exp\{\lambda \sum_{i=1}^{n}\delta_i + \sum_{i=1}^{n}[\delta_i(\alpha-1)\log y_i - \exp(\lambda) y_i^{\alpha}]\}\\
&\pi(\alpha, \lambda \mid y_i, \delta_i, \mathbf{x}_i) \propto L(\alpha, \lambda \mid y_i, \delta_i, \mathbf{x}_i) \times g(\alpha) \times g(\lambda)\\
&\qquad \propto \prod_{i=1}^{n} f(y_i \mid \alpha,\lambda)^{\delta_i} S(y_i \mid \alpha,\lambda)^{1-\delta_i} \times g(\alpha) \times g(\lambda)\\
&\qquad \propto \alpha^{\alpha_0 + \sum_{i=1}^{n}\delta_i - 1} \times\\
&\exp\{\lambda \sum_{i=1}^{n}\delta_i + \sum_{i=1}^{n}[\delta_i(\alpha-1)\log y_i - \exp(\lambda) y_i^{\alpha})] - \kappa_0\alpha - \frac{1}{2\sigma_0^2}(\lambda-\mu_0)^2\}
\end{aligned}
\tag{3.49}
$$

where the priors are $\alpha \sim \text{Gamma}[\alpha_0, \kappa_0]$ and $\lambda \sim \text{Normal}[\mu_0, \sigma_0^2]$. Then introduce the predictors through λ, i.e., $\lambda_i = \mathbf{x}'_i\boldsymbol{\beta}$. Let the prior $\boldsymbol{\beta} \sim \text{Normal}[\boldsymbol{\beta}_0, \Sigma_0]$, the joint posterior is derived:

$$\pi(\boldsymbol{\beta}, \alpha \mid y_i, \delta_i, \mathbf{x}_i) \propto \alpha^{\alpha_0 + \sum_{i=1}^{n} \delta_i - 1} \times \exp\{\sum_{i=1}^{n} [\delta_i \mathbf{x}_i' \beta + \delta_i(\alpha - 1)\log(\mathbf{y}_i) - \mathbf{y}_i^{\alpha} \exp(\mathbf{x}_i'\beta)] - \kappa_0 \alpha - \frac{1}{2}(\beta - \beta_0)' \Sigma_0^{-1} (\beta - \beta_0)\} \tag{3.50}$$

Estimation of $\boldsymbol{\beta}$ is carried out with MCMC. Weibull is a versatile distribution, with exponential as its special case (λ=1). With suitable parameterisation, it can be converted to an extreme-value distribution. Gamma, exponential and log-normal distributions are other candidate distributions for modelling time to event data.

3.4.3 Longitudinal and Clustered Data

In situations where the data are clustered or collected repeatedly over time, then some hierarchical structures are present. See subsection 2.3.3 for details. To facilitate discussion, consider outcome y_{ij} related to predictors $\mathbf{x}_{ij}$ for observations i=1, 2, …, n_j within clusters j=1, 2, …, k. In this case, while the clusters are likely to be independent, the evidences within each cluster are not. Conditional on the cluster effect υ_j, y_{ij} belongs to the exponential family distribution. The posterior is developed as below [110]:

$$f(y_{ij}; \theta_{ij} \mid \upsilon_j) = \exp\{[\theta_{ij} y_{ij} - \psi(\theta_{ij})] + \gamma(y_{ij})\} \quad i=1, 2,\ldots, n_j; \quad j=1, 2, \ldots, k$$
$$\pi(\boldsymbol{\beta}, \boldsymbol{\upsilon}, \boldsymbol{\Sigma} \mid y_{ij}, \mathbf{x}_{ij}) \propto [\prod_{i=1}^{n_j} \prod_{j=1}^{k} L(y_{ij} \mid \beta, \upsilon_j, x_{ij}) g(\beta)] \prod_{j=1}^{k} g(\upsilon_j \mid \Sigma) h(\Sigma) \tag{3.51}$$

where $\boldsymbol{\upsilon}=(\upsilon_1, \upsilon_2,\ldots \upsilon_n)'$ and $\boldsymbol{\Sigma}$ are the cluster-effect parameters and covariance matrix, respectively.

It may be useful to consider the conditional posteriors in modelling:

$$\pi_c(\boldsymbol{\beta} \mid \boldsymbol{\upsilon}, \boldsymbol{\Sigma}, y_{ij}, \mathbf{x}_{ij}) \propto \prod_{i=1}^{n_j}\prod_{j=1}^{k} f(y_{ij} \mid \beta, \upsilon_j, x_{ij}) g(\beta)$$
$$\pi_c(\upsilon_j \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, y_{ij}, \mathbf{x}_{ij}) \propto \prod_{i=1}^{n_j} f(y_{ij} \mid \beta, \upsilon_j, x_{ij}) h(\upsilon_j \mid \beta)$$
$$\pi_c(\boldsymbol{\Sigma} \mid \boldsymbol{\upsilon}, \boldsymbol{\beta}, y_{ij}, \mathbf{x}_{ij}) \propto \prod_{j=1}^{k} w(\upsilon_j \mid \Sigma) z(\Sigma) \tag{3.52}$$

The conditional posterior (π_c) of a particular parameter is derived by holding other parameters fixed. With specific distributions and functions chosen, the above set-up allows analyst to model clustered data with different nature, say counts, rates and continuous. MCMC is required for summarising the posterior results.

The above-mentioned two-level hierarchical data structure is a special case of the multi-level framework. But it is realistic enough to handle many real-life scenarios [96].

3.5 Specific Modelling Issues

3.5.1 Sensitivity Analysis

The quality of a Bayesian model depends on several criteria. Besides being parsimonious and interpretable, a reasonable Bayesian model should be resistant to departures in the specification of prior distribution. A common criticism of Bayesian analysis is that the prior distributions can never be correctly quantified, elicited and fitted, especially when time is limited. The allegation is not unfounded.

Therefore, it is important to assess the appropriateness of prior distribution in Bayesian analysis. The most straight-forward approach is to employ sensitivity analysis, a standard term refers to the process of investigating changes in the conclusions (posterior) caused by changes in the initial assumptions (prior).

Global sensitivity analysis is a broad approach that evaluates a wide range of alternative prior specifications, forms of the link function, error sensitivity and perturbations of the prior specifications and the likelihood. Unfortunately, this idea is too broad to be useful in practice. In a narrower sense, the analyst may specify an alternative non-informative prior over the support of the parameter of interest and compare it with the stipulated informative prior. A substantive change to the form of the posterior is a signal for caution.

In local sensitivity analysis, one specifies a more diffuse form of the prior than that originally specified. If no appreciable changes to the posterior are observed, one is confident that the initial prior is not specified incorrectly. Compared to the global approach, this approach incurs a lower cost in terms of effort and reporting.

It must be pointed out that if the initial prior is supported firmly on theoretical ground, then large changes to the posterior, as indicated by the sensitivity analysis, are not a sufficient reason to discard the chosen prior.

3.5.2 Robust Analysis

A different approach to minimise the effect of misspecified prior on posterior is based on the concept of robustness. It is not uncommon for clinicians to specify a wrong prior as they might base their prior evidence on faulty reasoning or beliefs. They may also be presented with past evidences that were generated from different contexts. With strong confidence in the prior, one may specify a highly precise yet incorrect prior.

The suggestion offered by the robust approach is to introduce a mixture prior:

$$g_{mixture}(\theta) = p_0 g_0(\theta) + p_1 g_1(\theta) \quad p_1 = 1 - p_0 \tag{3.53}$$

where $g_0(\theta)$ is the initial prior, $g_1(\theta)$ the more widely-spread alternative prior, p_0 the probability that $g_0(\theta)$ is correct and p_1 the probability that $g_1(\theta)$ is correct. One usually fixes p_0 with a high value, say above 0.9, thus reflecting the belief that the initial prior is correct. This also means that one is giving a small chance that the initial prior is misspecified.

According to the Bayes' Theorem [26], the respective posteriors are $\pi_0(\theta \mid y)$ and $\pi_1(\theta \mid y)$, and the mixture posterior is:

$$\pi_{mixture}(\theta \mid y) = p'_0 \pi_0(\theta \mid y) + p'_1 \pi_1(\theta \mid y) \tag{3.54}$$

The mixture posterior is thus a mixture of two posteriors. Note that p'_0 and p_1' are based on observed evidence and are proportional to the prior probabilities (p_0 and p_1) multiplied by the probability density evaluated at the evidence that has occurred.

Also note that the mixture indicator, 0 for initial prior and 1 for alternative prior, is marginalised out.

If the initial prior is correct, the mixture posterior is very close to the initial posterior. However, if the initial prior is very different from the likelihood, the posterior probability p'_0 will turn out to be very small and the mixture posterior will then be dominated by the likelihood.

The use of mixture prior provides some protection against a misspecified initial prior. This is the essence of robust statistics [145], a branch of methodological thought that flourished in the 1980s.

3.5.3 AdaBoost

The next issue concerns a Bayesian model's predictive ability. It is useless to have a sophisticated model that provides an excellent fit to the data but fails to predict future or out-of-the-sample observations adequately.

The logit model (3.16) described above may be called a classifier as it helps to determine how patients are classified (e.g., alive/dead, recover/relapse, improve/deteriorate). Classifier is a technique that develops well-defined rules for assigning observations. The following section concerns a machine-learning algorithm, known as AdaBoost [146], that might help to improve its predictive performance.

The Bayesian logit classifier described earlier is "global" in the sense that it is designed to apply to the entire data set over all the variable space. Since a single classifier may not be ideal, a combination of classifiers from the same data set may help to identify more special features of the misclassified observations. Therefore, one may view such combination of classifiers as a way of improving the performance of single "weak" classifiers (see Figure 3.3).

To facilitate discussion, define **x** as the vector of predictors, y the observed outcome and n the sample size. Unlike the usual logit models, $y \in \{-1,1\}$ in this case. Initially, each of the observations $(\mathbf{x}_i, y_i)$ is assigned an equal weight, such that $w_i=1/n$. At each stage of the iterative process, a classifier is constructed using the weights w_i, which reflects the probability of occurrence of the observations. In the process, misclassified observations are up-weighted, while correctly-classified observations are down-weighted. An error index, corresponding to the sum of weights of the misclassified observations is then computed. It is noted that the weights corresponding to misclassifed observations always increase and the classifier is updated with these weights. The AdaBoost procedure is summarised as follows:

i. initialise the weights $w_i=1/n$, i=1, 2,…, n.

ii. for m=1, 2,…, M, construct a classifier $\varpi(\mathbf{x})$ from the training set with weights w_i; compute e_m as the sum of weights w_i corresponding to misclassified observations; if $e_m>0.5$ or $e_m=0$ then terminate the procedure, otherwise set $w_i=w_i(1-e_m)/e_m$ for misclassified observations and renormalise the weights so that they sum to unity, and continue the process.

iii. for a two-class classifier, in which $\eta(\mathbf{x})>0$ implies $\mathbf{x} \in c_1$, otherwise c_0, form a weighted sum of the classifiers, ω_m, m=1, 2,…, M, $\hat{\omega}=\sum_{m=1}^{M}\log(\frac{1-e_m}{e_m})\varpi_m(x)$ and assign **x** to c_1 if $\hat{\omega}>0$, where c_j is the outcome class or level.

The final classifier is a linear combination of the classifiers from each stage of the process. The condition of $e_m>0.5$ ensures that smaller weights, i.e. lower

$\log(1-e_m/e_m)$, are attached to classifiers with larger number of misclassified observations.

The underlying notion driving combined classifiers is that if the results are averaged (based on the weighting scheme), more accurate and stable models may be produced. Research on combining classifiers in a single analysis has demonstrated that there are potentially very substantial gains in predictive accuracy. The reduction in error rates for well-known data sets typically ranges from 5 to 40%; albeit, it is important to realise that improvements are not always guaranteed. Moreover, when classifiers are combined, the final product is no longer visually appealing and comprehensible.

Nevertheless, the algorithm may also be applied routinely to check if a single Bayesian classifier has produced the "best possible" predictive performance. However, one should note that the AdaBoost algorithm can be applied to any model and is not exclusively a complementary tool for Bayesian modelling.

3.5.4 Receiver Operating Characteristic Curve

The predictive accuracy of a model can be expressed quantitatively in terms of sensitivity and specificity. Sensitivity is the proportion of truly diseased patients identified by the predicting model, while specificity is the proportion of non-diseased patients correctly identified as non-diseased by the model.

Apparently, these two measures are inversely related. A graphical display of this relationship is called a Receiver Operating Characteristic (ROC) curve [147-148]. Potting sensitivity against specificity provides a visual interpretation to the predictive accuracy of a model. The model's predictive accuracy is quantified by the area under the curve (AUC). The higher the AUC the more accurate is the model's prediction.

To compare the predictive accuracy of two or more models one compares the differences in AUCs. A common practice is to generate the respective confidence intervals (C.I.) for the AUCs and declare a significant difference in the models' predictive accuracy if the C.I.s do not overlap. This is, however, a crude approach. One should take into account not only the numerical difference in AUCs and their standard errors, but also the correlation between ROC curves [149]. This is because the models are applied to the same data set.

ROC has other uses in data analysis. It is also useful for identifying the optimum cut-off, characterised by the highest possible sensitivity and specificity, of a screening index or a diagnostic test.

3.5.5 Elicitation of Utilities

To complete the formulation of a decision problem, the analyst must elicit the preferences and utilities from the decision-maker. Utilities are the numerical ratings of the desirability of health states that reflect a patient's preferences.

The following suggests how a decision maker's (patient) utility is assessed and measured systematically:

- list the possible outcomes that could occur
- rank the outcomes in order of decision maker's preference
- assign utility values to the boundary conditions (1: most preferred outcome, 0: least preferred)
- create a scenario such that the decision maker is indifferent between the boundary conditions ($x_{0.5}$)
- choose an appropriate utility function characterised by the risk tolerance

parameter (ρ)

- estimate ρ based on $x_{0.5}$
- generate utilities for all identified intermediate outcomes

The exponential utility function is strongly recommended here as it fulfils the delta property:

$$u(x) = \frac{1 - e^{-\frac{x}{\rho}}}{1 - e^{-\frac{H-L}{\rho}}} \tag{3.55}$$

where x is the condition, H and L the boundary conditions, and ρ the risk tolerance. There are existing tables for generating the required values for ρ.

The alternative with the highest expected utility (EU) is chosen as the preferred decision. See section 1.2 regarding the rationale of applying EU as the basis of decision making, despite the serious challenges posed by competing theories [8-12]. Since the objective of this dissertation is about probability encoding, less emphasis is given to utility elicitation, which may actually pose a bigger challenge to decision analysts in many real-life applications.

Relying on known utility to generate unknown utilities for all intermediate outcome states, the easy-to-use technique outlined above may be seen as a modified version of the full standard reference gamble approach [7], where patients are presented with a series of hypothetical scenarios. It is also different from the time trade-off technique [7] where patients are asked to express their attitudes towards various lengths of time in ill health. Both the time trade-off and the full standard reference gamble are not recommended in this dissertation because of their difficulty

in applications. A number of hypothetical scenarios must be generated in order to elicit the patients' utilities and this often adds unnecessary confusion to the patients and clinicians.

CHAPTER 4

CASE STUDIES

The following analyses utilised evidences scoured from case reports, medical textbooks, disease registries, audited pamphlets distributed by drug manufacturers, abstracts of scientific conference proceedings, peer-reviewed journal articles (randomised clinical trials, case-control studies, cohort studies, systematic reviews, etc.), government and health agency bulletins and other relevant published documents.

The first 5 studies involved clinicians as the principal decision makers. No patients were involved in decision making as the primary objectives were to i) identify the risk factors associated with the medical conditions of interest, ii) evaluate the comparative effectiveness of drugs and treatment procedures, iii) develop screening tools for early detection of diseases, iv) identify complications of certain diseases, and v) compare performance of predictive models, etc.

The above-mentioned analyses are decision problems in the broad sense, but the results may then be applied as either prior evidence for future analyses or encoded as probabilities for the chance nodes in problems involving patients as the primary decision makers. Making decisions about medical care is believed to be most effective when a clinician and a patient work together. The optimum decision is made when the clinician's experience, knowledge and access to evidence are combined with patient's values, wishes and understanding of his problems.

With this in mind, the remaining 5 case studies focussed on helping patients to deal with their health conditions. With their preferences and utilities elicited, the

Bayesian models were applied to generate probabilities for quantifying all uncertainties identified in the relevant decision trees or influence diagrams. The prior evidences were either gathered from reviews of published information or elicited from domain experts.

The choice of the case studies is based on the current challenges faced by local EBM practitioners. It is a well-known fact that more and more Singaporeans are suffering from depression, osteoporosis, colon cancer, dengue fever, strokes, obesity, heart diseases, renal failure and breast cancer. These case studies are original research in their own right. They were conducted not merely to illustrate the properties and features of the proposed Bayesian models. They help to shed light on specific medical problems faced by EBM practitioners and the evidences may serve as the cornerstone for future medical guideline development.

In the Bayesian literature the terms "weak" or "strong" likelihood are commonly used to describe widely-spread or concentrated likelihood, respectively [69]. In the case when the observed evidences are pointing to a similar direction or supporting a particular hypothesis, the resultant likelihood is said to be "concentrated". However, there is no objective measure of a likelihood's "concentration". To preserve the vividness of such descriptions, the terms are retained in the following case studies. But more commonly-used descriptions are also provided for readers who are not at ease with the vague terminology.

4.1 Evaluation of Antidepressants' Tolerability

4.1.1 Aim

Depression is a common disorder that is becoming better understood as an illness that can be chronic, recurrent, and refractory to treatment [150]. Generally defined as a mood disorder that impairs normal functions, depression may be caused by many factors, including biochemical disorders. The onset of depression may be affected by certain medications, hormones and the occurrence of medical illnesses. Losing a loved one, financial concerns, work stress, or relationship problems may all contribute to depressive disorders. One is also at risk if there is a family history of depression.

People with the condition typically have problems regulating certain brain chemicals called neurotransmitters. By working within the brain to increase the levels of either noradrenaline, serotonin or both, antidepressants help to reduce the symptoms of anxiety and negative thoughts usually experienced by sufferers. However, they do not act immediately and the lifting of moods typically takes up to 2 weeks or longer.

There is a wide class of antidepressants available for treating all forms of depression regardless of cause. The most common antidepressants are selective serotonin reuptake inhibitors (SSRIs) and tricyclic antidepressants (TCAs). Widely regarded as one of the most effective antidepressants, the latter had been the first-choice medication for over 30 years. TCAs work by preventing the uptake of norepinephrine and serotonin, thus building up the concentration of these transmitters and improving the communication between neurons. However, there are a number of problems with TCAs as their effects in the brain are not restricted to alleviating depression. They also interact with a number of other brain receptors, thus causing side-effects like dry mouth, drowsiness, dizziness, blurred vision, constipation,

urinary difficulties, tremor and tachycardia. In addition, they tend to lower blood pressure and consequently cause a feeling of faintness. Morbidity and mortality caused by TCA overdose are also widely reported [151-152]. As such, TCAs are usually prescribed in severe cases of depression.

On the other hand, SSRIs work only on the serotonin system. Introduced in the 1980s, SSRIs possess an improved side-effect profile over TCAs with their selective mode of action [153]. While retaining good clinical efficacy, they have few of the anticholinergic, antihistaminergic and cardiotoxic effects [154] and are probably safer in overdose than TCAs [155]. As a result, SSRIs are recognised to be better tolerated and more acceptable to patients [156-160]. However, there are other well-documented side-effects associated with the use of the drugs. These include nausea, vomitting, diarrhoea, decreased appetite, fatigue, increased sweating, sleep disturbances and impotency [159]. Moreover, SSRIs may also interact with other drugs so extra care must be taken in prescription.

Both TCAs and SSRIs achieve similar efficacy, with 60% to 80% of patients responding adequately [161-162]. However, their side-effect profiles vary substantially, so the choice of medication for treating depression depends primarily on patients' tolerability. As such, the main objective of this case study was to conduct meta-analyses on the tolerability of SSRIs and TCAs. The results of the proposed Bayesian hierarchical models (3.29-3.32) were compared with those of the conventional model.

4.1.2 Selection of Published Studies

There were two sets of analyses. The first was concerned with primary-care patients' premature discontinuation from treatments due to drug-related side-effects.

It is important to investigate patients' tolerability in the primary care setting since the majority of depression cases are first seen and treated in general practice [163-166]. Randomised controlled trials investigating the efficacy and tolerability of SSRIs (fluoxetine, fluvoxamine, paroxetine, sertraline and citalopram, etc.) against a TCA or an antidepressant with identical mechanism of action (amitriptyline, imipramine, dothiepin, clomipramine and lofepramine, etc.) in patients with depressive disorders were identified through MEDLINE and Cochrane Library search up to May 2004, previous meta-analyses [162, 167-172] and literature review. The patients' depressive disorders were assessed by means of the Research Diagnostic Criteria (RDC), Diagnostic Systems (DSM-III), Hamilton Rating Scale for Depression (HRSD), Clinical Global Impression Score (CGI), Clinical Anxiety Scale (CAS) and Montgomery-Asberg Depression Rating Score (MADRS), and so on. No language restriction was imposed in the search. Studies were excluded from analysis if there were insufficient information on study design, description of treatments and tolerability, source of patients, and so on.

In the second analysis, the Bayesian model was compared with a reported result published in the Cochrane Database of Systematic Reviews [172]. This analysis serves to provide a more complete assessment of the comparative tolerability of the antidepressants in the general setting. The published result was based on a search on the Cochrane Collaboration Depression, Anxiety and Neurosis Controlled Trials Registers (1977-1999), MEDLINE (1966-1999), EMBASE (1974-1999), specialist journals, previous systematic reviews, conference abstracts, government documents and reference lists of relevant papers [172].

For the abovementioned analyses, the parameter of interest was the overall or combined odds ratio (OR) of premature withdrawal from treatments due to drug-

related side effects. The combined and study-specific ORs were computed such that a value above unity suggests that SSRIs were better tolerated than TCAs. To allow direct comparison with published results, the odds ratios were computed by Peto's Observed-Expected (O-E) method [173]. In order to provide more insights into the issues, several priors reflecting different views of the comparative tolerability of the antidepressants were applied in the proposed Bayesian analyses. The data were entered into Stata 9.0 for analysis and all statistical tests were conducted at 5% level of significance.

4.1.3 Discontinuation from Primary-Care Due to Side-Effects

The data were extracted from 7 randomised double-blinded clinical trials [174-180]. The trials, mostly conducted at multi-centres, involved a total of 2,524 patients (SSRIs: 1,386, TCAs: 1,138). Information concerning drug treatments, inclusion criteria and basic demographics of the patients is depicted in Table 4.1. Several studies, including those reported in a similar meta-analysis [170, 181], were omitted from analysis because of insufficient information. In addition, 2 trials which recruited only elderly patients aged 65 years and above [165, 182] were also excluded.

Three of the selected studies, notably those with large sample sizes, were in favour of SSRIs (Table 4.2 and Figure 4.1). Combining all 7 studies, the conventional model [115-116, 183] produced an overall OR of 1.35 (95% C.I.: 1.06—1.73) (Table 4.3), thus suggesting that SSRIs were better tolerated than the TCAs. The conventional fixed-effect model was applied here because the assumption of homogeneity could not be discarded (p-value: 0.09).

Table 4.1: Selected clinical trials with primary-care patients

Study	Patient selection	Antidepressants SSRI/TCA	Duration	Profile of patients
Corne 1989	Met RDC; HAM-D≥17	Fluoxetine/ dothiepin	6 weeks	Gender: 70% female Mean age: 41.7
Stott 1993	MADRS≥16; CAS≥11	Paroxetine/ amitriptyline	8 weeks	Gender: 66.5% female Mean age: 42.8
Rosenberg 1994	HAMD≥14	Citalopram/ imipramine	6-22 weeks	Gender: 70% female Age range: 19-65
Doogan 1994	Met DSM-III-R; MADRS≥22; CGI≥4	Sertraline/ dothiepin	6 weeks	Gender: 70.5% female Mean age: 47.1
Moon 1996	Met DSM-III-R; MADRS≥18	Paroxetine/ lofepramine	6 weeks	Gender: 71.3% female Mean age: 43.7
Christiansen 1996	HAMD≥15	Paroxetine/ amitriptyline	8 weeks	Age range: 18-65
Ravindran 1997	MADRS≥20 CAS≥11	Paroxetine/ clomipramine	12 weeks	Gender: 73.5% female Mean age: 42.6

Table 4.2: Primary-care patients discontinued from treatments due to side-effects

Study	SSRIs No/Total	TCAs No/Total	OR (95% C.I.)
Corne 1989	7/49	2/51	0.29 (0.07—1.12)
Stott 1993	35/243	49/262	1.36 (0.85—2.18)
Rosenberg 1994 *	43/380	16/92	1.74 (0.87—3.46)
Doogan 1994	5/83	2/96	0.35 (0.08—1.60)
Moon 1996	5/60	4/62	0.76 (0.20—2.94)
Christiansen 1996	9/71	9/73	0.97 (0.36—2.59)
Ravindran 1997	54/500	84/502	1.65 (1.15—2.36)
Total	158/1386	166/1138	

* Evaluated at 22 weeks

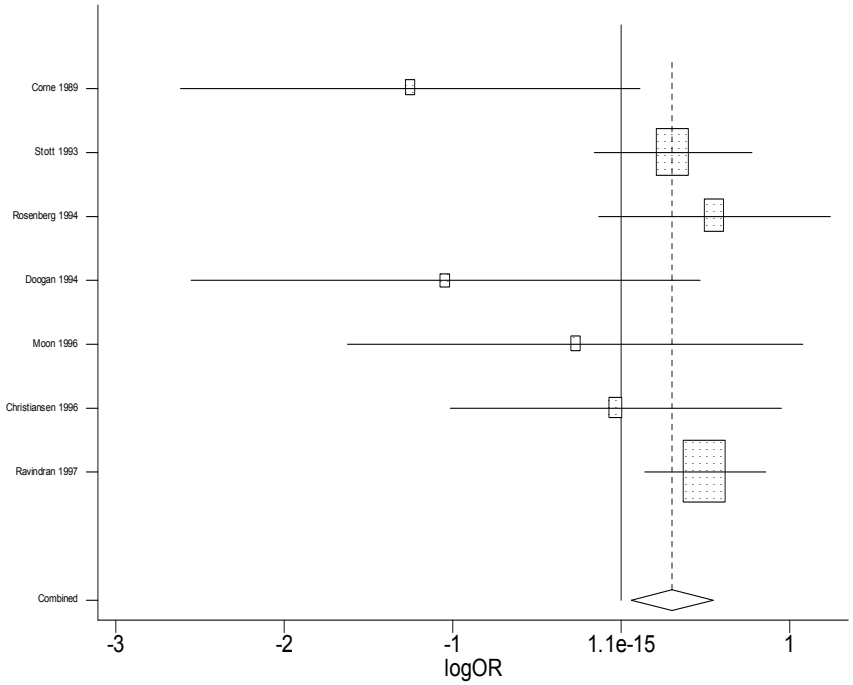


Figure 4.1: Graphical illustration of individual study results based on the conventional model

Table 4.3: Meta-analyses of the tolerability of SSRIs and TCAs in primary care

Models	Combined OR (95% Interval Estimate)
Conventional Fixed–Effect Model	1.35 (1.06—1.73)
Bayesian Model in favour of SSRIs (Prior OR: 1.50)	1.47 (1.21—1.77)
Bayesian Model in favour of SSRIs (Prior OR: 1.25)	1.23 (1.02—1.48)
Bayesian Model with 'Non-informative Indifferent' Prior (Prior OR: 1.00)	0.53 (0.11—2.52)
Bayesian Model with 'Non-informative' Prior in favour of TCAs (Prior OR: 0.75)	0.53 (0.11—2.52)

However, the result must be interpreted with care. The combined OR was largely influenced by one trial that favoured SSRIs strongly [180]. Omitting this trial would reduce the combined OR to 1.15 (95% C.I.: 0.83—1.60), thus suggesting that SSRIs were not significantly better tolerated than TCAs.

The Bayesian models were built next. Recall that 4 parameters must be fixed for the prior distributions for θ (combined odds ratio) and τ (between-study precision). Different prior values for θ reflect the different beliefs of the comparative tolerability of SSRIs and TCAs. To induce normality, the observed ORs were logarithmically transformed and the prior for θ refers to combined log OR. The final results were reported as OR by performing the necessary back-transformation. Next, the priors for τ were standardised as Gamma[c:0.01, d:0.01]. The choice of this distribution reflected the lack of prior information regarding between-study precision. Also, the number of burn-ins was set a priori at 500 and the Markov chain would thereafter be run another 1,000 times before the final analyses were conducted.

In the first attempt, the prior for θ was chosen as Normal[a:0.4055, b:100]. This reflects a highly-concentrated normal distribution with the overall OR believed to be 1.50, i.e., the SSRIs were better tolerated that TCAs. The selection of this prior was based on the general beliefs that SSRIs were associated with a significantly lower risk of toxicity [151]. The combined OR turned out to be 1.47 (Table 4.3). Since the 95% P.I. does not contain unity, SSRIs were interpreted to be better tolerated than TCAs.

In the second attempt, suppose an expert reported that SSRIs were better tolerated that TCAs, so the prior combined OR was fixed at 1.25, i.e. Normal[a:0.2231, b:100]. The combined posterior OR turned out to be 1.23 (Table 4.3). This is somewhat lower than that reported in the previous analysis.

In the next exploratory Bayesian analysis, a prior suggesting that SSRIs and TCAs were identical in terms of their tolerability was fitted. In this case, the "indifferent" prior for θ was chosen as Normal[a: 0, b: 0.000001]. Due to the low precision of 0.000001, this flat prior resembled that of a uniform distribution, thus suggesting that there was little prior information regarding θ. As shown in Table 4.3, the combined OR turned out to be 0.53 and the associated 95% P.I. was 0.11—2.52. As a result, one may interpret that SSRIs were not significantly better tolerated than TCAs. The fairly wide interval estimate was a result of the inclusion of 2 non-informative priors.

To further illustrate the properties of the proposed Bayesian model, a prior suggesting that TCAs were better tolerated than SSRIs was fixed next, i.e, Normal[a: -0.2877, b: 0.000001]. This was a "non-informative" prior, with OR fixed at 0.75. The result was identical to the previous analyses based on "indifferent" prior (Table 4.3).

Although the above analyses are unable to encode the probabilities for decision analysis directly, they do provide useful evidence on the risk of suffering from side effects by taking SSRI, while in comparison with TCA. For example, the Bayesian model favouring SSRI with a prior OR of 1.5, would deduce that a depressed patient is 1.47 times more likely to suffer from intolerable side-effects with SSRIs, when compared with his counterpart who is prescribed with TCAs.

The above analyses based on Bayesian models shared a very important common feature. The posteriors were dominated by their respective priors. This was the result of a less-concentrated likelihood, in view of the small number of selected trials. Moreover, Table 4.2 also shows that almost all individual ORs contain unity in their respective 95% C.I.s. (based on conventional model). Consequently the

combined posterior ORs were largely similar to the prior ORs. However, this point was not explicitly highlighted by the conventional model. In this case, the Bayesian analysis revealed more details of the data. It must also be emphasised that the above were conducted as sensitivity analysis. One must decide which prior to be fixed for final reporting.

In passing, note that the resultant posteriors based on the 4 different sets of priors were fairly normal and the Markov chains exhibited no obvious pattern of divergence after the burn-in values had been discarded (Figure 4.2).

4.1.4 Discontinuation from Treatment in the General Setting

It was not feasible to perform sub-group meta-analysis by drug class in the primary-care example as there were too few studies. In the next example, SSRIs were compared with 3 tertiary TCAs (amitriptyline, imipramine and clomipramine) separately. For all analyses, the priors for θ and τ were fixed as Normal[a:0.2231, b:100] and Gamma[c:0.01, d:0.01], respectively. The choice of priors reflected the belief that SSRIs were better tolerated than TCAs [184-185]. In addition, both the number of burn-ins and updates were set a priori at 1,000.

In the case where SSRIs were compared with amitriptyline, 31 studies were identified and selected. The conventional fixed-effect model was chosen to compare with the Bayesian model because statistical test suggested no strong evidence of study heterogeneity (p-value: 0.79). Both models showed favourable results for SSRIs but the Bayesian model's combined OR was substantially lower (Table 4.4).

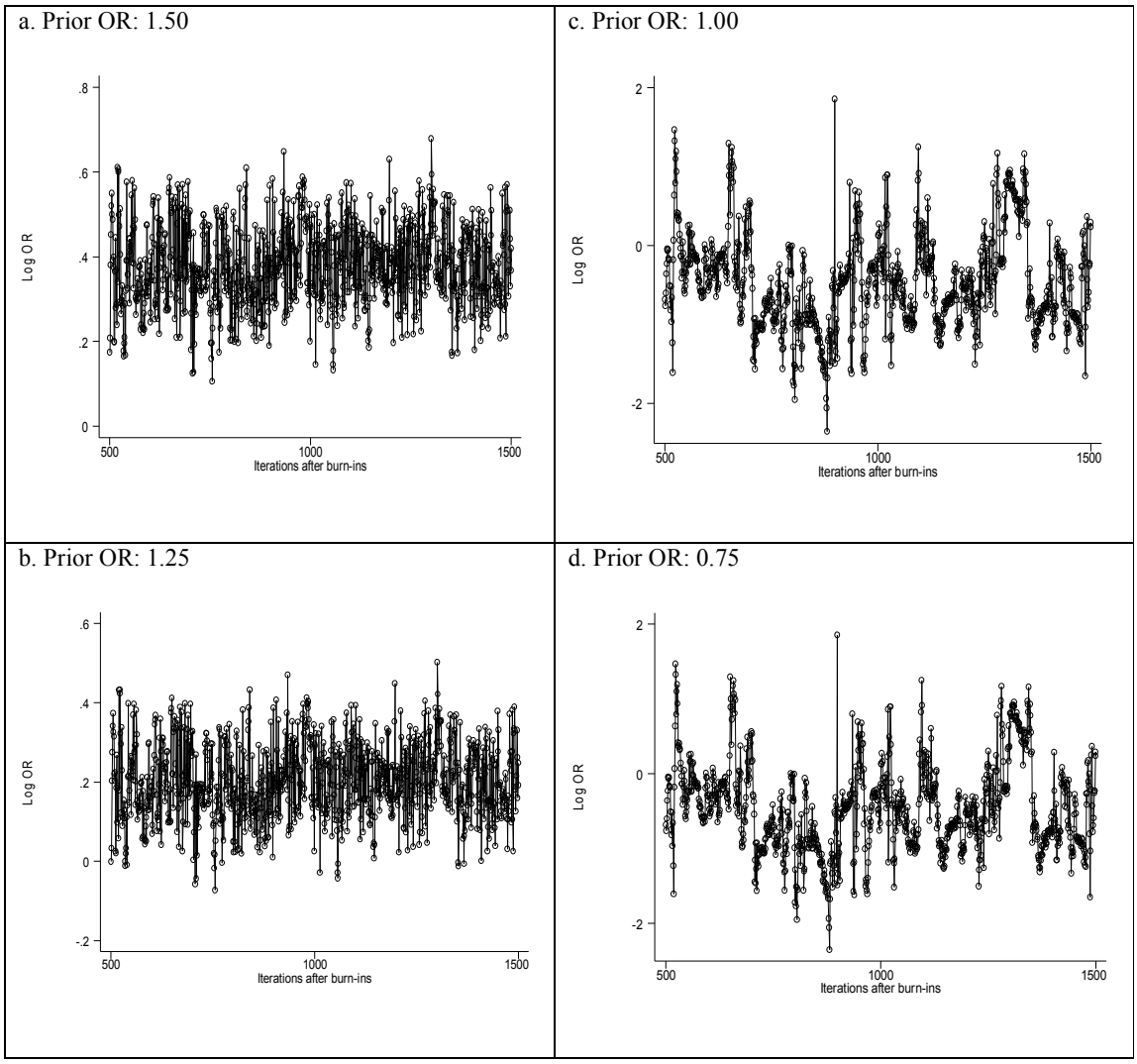


Figure 4.2: Iterative history of meta-analysis of the tolerability of SSRIs and TCAs in primary care

Table 4.4: Meta-analysis of the tolerability of SSRIs and TCAs in the general setting

Models	SSRIs vs. amitriptyline	SSRIs vs. imipramine	SSRIs vs. clomipramine
Conventional Model	OR: 1.57 (1.27—1.95)	OR: 1.48 * (1.09—2.01)	OR: 1.68 (1.24—2.26)
Bayesian Model in favour of SSRIs (Prior OR: 1.25)	OR: 1.34 (1.11—1.61)	OR: 1.26 (1.05—1.53)	OR: 1.29 (1.09—1.53)

* Based on random-effect model

A similar observation was made for comparison with imipramine where 29 studies were included for analysis. The Bayesian model's posterior combined OR was almost identical to the prior OR. The random-effect model was chosen for conventional analysis because there was strong evidence of study heterogeneity (p-value<0.01). In passing, note that the lower ends of the 95% C.I.s and P.I.s for both conventional and Bayesian analyses were close to unity.

Last but not least, there was again a large disparity in results when SSRIs were compared with clomipramine. Dominated by the prior, the posterior combined OR of the Bayesian model reported a less favourable effect for SSRIs when compared with the fixed-effect conventional model. With only 9 studies selected, there was no strong evidence of study heterogeneity according to the conventional analysis (p-value: 0.18).

The Markov chains exhibited no obvious pattern of divergence in the above-mentioned Bayesian analyses (figures not shown).

4.1.5 Discussion & Decision

Replication of experimental results has long been a central feature of scientific inquiry, and it raises questions concerning how to combine the results obtained. Meta-analysis is often defined as the statistical analysis of a collection of results from individual studies for the purpose of integrating the findings [186]. It involves the combination of quantitative evidence from studies that have investigated a common question.

The theoretical details of the conventional model for meta-analysis are well known [115-116, 118-119, 183, 187-188]. Following the rationale of conventional statistical theory, the parameter of interest is considered as an unknown but fixed

quantity that can be accurately estimated from data obtained from a proper literature search. Motivated by the current need for evidence-based medicine, the Bayesian model differs from the conventional approach in 2 aspects. First, it allows prior information—in the form of expert opinion—to be incorporated into analysis. Though subjective in nature, such information, it is argued, may provide a more realistic approach in data analysis. Many biomedical researchers may have accumulated a large amount of experience through practice and it is costly to ignore such information. Second, the analysis is conducted on the posterior distribution which summarises all the information, both prior- and data-based, that the analysts have about the unknown parameters.

As described, the proposed Bayesian model allows observed ORs to vary around their individual study-specific ORs, which in turn belong to a distribution characterised by the combined OR. It was preferred over a fixed-effect model for the above-mentioned analyses because there were differences in treatments (e.g., types of antidepressants, dose of drugs and treatment duration), types of patients (e.g., inclusion criteria, culture of drug compliance), experimental designs (e.g., with or without a placebo group) and type of statistical analyses applied. Consequently, it is naïve to assume that study heterogeneity does not exist even with the support of formal statistical tests. Moreover, such statistical tests may lack power in detecting the underlying differences among studies.

This case study aimed to analyse the tolerability of SSRIs and TCAs in patients with depressive disorders. As one of the most common illnesses that affect a large number of individuals in all countries, depression is a "whole-body" disorder affecting the nervous system, moods, thoughts and behaviour. As both SSRIs and

TCAs are effective in treating depression, the choice of medication depends mainly on patients' tolerability.

There was no convincing evidence from the primary care meta-analysis that SSRIs were better tolerated than TCAs. Of the 7 clinical trials considered, only 1 favoured SSRIs significantly. The Bayesian models demonstrated that the resultant posteriors were strongly influenced by the priors fixed before analysis. As such, the primary-care physicians must be vigilant when prescribing SSRIs (fluoxetine, praoxetine, citalopram and sertraline in particular).

In the general setting, however, the result was slightly more optimistic. Based on current findings, amitriptyline, imipramine and clomipramine were not as safe as SSRIs. This result conformed to the general beliefs. However, as in the analysis with primary-care patients, the posteriors of the Bayesian analyses were strongly dominated by the priors. Consequently, the safety of antidepressant therapies should be monitored carefully as patients who suffer from depression may experience different tolerability profiles.

The proposed Bayesian model (3.29-3.32) provides biomedical researchers an alternative approach for conducting meta-analysis. For future research, one may consider newer antidepressants and different types of patients such as the elderly.

4.2 An Alterative Screening Tool for Osteoporosis

4.2.1 Aim

Osteoporosis, a systemic skeletal disease characterised by low bone mineral density (BMD), causes considerable morbidity and mortality. Older persons with the condition are at much greater risk for developing hip fractures from accidents such as falls. It is estimated that about 26% of the elderly suffering from osteoporotic hip fractures would not survive within a year after the injury [189].

This becomes a major concern for Asian countries where populations are greying rapidly. It is projected that 50% of the world's hip fractures will occur in the continent by end of year 2050 [190]. In view of the ageing demographic structure and rising osteoporotic hip fractures rate, Singapore is expected to experience substantial inflation in related health care costs [189, 191]. This poses a serious challenge to the nation.

Due to a natural decline in bone density after menopause, the majority of osteoporotic fractures occur in older women. To permit prevention and early intervention, it is therefore important to identify postmenopausal women at risk of developing osteoporosis. However, mass screening using the dual x-ray absorptiometry (DXA)—widely regarded as the "gold standard" for diagnosing osteoporosis—is not feasible owing to its cost [192-193]. As such, it is useful to consider several well-cited indices [194-198] for identifying elderly subjects at risk of developing osteoporosis. These indices aim to assess how the risk factors, such as old age, low body weight, low level of estrogen and history of rheumatoid arthritis, are associated with low BMD.

Generally, these indices have moderate to high sensitivity but low specificity [199]. While widely applied in the Caucasian and Asian populations, most of the

indices were not validated in the Singapore context. As such, this paper aims to compare the sensitivity and specificity of the indices [194-198] for identifying Chinese elderly women (aged 55 years and above) in Singapore at risk of osteoporosis.

4.2.2 Data and Indices

One hundred and thirty-five free-living ambulant Chinese postmenopausal female subjects were recruited from a community in the eastern part of Singapore. Using housing type as a surrogate measure for socioeconomic status, the recruited subjects were representative of the socio-economic status of Singaporeans and had a wide range of body mass index (BMI). The sample size was determined based on estimation of the proportions of subjects in the different categories of BMI status (50% in the normal range of 18.5-22.9 kg/m^2, 30% in the overweight category of ≥23 kg/m^2 and 20% in the underweight category of <18.5 kg/m^2) [200-201]. Dropouts were replaced by matching the gender, housing type and BMI of subjects. Approved by the Medical Ethics Committee, Health Promotion Board (HPB), Republic of Singapore, the study took place in March 2003.

Upon given their informed consent to participate in the study, eligible subjects were invited to Changi General Hospital (CGH), Singapore, to complete a short screening questionnaire and BMD measurements. The subjects were interviewed by a trained interviewer on diet, physical activity, quality of life, smoking status and current medications.

BMD measurement of the femoral neck was performed by dual-energy x-ray absorptiometry (DXA) using a Hologic QDR 4500W densitometer (S/N 49088). All DXA measurements were conducted by a qualified radiologist. Body weight was measured on a calibrated mechanical scale, and height was measured with a wall-

mounted stadiometer, with subjects wearing light indoor clothing only (without shoes). Food intakes were also assessed using a validated interviewer-administered food frequency questionnaire (FFQ) [202].

The various indices considered in this study were Simple Calculated Osteoporosis Risk Estimation (SCORE), Osteoporosis Risk Assessment Instrument (ORAI), Age Bulk One or Never Estrogens (ABONE), Body Weight (WEIGHT) and Osteoporosis Self-Assessment Tool for Asians (OSTA) [194-198]. Table 4.5 shows the indices' scoring systems and published results.

SCORE, one of the first attempts to develop predictive rules for screening osteoporotic patients, was developed in an American cohort of 1,102 postmenopausal women of all ethnic background aged 45 years and above [194]. ORAI was developed and validated in the Canadian Multicentre Osteoporosis Study comprising 926 non-institutionalised female subjects aged 45 years and above [195]. ABONE was developed with the data of 1,610 postmenopausal white women [196] and WEIGHT was based on 175 randomly-selected women (aged 28-74 years) in Sweden [197]. OSTA was developed with 860 postmenopausal Asian women in 8 countries (PR China, Taiwan, Hong Kong, Korea, Malaysia, Singapore, Thailand and the Philippines) [198]. It had been validated in various studies consisting of post-menopausal Japanese, Korean and Hong Kong women [198, 203-204].

The primary outcome considered in this case study was femoral neck BMD on the left region determined by DXA. Based on the World Health Organisation (WHO) criteria [205], the subjects were classified as osteoporotic (BMD T-score$\leq$-2.5) or non-osteoporotic (BMD T-score>-2.5).

An OSTA-like index based on age and body weight was also constructed based on Bayesian logistic regression (3.16). This helps to facilitate comparison

between Bayesian and conventional analysis. A concentrated prior Beta[5, 10] was adopted after detailed discussion among the investigators, with expert opinion given by a consultant chemist specialising in the subject matter. The choice of beta prior reflected the strong belief that about 30% of the aged subjects were osteoporotic.

With the published cut-off points (Table 4.5), the sensitivity and specificity for the indices were computed in the study sample of 135 subjects. The Receiver Operating Characteristic (ROC) curves [147-148] were then generated to determine empirically the indices' optimal cut-off points, sensitivity and specificity in the same study sample. The optimal cut-off of an index was identified at the point nearest to the top left-hand corner of the ROC curve [206]. The indices were compared by means of area under the curves (AUC) at 1% level of significance, taking into account the numerical difference in AUCs and their respective standard errors, which in turn depend on the correlation between ROC curves [149]. Other things being equal, a low correlation will result in a large standard error, thus suggesting the AUC difference to be non-significant.

4.2.3 Comparison of Indices

Interrupted by the SARS outbreak, the study was concluded in August 2003. Table 4.6 summarises the characteristics of subjects included in the study. The mean age of the postmenopausal women was about 68 years. Overall, the proportion of subjects found to be osteoporotic based on femoral neck BMD T-scores of ≤-2.5 was about 24%. The mean body weight and mean BMI were 58.8 kg and 25.4 kg/m^2, respectively.

The sensitivity and specificity based on the published cut-off points of the indices (with femoral neck BMD) are presented in Table 4.7. SCORE, ORAI and

ABONE attained the maximum sensitivity of 100%, but their ability to correctly identify non-osteoporotic subjects was far from satisfactory. The body weight (WEIGHT) criterion also had a high sensitivity of 97.0% and a low specificity of 18.6%. OSTA had the same sensitivity as WEIGHT, but its specificity was highest among all 5 indices (43.1%).

The results based on generated ROC curves are presented in Table 4.8. For OSTA and WEIGHT, lower cut-off values represent higher risk of osteoporosis (low BMD). On the other hand, higher values for the other indices, namely ORAI, SCORE and ABONE, indicate higher risk of developing osteoporosis.

OSTA had the highest discriminatory power, with an estimated AUC of 0.82 (Figure 4.3). At the cut-off point of -2, OSTA achieved a sensitivity and specificity of 90.9% and 58.8%, respectively (Table 4.8). In passing, note that the identified optimal cut-off was lower than the published cut-off point of -1 (Table 4.5).

With Bayesian augmentation, OSTA achieved a marginally higher specificity at 62.8% (Table 4.8), while its AUC and sensitivity remained almost unchanged. One possible explanation is that the collected data were fairly consistent. Unlike the original OSTA, the Bayesian-augmented index may be used in the clinical context for advising patients with respect to their probability of contracting osteoporosis. This is crucial for making decisions regarding patients' need to attend special treatment.

For example, an 80 year-old female weighing 40 kg is deemed to be at risk as she has a 36% chance of contacting osteoporosis (Table 4.9). Other probabilities may be encoded with different combinations of age and body weight. The model is found to have provided a satisfactory fit to the collected data. No influential outliers were detected that might affect the fit. Sensitivity analysis with "non-informative" prior, say Beta[1, 1], suggested minor changes to the model.

The next index with the largest estimated AUC was SCORE, with ≥8 identified as the optimum cut-off point (sensitivity: 93.9%; specificity: 60.8%). This optimal cut-off point, determined empirically from the analysis, was higher than the published result (Table 4.5).

The body weight criterion, with an estimated AUC of 0.78, identified 54.0 kg as the optimum cut-off point. As such, the criterion suggests that subjects under 54.0 kg were at risk. While its sensitivity was about 70%, the criterion managed to identify 77.5% of the non-osteoporotic subjects correctly.

With an estimated AUC of 0.76, ORAI correctly identified about 76% of the osteoporotic subjects at an optimum cut-off point of ≥20. Its specificity was about 67%.

The AUC of ABONE was 0.70. In terms of sensitivity, ABONE correctly identified about 82% of the osteoporotic subjects at cut-off point 3 (possible range: 1—3). However, its specificity of 55.9% was lowest among the indices considered in this study.

There was no significant difference in AUCs among OSTA, SCORE, ORAI, ABONE and WEIGHT in the abovementioned analysis. Although the numerical difference in AUCs between OSTA and ABONE was more than 10% where femoral neck BMD was considered (Table 4.8), the result was not significant in view of the low correlation between the two ROC curves [149].

Table 4.5: Indices for identifying osteoporotic subjects

Index	Published Cut-Off Point/ Results	Scoring System
Simple Calculated Osteoporosis Risk Estimation (SCORE)	Score≥6 Sensitivity: 89.0% Specificity: 50.0%	Race: 5 if not black Rheumatoid arthritis: 4 if applicable History of minimal trauma fracture after age 45 years: 4 for each fracture of the wrist *, hip or ribs (maximum points=12 points) Age: 3 times first digit of age in years Estrogen therapy: 1 if never used Body weight: - 1 times weight in pounds (lb) divided by 10 and truncated to an integer
Osteoporosis Risk Assessment Instrument (ORAI)	Score≥9 Sensitivity: 97.0% Specificity: 41.3%	Age: 15 if ≥ 75 years 9 if 65 – 74 years 5 if 55 – 64 years Body weight: 9 if < 60 kg 3 if < 60.0 – 69.9 kg Estrogen use: 2 if not currently taking estrogen
Age, Bulk, One or Never Estrogens (ABONE)	Score≥2 Sensitivity and Specificity not published	Age: 1 if > 65 years Body weight: 1 if < 63.5 kg Estrogen use: 1 if never used oral contraceptives or estrogen therapy for at least 6 months
Body Weight Criteron (WEIGHT)	≤70 kg Sensitivity: 94.0% Specificity: 36.0%	Body weight ≤ 70 kg
Osteoporosis Self-Assessment Tool for Asians (OSTA)	Score≤ –1 Sensitivity: 91.0% Specificity: 45.0%	0.2 × [body weight (kg) – age (years)]

* Forearm/wrist was included as a history of wrist fracture

Table 4.6: Characteristics of study sample

Variables	Mean (sd) *	Range
Age (years)	68.4 (5.5)	56.9 — 80.1
Body weight (kg)	58.8 (10.1)	35.5 — 86.5
Height (m)	1.52 (0.05)	1.40 — 1.64
Body Mass Index (kg/m²)	25.4 (3.9)	15.9 — 35.5
Femoral Neck BMD (g/cm²)	-1.70 (1.01)	-4.5 — 1.9
Classified Status of Osteoporosis (%) †		
Non-Osteoporotic	102 (75.6%)	—
Osteoporotic	33 (24.4%)	
Current Smoking Status (%)		
Non-Smokers	128 (94.8%)	—
Smokers	7 (5.2%)	
Estrogen Use (%)		
No	133 (98.5%)	—
Yes	2 (1.5%)	

* Unless otherwise specified
† Based on WHO guideline

Table 4.7: Sensitivity and specificity based on published cut-off Points for identifying osteoporotic subjects with femoral neck BMD T-score≤-2.5

Index	Published Cut-Off Point	Sensitivity	Specificity
Simple Calculated Osteoporosis Risk Estimation (SCORE)	Score≥6	100.0%	30.4%
Osteoporosis Risk Assessment Instrument (ORAI)	Score≥9	100.0%	9.8%
Age, Bulk, One or Never Estrogens (ABONE)	Score≥2	100.0%	16.7%
Body Weight Criterion (WEIGHT)	≤70kg	97.0%	18.6%
Osteoporosis Self-Assessment Tool for Asians (OSTA)	Score≤ –1	97.0%	43.1%

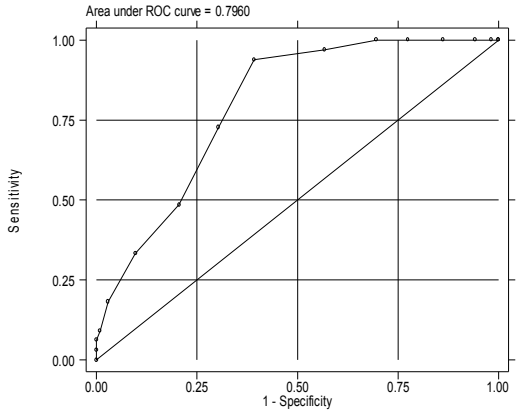
Table 4.8: Empirically-determined cut-off points, sensitivity and specificity based on ROC curves for identifying osteoporotic subjects with femoral neck BMD≤-2.5

Index	Range	Suggested Cut-Off Point	AUROC (95% C.I.)
Simple Calculated Osteoporosis Risk Estimation (SCORE)	1—15	Score≥8 Sensitivity: 93.9% Specificity: 60.8%	0.80 (0.72—0.87)
Osteoporosis Risk Assessment Instrument (ORAI)	7—26	Score≥20 Sensitivity: 75.8% Specificity: 66.7%	0.76 (0.68—0.84)
Age, Bulk, One or Never Estrogens (ABONE)	1—3	Score=3 Sensitivity: 81.8% Specificity: 55.9%	0.70 (0.63—0.78)
Body Weight Criterion (WEIGHT)	35.5—86.5 kg	Body weight<54.0 kg Sensitivity: 69.7% Specificity: 77.5%	0.78 (0.69—0.87)
Osteoporosis Self-Assessment Tool for Asians (OSTA)	-7—3	Score≤-2 Sensitivity: 90.9% Specificity: 58.8%	0.82 (0.75—0.90)
Bayesian-Augmented Osteoporosis Self-Assessment Tool for Asians (OSTA)	-0.9— -0.6	Score≤-0.7 Sensitivity: 90.9% Specificity: 62.8%	0.82 (0.75—0.89)

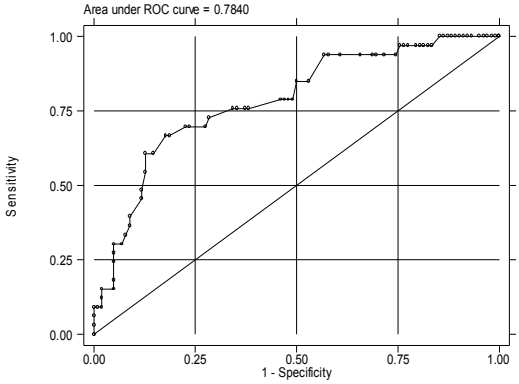
Table 4.9: Bayesian logit analysis of osteoporosis (based on OSTA findings)

	Coefficient	95% P.I.
Age (years)	5.21×10^{-3}	1.99×10^{-3}—8.44×10^{-3}
Body weight (kg)	-4.36×10^{-3}	-6.18×10^{-3}— -2.55×10^{-3}
Intercept	-0.82	-1.08— -0.56

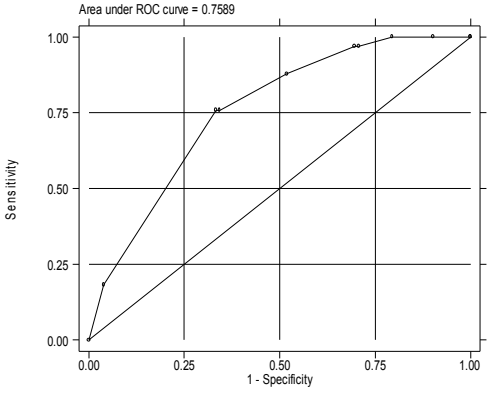
a. Simple Calculated Osteoporosis Risk Estimation(SCORE)



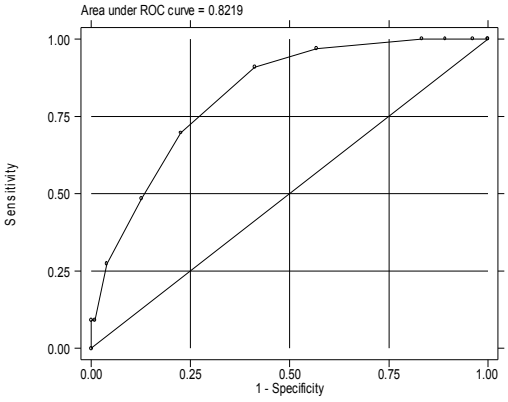
e. Body Weight Criterion (WEIGHT)



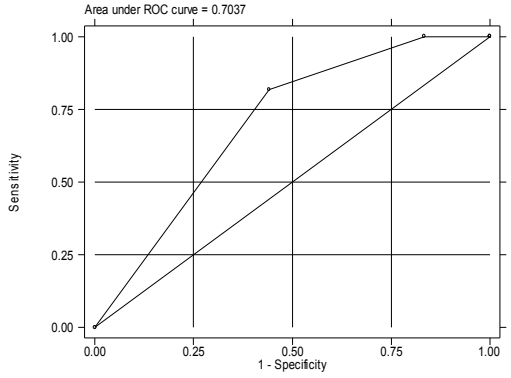
b. Osteoporosis Risk Assessment Instrument (ORAI)



d. Osteoporosis Self-Assessment Tool for Asians (OSTA)



c. Age, Bulk, One or Never Estrogens (ABONE)



d. Osteoporosis Self-Assessment Tool for Asians (OSTA) with Bayesian augmentation

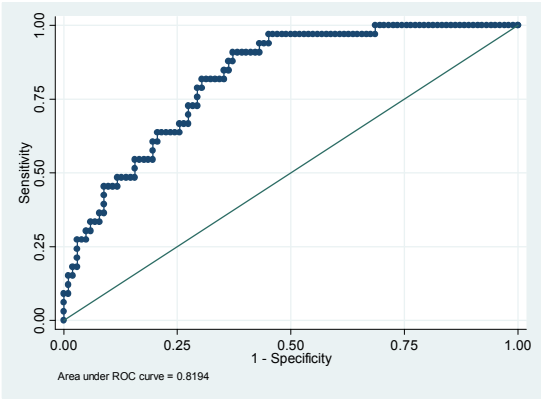


Figure 4.3: ROC curves based on femoral neck BMD≤-2.5

4.2.4 Discussion

Simple risk assessment tools provide a quick and inexpensive way for identifying persons at risk of osteoporosis and hip fractures. Such indices are extremely useful in communities where access to BMD measurement is limited and costly.

This empirical study showed that the various indices considered were useful in identifying postmenopausal elderly Chinese females with osteoporosis. Based on ROC analysis with femoral neck BMD, the sensitivity of the indices was above 69% (Table 4.8). A high sensitivity is essential as it provides reliable evidence for clinicians to start early treatment for patients at risk of osteoporosis. All indices yielded a specificity of no less than 55% (Table 4.8).

Numerically, OSTA yielded the highest AUC at 0.82, with a sensitivity and specificity of 91% and 59%, respectively (Table 4.8). While its AUC was not significantly higher than SCORE, OSTA is a more convenient tool in the sense that only age and body weight are required in computation. This is a desirable feature of OSTA as other indices require more detailed information such as the use of estrogen and history of rheumatoid arthritis. While more complicated than WEIGHT, OSTA yielded a higher AUC and sensitivity.

There was no significant difference in AUCs among OSTA, SCORE, ORAI, ABONE and WEIGHT. This could be partly due to the limited sample size [200]. Moreover, the correlations between ROC curves were generally low. It is also not clinically significant in detecting a less than 10% difference in AUCs (Table 4.8).

It is worthwhile to note that the original published cut-off of OSTA was $\leq$-1, based on femoral neck BMD of a cohort of 860 women in 8 Asian countries (Table 4.5). On the other hand, this study suggests a lower cut-off at -2 (Table 4.8). The

difference in cut-off points may be explained by demographic differences in the samples. The cohort in the original OSTA study included younger women (reported age range: 45–88 years) while this study only considered postmenopausal women aged 55 years and above. This highlights that OSTA's optimal cut-off point may vary with different age groups. Moreover, the original OSTA study used sensitivity to select the optimum cut-off while this study adopted the conventional approach by giving equal attention to both sensitivity and specificity.

As such, further studies may be carried out to determine how OSTA perform in different age groups (e.g., 55-69 years, 70-80 years and ≥80 years) among the postmenopausal women. In addition, subjects from other ethnic groups may also be included in future studies. With diversified demographic characteristics in subjects and a larger sample, a more conclusive answer to the index's generalisability and applicability may be derived.

4.2.5 Decisions

This empirical study showed that OSTA is an effective index for identifying postmenopausal Chinese women at risk with osteoporosis. That is, OSTA has a high discriminatory power characterised by high sensitivity and specificity. It is therefore applicable in the clinical context. Further work may be carried out to evaluate its performance in different age range and other ethnic groups in Singapore. The Bayesian logit model may also be applied in the development of new indices, which may involve more variables such as the use of estrogen therapy and past history of fracture.

4.3 Sulindac as an Effective Treatment for Colonic Polyps

4.3.1 Aim

The colon is the part of the digestive system where the waste material is stored. Tumors of the colon are growths arising from the inner wall of the large intestine. Benign tumors (polyps) do not invade nearby tissues or spread to other parts of the body and can be easily removed from colonoscopy. However, polyps can become malignant if not removed from the intestine. It is believed that most cancers of the large intestine are developed from polyps [207-209].

Polyps vary considerably in size; the larger the polyp the greater the risk it becomes cancerous. They usually do not cause symptoms, but when they do, the most common is bleeding from the rectum. A large polyp may cause abdominal pain or obstruction. Adenomatous polyps, which consist primarily of glandular cells inside the large intestine, are likely to become malignant. A family history of polyps increases the risk of colon cancer and the term "familial polyposis" refers to the condition where 100 or more polyps develop through the large intestine during childhood or adolescent. Caused by a germ-line mutation in the adenomatous polyposis coli (APC) gene, familial adenomatous polyposis (FAP) is therefore a predisposition syndrome characterised by the formation of a large number of pre-cancerous colonic polyps. In nearly all untreated people, the polyps eventually develop into colon cancer in their middle age.

Colon cancer is common among Singaporeans aged 50 years and above, as over 1,000 incident cases are diagnosed every year. As such, persons of this age group are constantly reminded to screen for the disease. As mentioned before, persons with a family history of colon cancer have a higher risk of developing the cancer themselves. Smoking, over-eating, physical inactivity and insufficient calcium

and fiber intake diet are other prominent risk factors. Exposure to air and water pollution, particularly to industrial cancer-causing substances (carcinogens), also play a role. Although it is a common cancer among developed countries, there is no proven cure for the disease. Shrouded with uncertainties, the best advice is to identify the cancer early and have it removed before it begins to spread in the body. As a result, early treatment of polyps is helpful.

This case study aims at reviewing the results of a well-known randomised, double-blind, placebo-controlled trial testing whether or not sulindac, a non-steroid anti-inflammatory (NSAID) agent, could reduce the size of colonic polyps in patients with FAP [210]. The mechanism by which sulindac causes polyp regression is unknown [211].

Essentially, this case study attempts to re-analyse the published trial data by combining with other available evidences. This may shed light on the effectiveness of sulindac based on the Bayesian perspective.

4.3.2 Data

Altogether, 22 patients were randomised to receive sulindac (150 mg orally twice daily for 9 months) or placebo in this trial. There were 13 (59.1%) male subjects in the sample. With an average age of 24.1 (range: 13—50 years), the sample contained an equal number of subjects with and without sulindac. The average baseline poly size of the sample was 3.29 mm (range: 1.7—5.5 mm). This was reduced to 2.51 mm (range: 0.4-4.4 mm) after 12 months. The complete listing of data can be found in reference [210]. In the following analysis, only the 12-month data were considered and the primary end-point was polyp size (mm). Baseline polyp size was included as one of the predictors in the analysis.

To ascertain whether sulindac was effective in reducing the size of polyps, the result from a Bayesian linear relational model (3.42-3.44) was compared with that of the conventional regression model [91, 206]. Analysed with Stata 9.0, all statistical analyses were conducted at 5% level of significance.

4.3.3 Effectiveness of Sulindac

The conventional model shows that sulindac treatment was effective in reducing the polyp size at 12 months (see Table 4.10). The average polyp size for patients receiving sulindac was about 1.3 mm lower than those receiving placebo, after adjusting for the baseline measurements. The analysis was based on 19 observations (sulindac: 9; placebo: 10) as there were no data recorded for 3 patients at 12 months.

However, there are mixed findings from other studies [207-209, 211-212]. To be conservative, consider a fairly precise "sceptical" prior $\boldsymbol{\beta}_0$=[0, 0, 0]' with $\boldsymbol{\Sigma}_0$:= $\text{diagonal}_{3\times3}(10^{-1})$ for the following Bayesian analysis. This prior suggests that sulindac is not effective in reducing polyp size. Moreover, baseline polyp size was also not associated with the subsequent measurements at 12 months. The inverse gamma distribution for σ^2 was fixed as Inverse Gaussian[3, 3]. As shown in Table 4.10, the posterior coefficients are closer to the priors, with slightly smaller standard errors, and consequently narrow 95% P.I.s. The Bayesian model shows that there could be no sufficient evidence suggesting that sulindac is effective in reducing polyp size for patients with FAP. The drug may not have the colon cancer prevention properties once hoped for.

Table 4.10: Linear regression analyses on 12-month polyp size

	Conventional Model Coefficient	95% C.I.	Bayesian Model Coefficient	95% P.I.
Baseline polyp size (mm)	0.21	-0.38—0.80	0.60	0.20—1.00 *
Treatment				
0: Placebo	—	—	—	—
1: Sulindac	-1.29 *	-2.55— -0.03	-0.28	-1.29—0.73
Intercept	2.42 *	0.24—4.60	0.43	-0.71—1.57

* Statistically significant at 5%

The residuals of the Bayesian model were fairly symmetrical. There was no evidence of violation of model assumptions such as heteroscedasticty. No influential outliers were identified (Figure 4.4a).

4.3.4 Probability Encoding

The Bayesian model may be used for predicting the prognosis of patients with or without sulindac. For example, a patient with a baseline polyp size of 2 mm is expected to reduce to 1.3 mm after 12 months on sulindac. However, the expected size of polyp is expected to reduce to only 1.6 mm if he were on placebo.

With these expected values and assuming the underlying normal distribution is valid, one is able to develop a prototype decision model (Figure 4.4b). For the above-mentioned patient with a baseline polyp size of 2 mm, the probability of reducing the polyp size with sulindac at 12 months is 0.71. This also imply that his condition deteriorates with a probability of 0.29. On the other hand, if the patient decides not to have sulindac, his probability of a better prognosis is 0.62.

Based on expert testimony, the probability of eventual development of colon cancer is assessed to be 0.5 should the patient's condition fail to improve after 12 months (Figure 4.4b). By inspection, it is obvious that a rational patient should choose sulindac for treatment even without eliciting his utilities.

4.3.5 Recommendations

However, one must be reminded that this is an incomplete representation of the relevant decision analysis as the tree diagram should also include other decision and chance nodes like colonscopy procedure, drug safety, future complications or side-effects, and possibly the outcomes of surgery, chemotherapy and radiotherapy

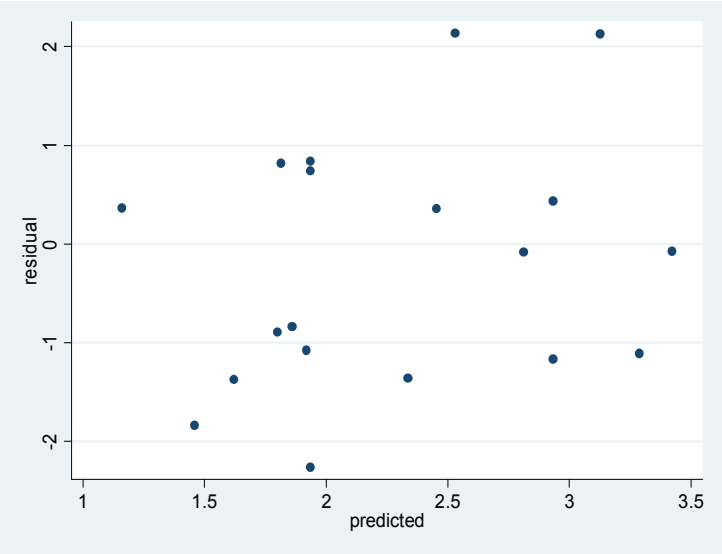
should colon cancer develops. Besides taking NSAIDs, polyps may be removed with colonoscopy procedure using a cutting instrument or an electrified wire loop. If the polyps cannot be removed during colonoscopy, abdominal surgery may be recommended.

If new polyps appear rapidly the rectum must also be removed. A patient may suffer from much inconvenience in daily life as an opening is created through the abdominal wall from the small intestine. This procedure is called ileostomy. Body wastes are eliminated through the ileostomy into a disposable bag.

Surgery is the only recommended treatment of colon cancer. Should the cancerous cells divide rapidly and spread beyond the colon, chemotherapy or radiotherapy may be applied after surgery. The former involves the use of drugs to kill the cancerous cells, while the latter applies radiation on the original site of the cancer in an attempt to control the disease.

Unfortunately, there is a lacuna of published evidences in the above-mentioned treatment and surgical procedures. The results obtained from this analysis may therefore serve as prior evidence for future studies.

a. Residual plot



b. Prototype decision tree

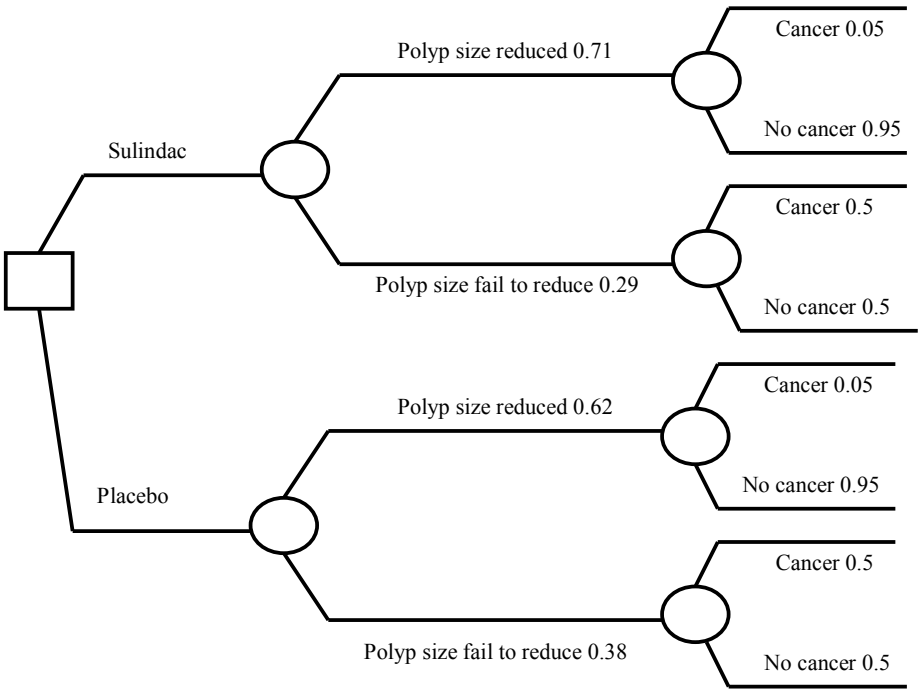


Figure 4.4: Residual plot and decision tree based on Bayesian linear relational model

4.4 Ocular Complications of Dengue Fever

4.4.1 Aim

Dengue fever (DF) is a viral infection caused by Flavivirus in humans and the mode of transmission is via the bite of an infected Aedes aegypti mosquito [213]. Accounts for worldwide cases of illness in excess of 100 million per year, the infection is common in the tropics, subtropics and warm temperate regions [214]. With a relatively high average temperature at around 28°C, Singapore is a potential breeding ground for DF and year-round transmission is often observed. A surge in cases has been observed in 1992, 1998 and 2004, which recorded an alarming annual incidence of 9,459 cases [215].

DF is characterised clinically by abrupt onset of fever after 2 to 7 days of incubation. Patients often suffer from severe malaise, headache, arthralgia, cough, sore throat, nausea, vomitting, anorexia and altered taste sensation. The disease is more severe in adults who usually suffer from high fever, headache and intolerable body aches. It is so painful that DF is often described as "breakbone fever". A transient macular rash may be seen on day 1 or 2 of illness, followed by a second maculopapular rash on day 3 to 6 of illness which typically involves the trunks, limbs and face but sparing the palms and soles. Blood dyscrasias may include thrombocytopenia and neutropenia (leukopenia). DF is usually self-limiting, but patients with the severe form (commonly known as dengue hemorrhagic fever or dengue shock syndrome) may present with bleeding and shock.

Recently, there has been increasing reports of dengue-related ocular complications in the literature [216-228]. Many of the reported cases had visual symptoms either in the form of blurring of vision or scotomas, which could represent one spectrum of the disease in which the ophthalmic complications were significant

enough to cause visual symptoms. However, there were some cases who presented with no visual symptoms. Aimed at detecting the incidence of dengue related ocular complication, the following analysis involved patients who were diagnosed and admitted to the Communicable Disease Centre, Republic of Singapore, with DF from late September 2005 to early January 2006.

4.4.2 Methods

This was a prospective cohort study. The diagnosis of DF was based on a combination of clinical findings correlated to positive results from dengue serology (dengue immunoglobulin IgM and IgG sero-conversion), polymerase chain reaction (PCR) or both. Patients were recruited to the study at day 6 to 8 of illness. This timing was chosen because in several of the previous case series it was found that the onset of visual symptoms and detection of ocular complications occurred around 7 days from the onset of DF [217-219, 224, 228].

All patients were asked to describe, with the aid of an Amsler grid, any visual symptoms experienced, whether they were scotoma, metamorphopsia or any other conditions. This was followed by fundus photography using a Zeiss FF450 (Carl Zeiss Inc, Germany) fundus camera with a Kodak DCS620 (Kodak Inc, USA) digital back to obtain 50° field image per eye. The images taken were reviewed by an ophthalmologist on the same day. Patients with abnormal fundal photos were reviewed by an ophthalmologist the following day who took detailed history and performed dilated fundal examination using slit lamp biomicroscopy. In order to eliminate unnecessary confounding factors from the analysis, patients with pre-existing ocular problems were excluded. Patients with both abnormal fundus photography and fundal biomicroscopy were considered as having ocular

complication. Further clinical investigation and management were made on a case by case basis.

To determine if the identified factors (age, vasculitis, and visual symptoms, etc.) were associated with ocular complications, Bayesian logistic regression (3.16) was applied. To assess the model's out-sample predictive ability, the original data set was randomly split into two sub-samples. Based on the training sub-sample with 80 observations, an auxiliary Bayesian logit model was built. It was applied to generate the probabilities of contracting ocular complications and the results were compared with the true status in the testing sub-sample. To quantify its predictability, a ROC curve [147-148] was generated. Analysed with Stata 9.0 and R 2.4.1, all statistical tests were conducted at 5% level of significance.

4.4.3 Results

A total of 131 patients admitted to hospitals during the study period for management of fever with clinical suspicion of DF were recruited into the cohort study. Of these, 7 were later found to be negative for dengue serology and were subsequently excluded from analysis.

Of the remaining 124 patients with DF confirmed by clinical symptoms and positive results from dengue serology, PCR or both, 22 (17.7%) reported an abnormal fundus photography. Of these, 12 (9.7%) were found to have retinal abnormalities on dilated fundal examination using fundus biomicroscopy. The discrepancy between the number of patients with abnormal fundus photography and abnormal fundal biomicroscopy examination was partly due to photographic artefacts.

The mean age of the recruited patients was 32 years (s.d.: 9.37; range 16-62) and the majority were males (76.2%). Twenty-four patients dropped out of the study

either before having fundus photography or fundal examination, for reasons including not feeling well enough to undergo the examination or that they were discharged from the hospitals before the examination was conducted.

Of the 100 patients who completed the study, 6 had bilateral retinal abnormalities at the time of examination, while 6 were presented with unilateral abnormalities. It was also observed that some patients presented with more than 1 type of retinal abnormality. As for the types of retinal abnormalities found, 9 had retinal haemorrhage, 5 had cotton wool spots, 2 had vascular sheathing, 1 had macular oedema and 1 had optic disc haemorrhage. Four patients reported visual symptoms on direct questioning and testing with the Amsler grid. The symptoms included blurring of vision (2), scotoma (1) and metamorphopsia (1). Of these 4 patients, 2 were confirmed to suffer from ocular complications.

There was no directly relevant prior evidence for analysing the data with proposed the Bayesian analysis, although the decision-maker and analysts believed that the complication rate could be around 10%. This was based on case series reported prior to the study. As such, the model utilised Beta[1, 9] as the prior distribution.

DF patients with visual symptoms were found to be likely to develop ocular complications when compared with those who were asymptomatic. However, the result was not statistically significant (95% P.I.: -0.02—0.41). Age was also not significantly associated with the occurrence of complications (95% P.I.: -2.87×10^{-3} — 8.22×10^{-3}). The results are depicted in Table 4.11. As there were only 2 patients with vasculitis, a disorder developed from inflammation of the blood vessels, the predictor was omitted from analysis.

Therefore, a 40 year-old DF patient presented with visual symptom would have a 12% chance of developing ocular complications based on the Bayesian logit model (Table 4.11). The model was found to be satisfactory, based on sensitivity analysis. Unfortunately, its out-sample predictive ability was far from excellent as the area under the ROC curve (AUC) was 0.53.

4.4.4 Discussion

There is an increasing awareness of the ocular complications associated with DF in Singapore [216-218, 229]. Previous data reported in the indexed medical literature have been mainly limited to case reports and case series. To the best of the analysts' knowledge, there has not been any reported data on the incidence of ocular complication in DF.

In this study, the incidence of ocular complications associated with DF was found to be around 10%. Bleeding tendency as a result of thrombocytopenia in DF may lead to retinal and disc hemorrhages, and the onset of ocular complications was closely correlated to the nadir of thrombocytopenia [217-218]. However, some of the other ocular complications observed like vascular sheathing and macular oedema could point to an inflammatory component of the complication. It has been suggested that the pathogenesis of DF could involve immune clearance by way of induction of cross-reactive T-cell memory, T-cell proliferation and recognition of dengue viral antigens on infected monocytes by sensitized CD4+CD8- and CD4-CD8+ cytotoxic T cells, which result in the release of cytokines with vasoactive and procoagulant properties [230-231]. However, the exact pathogenesis of ocular complications of DF is not clear and is beyond the scope of the present study.

The results of this study need to be interpreted with extra care because of several limitations. The selection of a cohort group of an inpatient population might have reflected one spectrum of the diseased population. Also, the use of fundal photography allowed only a limited view of the fundus and abnormalities in the peripheral retinal could be missed. Furthermore, the 2-dimensional images obtained by fundus photography might not be able to detect subtle changes like macular oedema. Finally, ocular complications like anterior uveitis, which has been previously described in dengue-related ocular complication would not have been easily detected. Nevertheless, the study could serve as a baseline for comparison with further studies.

4.4.5 Decisions

Healthcare workers dealing with DF patients need to be aware of the ocular manifestations of the disease, even in those who reported no visual complaints. There is an urgent need to further the understanding of the epidemiology and pathophysiology of such ocular complications.

Table 4.11: Bayesian logit analysis of ocular complications of dengue fever

	Coefficient	95% P.I.
Age (years)	2.68×10^{-3}	-2.87×10^{-3} — 8.22×10^{-3}
Visual Symptoms		
0: Without	Reference	—
1: With	0.19	-0.02 — 0.41
Intercept	-2.30	-2.49 — -2.12

4.5 Predicting Mortality after Intracerebral Haemorrhage

4.5.1 Aim

Stroke is the third leading cause of deaths in Singapore [232]. Also known as brain attack, stroke is a sudden interruption in the blood supply of the brain. Most strokes are caused by an abrupt blockage of arteries leading to the brain. However, intracerebral haemorrhage (ICH) occurs when a diseased blood vessel within the brain bursts, thus allowing blood to leak inside the brain. The sudden build-up in blood pressure within the brain can cause damage to the brain cells, and subsequently unconsciousness and death. The main risk factors are hypertension, smoking, hyperlipidemia and diabetes mellitus, and the disease may occur in all age groups. Less common than ischaemic stroke, ICH makes up about 12% of all stroke cases in Singapore. This is substantially lower than other countries with a predominant Chinese population, such as PR China (46.0%), Hong Kong (27.1%) and Taiwan (23.0%). Different subtypes of strokes have different pathophysiological mechanisms, morbidity and mortality.

This primary aim of this study was to identify the predictors of 30-day mortality after ICH. This is a crucial task to neurologists and the public health service at large because stroke patients need to be evaluated closely at all stages. Otherwise, the hospitals concerned are unable to gear up to cope with the increase in stroke load. The analysis will help the relevant emergency departments to identify patients who might benefit from intensive care. Moreover, a timely prediction of patients' prognosis also enables care-givers to decide what intensive rehabilitation is required for post-stroke care. Based on the encoded probabilities, the patients are informed of their chance of survival. For those with an unfavourable prognosis, advanced directives (including end-of-life issues) may be established because the recurrence

and progression of ICH are unpredictable. This may help neurologists and other care-givers to determine the most cost-effective treatment should the patients become unable to make decisions.

The secondary objective of this study was to evaluate the predictive performance of various candidate statistical and data-mining models. This will not only help to determine which model is most reliable and useful in predicting primary ICH patients' mortality, but also enable the analysts to develop more insight into the problem.

4.5.2 Methods

The data were obtained from a registry established at the National Neuroscience Institute (NNI), Republic of Singapore. Founded in 1999, NNI is a national specialist centre for managing patients who suffer from all diseases of the nervous system. The investigation team for this study comprised two neurologists and two medical statisticians.

Multivariate logistic regression (logit) based on maximum likelihood [133] was employed to ascertain how the identified factors were associated with mortality at 30 days (1: dead; 0: alive). Backward elimination was used to identify the optimum model.

The Bayesian logit model based on prior evidences concerning the underlying binomial distribution (3.16) was built next. Expressed as proportions, the prior evidences were elicited from published data [233-234] with model (3.36-3.37). The decision makers specified the prior for between-study precision as Exponential[1] and that of the underlying proportion of 30-day mortality as Beta[1, 1]. The burn-ins and updates of the MCMC (Gibbs Sampler) were taken as 100 and 500, respectively. A

thorough search on the MEDLINE identified a relevant article [233]. This in turn helped to extract 8 most relevant and recent studies with detailed information on prognostic models concerning short-term mortality associated with ICH. In addition, the extracted article also reported the results of 122 ICH patients (aged 18 years and above) admitted in the authors' hospital between January 1988 and December 1997.

Conventional generalised additive model (GAM) [235] and two data mining models, namely Classification Tree (CART) [236] and Logistic Trees with Unbiased Selection (LOTUS) [237] were also applied. To validate the models externally, the original sample was randomly divided into two sub-samples (training set: 60%, test set: 40%). Receiver operating characteristic curves (ROC) [147-148] were generated to compare the models' predictive accuracy with observations in the test set. The models' areas under the curves (AUC) were compared [149]. Univariate analyses were performed with chi-square tests or Mann-Whitney tests [238].

Last but not least, the celebrated AdaBoost procedure [146] was proposed as a diagnostic check of the predictive performance of the candidate models. Analysed with Stata 9.0, R 2.4.1 and LOTUS 2.3, all statistical tests were conducted at 5% level of significance.

4.5.3 Comparison of Models

The sample characteristics and results based on univariate analyses are presented in Table 4.12. Old age, history of stroke, known atrial fibrillation, use of warfarin, glucose level, presenting Glasgow Coma Scale (GCS) [239] and pupil abnormality, post-resuscitation GCS and pupil abnormality, 1st International Normalised Ratio (INR) and PT results, vomiting, seizure, total volume of clot, ventrical extension and hydrocephalus were significantly associated with mortality.

Based on the best eye, verbal and motor responses, GCS is a reliable neurological scale for measuring the conscious state of a person (GCS≤8: severe coma, 9-12: moderate, ≥13: minor).

Conventional multivariate logit with backward elimination showed that only age, presenting GCS, 1st INR result and total volume of clot were significantly associated with 30-day mortality (Table 4.13). The model was found to be satisfactory by means of the Hosmer-Lemeshow test (p-value: 0.27) [240].

Based on selected published references (Table 4.14), the Bayesian logit model utilised a prior beta distribution, namely Beta[6, 8], for handling the unknown parameter of the underlying binomial distribution. The prior, generated from the published evidences based on model (3.36), was found to be reasonable according to the robust analysis. Figure 4.5 suggests that the Markov chain did not diverge. The close-form estimated ORs of the Bayesian logit model were slightly different from that of the conventional logit model and the 95% P.I.s were much tighter (Table 4.13).

The other candidate models did not generate ORs for analysis. Instead, they generated probabilities for predicting patients' mortality directly. For CART, the optimal tree was first obtained with a least CP criterion with 1 standard error, after pruning with 10 folds of cross validation. The same result was obtained should the deviance was used as the criterion. Presenting GCS (cut-off≤7.5) was identified as the only significant variable.

In terms of out-sample prediction (Figure 4.6), the AUC for Bayesian model was 0.83 (95% P.I.: 0.76—0.90). This was comparable to the conventional Logit (0.87; 95% C.I.: 0.81—0.94) and LOTUS (0.87; 95% C.I.: 0.81—0.92), but lower than GAM (0.92; 95% C.I.: 0.88—0.96). However, it performed significantly better

than CART (0.79; 95% C.I.: 0.73—0.85). In fact, the AUC of CART was significantly lower than that of the other 4 candidate models [149].

There was no obvious improvement in their predictability when AdaBoost was applied. The predictive performances based on AdaBoost were identical to their respective single models.

4.5.4 Decisions

This study was conducted in an attempt to supplement the published literature on performance of prognostic models for predicting 30-day mortality after ICH, a fatal condition faced by millions of patients world-wide. Unlike the relevant published references [233-234], this study employed several models based on conventional statistics, nonparametric statistics, data-mining and Bayesian statistics.

Neurologists should pay attention to ageing ICH patients with hydrocephalus, who presented with GSC≤8 (severe coma), high 1st INR result and high total volume of clot. Demmed to have a poor prognosis, these patients were at risk of death 30 days after suffering from ICH. For example, a 60 year-old patient with 1st INR=60, GSC≤8, and total volume of clot of 60 has a 33.3% chance of death within 30 days.

However, there was no clear-cut evidence that which candidate model was most superior. As a result, the analyst may choose any of the model (with the exception of CART) based on personal preference and ease of implementation. It is also not necessary to perform complicated analyses with the AdaBoost procedure as the single classifiers had attained their "best possible" predictive performance. However, it is worthwhile to recruit more patients to the on-going study and conduct another comparison in future.

Table 4.12: Sample characteristics of ICH patients

	30-Day Mortality Died (n=179)	Alive (n=455)
Age (years) *		
Mean (Range)	67.7 (17-95)	63.5 (18-109)
Gender		
Male	100 (27.0%)	271 (73.0%)
Female	79 (30.0%)	184 (60.0%)
History of stroke *		
Yes	59 (35.5%)	107 (64.5%)
No	119 (25.5%)	348 (74.5%)
Known hypertension		
Yes	125 (27.2%)	334 (72.8%)
No	54 (31.0%)	120 (69.0%)
Known atrial fibrillation *		
Yes	19 (42.2%)	26 (57.8%)
No	160 (27.2%)	429 (72.8%)
Use of warfarin *		
Yes	26 (57.8%)	19 (42.2%)
No	153 (26.0%)	436 (74.0%)
Artiplatelet		
Yes	31 (30.4%)	71 (69.6%)
No	136 (26.4%)	379 (73.6%)
Glucose level (mmoL) *		
Mean	9.0	8.5
Range	4.9-19.7	3.7-23.4
Presenting Glasgow Coma Scale (GCS) *		
≤8	103 (70.6%)	43 (29.4%)
>8	73 (15.2%)	407 (84.8%)
Post resuscitation GCS *		
≤8	99 (68.3%)	46 (31.7%)
>8	34 (9.5%)	322 (90.5%)
Presenting pupil abnormality *		
Yes	37 (88.1%)	5 (11.9%)
No	97 (20.1%)	386 (79.9%)
Post resuscitation pupil abnormality *		
Yes	42 (91.3%)	4 (8.7%)
No	82 (20.5%)	319 (79.5%)
1st PT result *		
Mean	17.0	14.0
Range	11.7-59.4	10.2-67.6
1st PT result		
Mean	32.3	30.1
Range	20.9-115.0	3.0-101.5
1st INR result *		
Mean	1.4	1.1
Range	0.9-4.9	0.6-13.0
Vomiting *		
Yes	47 (36.7%)	81 (63.3%)
No	132 (26.1%)	373 (73.9%)
Seizure *		
Yes	5 (27.8%)	13 (72.2%)
No	127 (28.1%)	325 (71.9%)
Total volume of clot *		
Mean	56.7	16.7
Range	0.3-224.5	0.0-176.9
Ventrical extension *		
Yes	111 (44.4%)	139 (55.6%)
No	61 (16.4%)	312 (83.6%)
Hydrocephalus *		
Yes	62 (45.3%)	75 (54.7%)
No	108 (22.3%)	376 (77.7%)

* Significant at 5% level of significance

Table 4.13: Comparison between Logit and Bayesian Logit

Covariates	Conventional Logit OR (95% C.I.)	Bayesian Logit OR (95% P.I.)
Age (years)	1.03 (1.01-1.05) *	1.00 (1.00—1.01) *
Presenting GCS		
≤8	Reference	Reference
>8	0.11 (0.06-0.22) *	0.92 (0.90—0.93) *
1st INR result	1.44 (1.09-1.90) *	1.01 (1.01—1.02) *
Total volume of clot	1.03 (1.02-1.04) *	1.00 (1.00—1.01) *
Hydrocephalus		
Yes	Reference	Reference
No	0.92 (0.48-1.74)	0.99 (0.98—1.02)

* Significant at 5% level of significance

Table 4.14: Selected studies for prior elicitation

Study	30-Day Mortality n/N
Mase 1995	38/138
Razzaq 1998	58/146
Tuhrim 1999	28/129
Phan 2000	29/100
Hemphill 2001	68/152
Nilsson 2002	124/341
Szczudlik 2002	59/152
Ariesen 2006	49/122

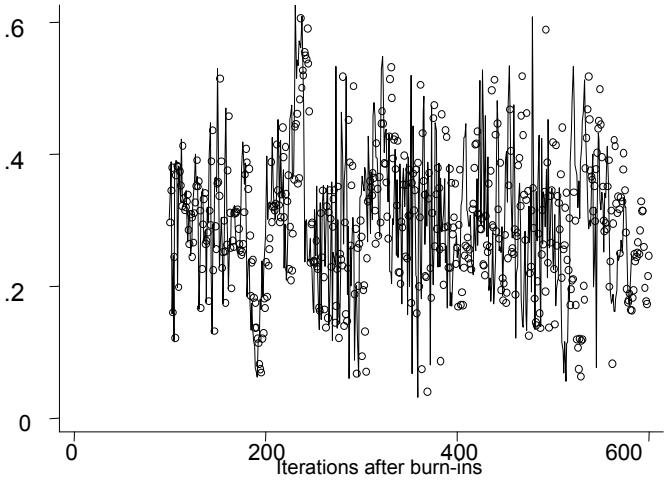


Figure 4.5: MCMC iterative history

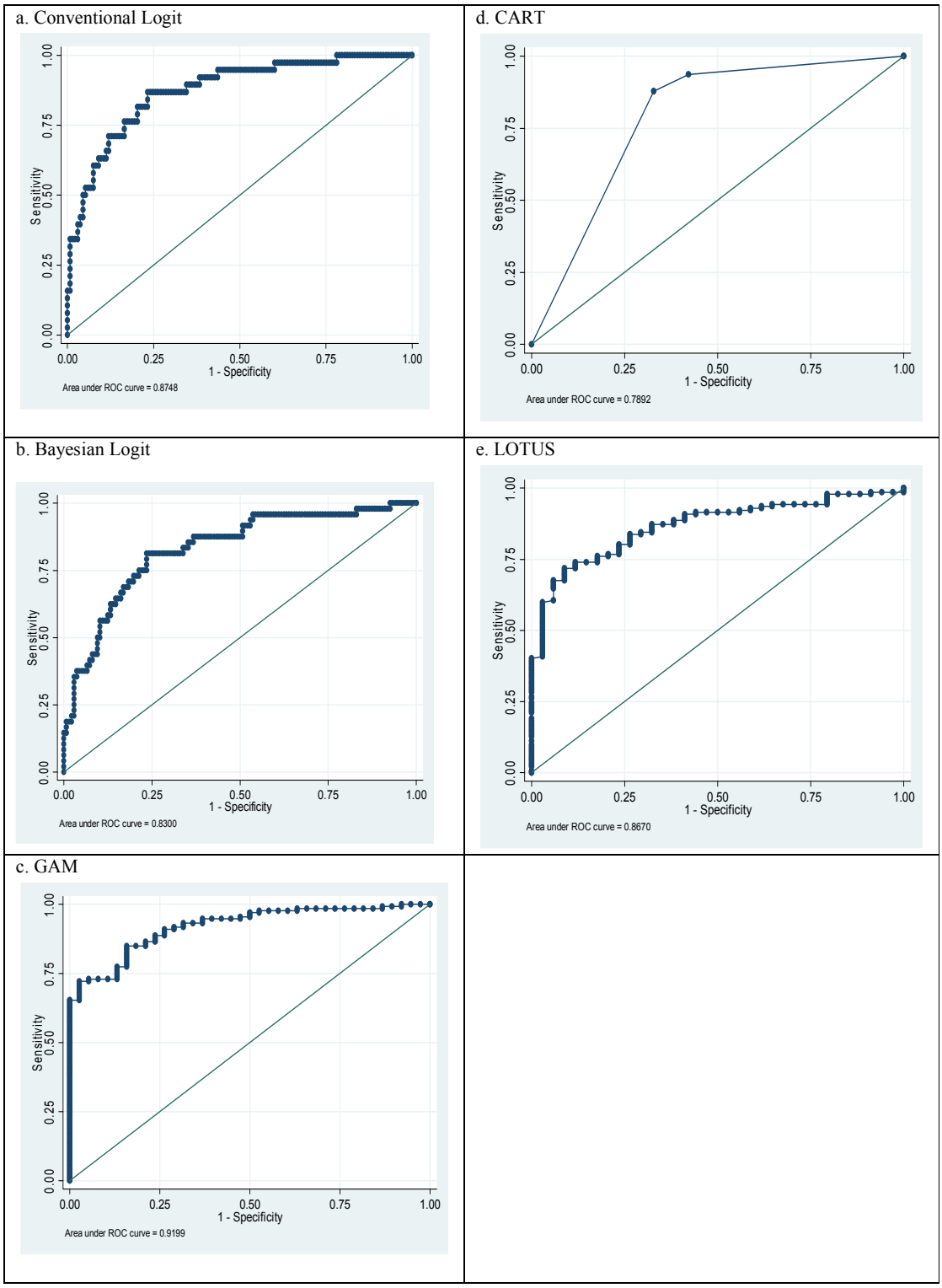


Figure 4.6: ROC curves for comparing candidate ICH models' out-sample predictive performance

4.6 Body Weight Reduction

4.6.1 Background

Associated with many disorders, obesity is defined as the accumulation of excessive body fat. The body mass index (BMI) is often used to define obesity. According to the latest guideline developed for Asian populations, a person carries a moderate health risk with a BMI≥23 kg/m^2, while a BMI≥27.5 kg/m^2 signifies high risk. This health risk refers to the risk of developing chronic health problems if weight is not controlled in the longer term. In Singapore, the proportion of population aged 18-69 years with BMI≥30 kg/m^2 was 6.4% in 2004. This was higher than the 6.0% reported in 1998.

Becoming common among urban populations in affluent societies, obesity is now considered by public health practitioners as a major world-wide health issue. "Obesity pandemic" is now used to describe this worrying trend. There are strong evidences suggesting that obesity is associated with many life-style diseases and disorders (back pain, sleep apnea, depression, high blood pressure, diabetes, heart failure, hyperlipidemia, stroke, menstrual disorders, skin disorders, osteoarthritis, gout, gallbladder disease and cancers of the ovaries, breasts and uterus, etc.), accelerated ageing and excess deaths [241]. Obesity is often caused by consumption of high-fat foods, physical inactivity, emotional disturbances, genetic and environmental factors, and the use of certain drugs (antidepressants, antipsychotics, and antihypertensives, etc.).

Accumulation of excessive body fats changes overall physical appearances. Obese people may also find their physical or social activities restrcited because of fatigue, depression, lack of mobility and other complications. This is a major concern among persons who are conscious about their physical attractiveness. The common

advice for weight reduction usually concerns lifestyle changes, and these include regular exercise and reduced intake of calories. For those who are severely obese (BMI>40 kg/m^2), clinicians may recommend surgery as the choice of treatment. For persons who are moderately obese, weight-loss drugs and medicinal herbs that aim at increasing metabolism and/or to reduce appetite may be prescribed.

Although usually harmless, the common weight-reducing drugs available in the market are reported to be ineffective. Clinicians usually recommend weight-management programme as part of the treatment plan. However, development of new drugs, especially those with natural contents, presents a lucrative business for the pharmaceutical industry.

In an attempt to fulfill the objective of this study, the effectiveness of Xändo, a widely-publicised drug, was evaluated. Its advertising strategy included the use of popular celebrity spokespersons. Promised to be a weight-reducing agent of natural origin, Xändo comprises a number of ingredients (α-amylase inhibitor, inulin and Garcinia cambogia) that work together to capture excess sugars, reduce their storage as body fat and eliminate them from the body by excretion [242]. However, the statistical analysis presented in the original article was incorrect. As a result, one is not sure if the developer's claim such as "3× more weight loss" was valid. This will be dealt with in the relevant section.

4.6.2 Aim

The primary aim of this study was to assist a male obese teenager to decide how to reduce his body weight within 3-4 months. This is a valid concern among teenagers because they are likely to remain obese when they grow up. Young men with morbid obesity have a 10-fold excess mortality compared with their normal

weight counterparts. As mentioned before, this study also helped to re-analyse the presented data of Xändo [242] with an appropriate Bayesian model. Next, the study made use of data provided by the Singapore Armed Forces (SAF) in order to ascertain whether a prolonged basic military training (BMT) for the decision maker would lead to reduction in BMI and increase in injury rate.

4.6.3 Decision Problem and Data

The decision maker of this study was a Chinese, pre-university obese teenager (BMI: 35.5 kg/m^2) who was about to be enlisted. At the point when the decision problem was surfaced he had only about 15 weeks to reduce his excessive body weight. Initially, he was presented with two choices—consume Xändo tablets regularly or embark on a weight-management programme. The first choice was a relatively comfortable decision as he only needed to adhere to the developer's guidelines. Table 4.15 depicts the published results of the original article on Xändo [242]. With 40 healthy obese volunteers (Xändo: 20, placebo: 20) enrolled, the study was concluded in 12 weeks (end-point). With Xändo, the decision maker did not have to make substantive changes to his current lifestyle, but was uncertain of the effectiveness of the new drug. Moreover, there might be complications and side-effects, although the developer promised there were none [242]. The second choice could turn out to be less tolerable as he might have to give up many of his hobbies. The clinicians also estimated that there was only a slim chance (0.05) to reduce his BMI to the targeted 27 kg/m^2 in 12-15 weeks, given his unique condition.

Table 4.15: Selected original published results of Xändo

	Week 0 (Baseline) Mean BMI±s.d.	Week 12 (End-Point) BMI±s.d.	Change
Xändo (n: 20)	31.0 ± 3.2	29.7 ± 3.2	1.3 ± 1.2
Placebo (n: 20)	31.7 ± 2.9	31.2 ± 2.7	0.5 ± 0.9

If he failed to reduce his BMI to below 27 kg/m^2 he would be required to undergo the compulsory strenuous military training designed for obese servicemen. He would then face a prolonged service that could delay his discharge, or more seriously an enhanced likelihood of suffering from physical injuries. Obesity is believed to pose additional risks during strenuous military training.

In the past, pre-enlistees with BMI≥27.0 kg/m^2 underwent a 16-week basic military training (BMT). The severe and extreme obese enlistees underwent similar training with the mildly and moderately obese recruits. The BMT was then prolonged to 26 weeks for the severely obese enlistees in 2004. The burning question was if an increase in duration of training would result in improved weight loss profiles, reduced injury rates and improved fitness levels. After BMT, the recruits were given a PES grading based on their BMI and/or body fat percentage according to the SAF definition of obesity. This PES grading would in turn determine their deployability into various military vocations. A reduction in BMI had to be significant enough, medically speaking, to result in a change in PES grading. The retrospective data of 716 recruits with BMI≥35 were analysed (16-week BMT: 338; 26-week BMT: 378).

Bayesian linear regression model (3.44) was used to analyse the two data sets described above in order to encode the required probabilities for the decision problem. With non-informative priors, the effectiveness of Xändo was evaluated. While the chemical effects of inulin and hydroxycitric acid (HCA) are fairly well-known, there was insufficient evidence of their effectiveness in human bodies. With this in mind, a diffuse prior for the covariance matrix was used, i.e., Σ:=diagonal$_{3\times3}$(1,000). The prior for coefficients was taken to be $\boldsymbol{\beta}^{\mathbf{0}}$=[0.8, -2.0, 2.0]'. This prior suggested that Xändo might be effective when compared with placebo. The average active group was expected to be 2.0 kg/m^2 lower than that of the placebo group after 12 weeks.

The baseline and end-point BMI were expected to be correlated at 0.8. This is reasonable because the study's time frame was relatively short.

The obese recruits' end-point BMI between training groups, after correcting for baseline BMI, were also analysed with informative priors. In a well-cited study involving 27 severely obese subjects (average baseline BMI: 44 kg/m^2) who engaged in an intensive lifestyle intervention composing of physical activity, dietary changes and personal development, the 15-week result was significant [243]. Based on the evidence, the average end-point BMI was expected to reduce by 1.0 kg/m^2. It was also expected that the baseline and end-point BMI (26 weeks) were correlated at 0.8. As such, the priors were taken as $\boldsymbol{\Sigma_0}$:= $\text{diagonal}_{3\times3}(10)$ and $\boldsymbol{\beta_0}$=[0.8, -0.1, 2.0]'.

Last but not least, Bayesian logit model (3.16) with informative prior, i.e., Beta[5, 5] was applied to analyse the occurrence of injuries (1: yes, 0: no). The relevant priors are provided by a SAF medical officer specialised in weight-reduction programmes. Analysed with Stata 9.0 and Microsoft Excel 2002, all statistical tests were conducted at 5% level of significance. Post-hoc analysis suggested that the sample size provided by SAF was sufficient for all relevant statistical testing at 90% power and 5% level of significance, after taking into consideration the potential correlation between baseline and end-point BMI of the enlistees.

4.6.4 Results

The demographic characteristics of the recruits in the two training groups were comparable (See Table 4.16). There was no significant difference in average baseline BMI between the two groups (p-value: 0.10). Both groups demonstrated significant decline in average BMI at end-point (16-week BMT: 39.0 $kg/m^2 \rightarrow$34.3 kg/m^2; p-

value<0.001; 26-week BMT: 39.5 kg/m^2→34.4 kg/m^2; p-value<0.001). The baseline BMI was normally distributed but became slightly skewed at end-point.

The decline in average BMI of the 26-week group was significantly larger than that of the 16-week group, after adjusting for ethnicity and educational attainment (Table 4.17). The non-informative Bayesian model was found to be satisfactory.

Although there were more injured cases (9.3%) reported for the 26-week training group when compared with the 16-week group (8.0%), the difference did not turn out to be significant, after adjusting for ethnicity, educational attainment and baseline BMI (Table 4.18). The model made use of an informative prior Beta[5, 5] given by the analyst. There were significantly fewer injured cases among the Malay recruits (5.3%), when compared with their Chinese (9.9%) and Indian (17.8%) counterparts. On the other hand, there were more injured cases reported for recruits with primary and lower education (40.0%), when compared with those holding higher qualifications (secondary: 8.7%, pre-university: 8.9%, tertiary qualifications: 7.3%).

The analysis of the Xändo called for some re-construction based on the published aggregate evidence (Table 4.15) as no subject-level data was available. The analysis provided by the article was wrong because it failed to compare the groups' end-point BMI directly. Instead it analysed the groups' end-point BMI separately. Statistically speaking, this is unacceptable. A proper analysis should include the baseline (week 0) BMI as it is expected to be associated with the end-point (week 12) BMI. The proposed Bayesian technique required the analyst to compute the conventional linear regression model's coefficients first. Usually the task is impossible without subject-level data but fortunately the original article [242] threw a lifeline by providing the standard deviations of change in end-point BMI for both

groups. The most important element is the correlation between baseline (week 0) BMI and end-point (week 12) BMI as it is required to compute all the necessary quantities (intermediate and final) and coefficients for the regression analysis. The technique works as follows:

i. generate the pooled variance of baseline BMI and end-point BMI of the two groups (0: placebo, 1: Xändo)

ii. compute the pooled coefficient of correlation between the baseline BMI and end-point BMI by making use the fact that $V[BMI_1 - BMI_0] = V[BMI_1] + V[BMI_0] - 2(\text{correlation})\sqrt{V[BMI_1]V[BMI_0]}$

iii. work out all the cross products required for regression analysis based on ii and the known features and quantities of the study

iv. compute the regression coefficients based on ii and iii

v. generate the coefficients' standard errors, sums of squares (total, model and residual), p-values, 95% C.I.s, F-value and adjusted coefficient of determination (R^2)

The re-construction of the regression analysis was based on the known inter-relationships of all the relevant quantities. The quantities generated are based on exact methods [34].

The results are shown in Table 4.19. It is interesting to note that the developer's claim was fairly valid, i.e., there was significant change in subjects who took Xändo, when compared with those on placebo. The end-point BMI of the Xändo group was about 0.86 kg/m^2 significantly lower than the placebo group. The conventional model provided a good fit to the data as the covariates were able to explain about 90% of variations in end-point BMI.

Table 4.16: Sample characteristics of SAF recruits

	16-week BMT n: 338	26-week BMT n: 378
Ethnicity:		
Chinese	187 (55.3%)	206 (54.5%)
Malay	122 (36.1%)	144 (38.1%)
Indian	20 (5.9%)	25 (6.6%)
Others	9 (2.7%)	3 (0.8%)
Educational attainment:		
Primary & below	4 (1.2%)	6 (1.6%)
Secondary	161 (47.6%)	207 (54.8%)
Pre-University	45 (13.3%)	45 (11.9%)
Tertiary	128 (37.9%)	120 (31.7%)
Occurrence of injuries:		
No	311 (92.0%)	343 (90.7%)
Yes	27 (8.0%)	35 (9.3%)
Baseline BMI (kg/m^2)	39.0 (s.d.: 4.0) Range: 32.7—58.8	39.5 (s.d.: 4.1) Range: 29.8—57.4
End-Point BMI (kg/m^2) ✞	34.3 (s.d.: 3.6) Range: 27.5—52.5	34.4 (s.d.: 4.0) Range: 24.8—49.6

✞ Based on 654 injury-free cases: 311 (16-week) and 343 (26-week)

Table 4.17: Informative Bayesian linear regression analysis of recruits' end-point BMI

Covariates	Coefficient	95% P.I.
Baseline BMI (kg/m^2) *	0.79	0.83 — 0.90
Group *:		
1: 16-week BMT	Reference	—
2: 26-week BMT	-0.49	-0.78 — -0.26
Constant	4.21	-1.47 — 2.29

* Statistically significant at 5%

Table 4.18: Informative Bayesian logit analysis of occurrence of injury (1: yes, 0: no)

Covariates	Coefficient	95% P.I.
Baseline BMI (kg/m^2)	1.74×10^{-3}	-1.03×10^{-4} — 3.59×10^{-3}
Ethnicity:		
Chinese	Reference	—
Malay *	-0.02	-0.03 — -1.11×10^{-3}
Indian	0.03	-4.83×10^{-3} — 0.06
Others	-2.22×10^{-3}	-0.06 — 0.06
Educational attainment *:		
Primary & below	Reference	—
Secondary	-0.11	-0.17 — -0.04
Pre-University	-0.11	-0.18 — -0.04
Tertiary	-0.11	-0.18 — -0.05
Group:		
1: 16-week BMT	Reference	—
2: 26-week BMT	3.05×10^{-3}	-0.01 — 0.02
Intercept *	-0.11	-0.21 — -0.01

* Statistically significant at 5%

Table 4.19: Reconstruction of conventional linear regression analysis of end-point BMI with/without Xändo

Source	Sum of squares	Degrees of freedom	Mean sum of squares	n	40
Model	308.48	2	154.24	R^2	0.90
Residual	33.35	37	0.90	Adjusted R^2	0.90
Total	341.84	39	8.77	F (p-value)	171.11 (<0.01)

Covariates	Coefficients	Standard error	t-value	p-value	95% C.I.
Baseline BMI *	0.89	0.05	17.81	<0.01	0.79—0.99
Group *					
0: Placebo	Reference	—	—	—	—
1: Xändo	-0.88	0.30	-2.89	<0.01	-1.49 — -0.26
Intercept	2.90	0.03	108.76	<0.01	2.84 — 2.95

* Statistically significant at 5%

With the conventional model's coefficients generated the analyst was able to compute the Bayesian coefficients (3.44) since the latter are "shrunk" estimates of the former and the priors (Table 4.20). Again, Xändo was shown to be effective in reducing subjects' body weight after 12 weeks. In passing, note that the results were almost identical to that of the conventional model because the data were fairly consistent.

4.6.5 Decision

With the probabilities generated from the Bayesian models (Tables 4.17-4.18, 4.20), the tree diagram was completed (Figure 4.7). The encoded probabilities illustrated on the chance nodes are bracketed. The most preferred and least preferred scenarios were meeting target with Xändo (BMI<27 kg/m^2) in 12 weeks and picking up an injury with target unmet (BMI≥27 kg/m^2), respectively. Complication due to Xändo was not featured because the published article [242] reported no side-effects observed in the obese subjects in 12 weeks.

Based on some close discussion with the decision maker the following exponential utility function was derived:

$$u(x) = 1.78(1 - e^{-x/12.16}) \tag{4.1}$$

The procedure is documented in subsection 3.5.5. With his utilities elicited, it is obvious that he might benefit from choosing Xändo to reduce his body weight. The expected utility derived from using Xändo was 0.46. This was higher than that derived from participating in the weight-management programme. However, it was a rather formidable task to achieve his target (BMI<27 kg/m^2) within 12-15 weeks,

given his initial BMI of 35.5 kg/m^2. The probability of achieving the target was only 0.12 with Xändo. Perhaps he should have started his weight-reduction programme earlier. With his height at 1.80 m, reducing a BMI by 8.5 kg/m^2 was equivalent to reducing his body weight by 27.5 kg.

4.6.6 Discussion

Xändo seemed to have fulfilled its promise of helping obese consumers to attain significant weight loss after 12 weeks, although the original analysis presented in the article [242] was conceptually and technically wrong. However, nothing was said about its long-term effects. The problem of weight regain after termination of a therapy is a well documented issue. On average, most patients regain about 30% to 35% of their lost weight one year after on treatment. Published reports also suggested that 50% or more of participants would return to their baseline weight approximately 3 to 5 years after therapy [244-246]. One pair-matched study involving 24 subjects reported that they regained 11 kg 4 years after losing 10 kg [247].

Future studies may consider other weight-management programmes and more established weight-reducing drugs. The US Food and Drug Administration (FDA), as well as an expert panel convened by the National Heart Lung and Blood Institute, have recommended that weight-reducing medications be used only as an adjunct to a comprehensive program of lifestyle modification that includes diet, physical activity, and behavior therapy. Two medications, sibutramine and orlistat, are currently approved by the FDA for long-term use in obesity management. Sibutramine is a CNS agent that inhibits the reuptake of norepinephrine and serotonin. By contrast, orlistat is a gastric and pancreatic lipase inhibitor that works by blocking the

absorption of fats contained in a meal, and with the undigested fat excreted. Both sibutramine and orlistat have been used successfully in the induction of weight loss.

One limitation of the analysis based on SAF data was that the study was retrospective in nature. It did not allow for randomisation to ensure uniformity in the 2 study groups. This was, however, partially compensated with the use of multivariate statistical models, which adjusted for ethnicity, education attainment and baseline BMI. The 26-week BMT did prove to produce a more significant decrease in BMI, when compared with the 16-week group. This could be attributed to the additional 10 weeks of physical training in a controlled environment. However, as there was minimal change in training programme or diet between the 2 groups, the difference merely reflected weight loss as a result of overall increase in amount of exercise. It may be prudent to design a specific BMT training programme for the 26 weeks that addresses intensity, graduated increments as well as dietary alterations to fully exploit the benefits of the additional training.

While the chosen decision may seem straight-forward and obvious in view of the simple structure of the problem, it does illustrate the usefulness of the incorporation of various Bayesian models for probability encoding.

Table 4.20: Informative Bayesian linear regression analysis of recruits' end-point BMI with/without Xändo

Covariates	Coefficient	95% P.I.
Baseline BMI (kg/m^2) *	0.91	0.82—1.00
Group *:		
0: Placebo	Reference	—
1: Xändo	-0.86	-1.42 — -0.30
Intercept	2.36	-0.61 — 5.34

* Statistically significant at 5%

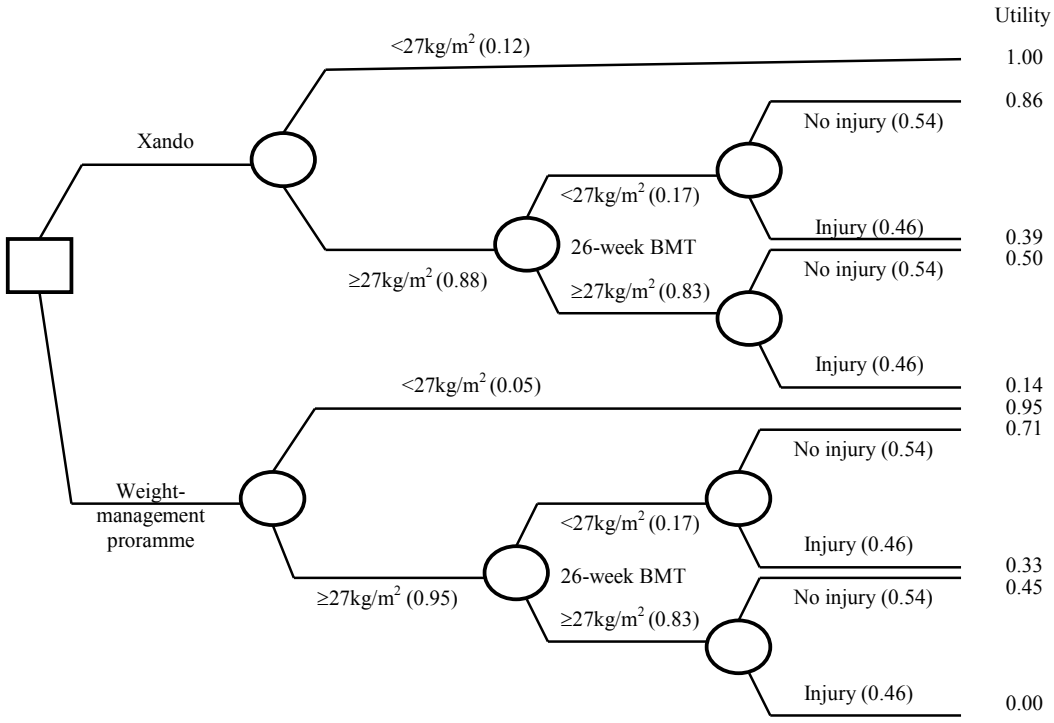


Figure 4.7: Xändo versus weight-management programme for weight-reduction

4.7 ACE Inhibitor for treating Ischaemic Heart Disease

4.7.1 Background

Angiotensin Converting Enzyme (ACE) inhibitors are a group of pharmaceuticals used primarily for controlling blood pressure and treating congestive heart failure. They help to slow down the enzyme activities, which in turn halt the production of angiotensin II, a potent chemical responsible for causing high blood pressure (hypertension). Also useful for preventing kidney damage in patients with hypertension or diabetes mellitus (DM), ACE inhibitors have become an important class of drugs for preventing death resulting from heart failure or heart attack. In addition, it is able to reduce the progress of diabetic nephropathy independent from their blood-pressure lowering effect.

The highly-regarded European Trial on reduction of cardiac event with Perindopril (EUROPA) demonstrated that ACE inhibitors significantly improved patients' prognosis, with or without hypertension and diabetes, and irrespective of age [248]. Involving patients from 424 centres across Europe, the trial was randomised, double-blinded, and placebo-controlled.

There are variations among the ACE inhibitors, such as salfhydryl-containing (capoten), dicarboxylate-containing (perindopril) and phosphate-containing (fosinopril). The difference lies in how they are eliminated from the body. Some inhibitors need to be converted into an active form in the body before they function. Others may inhibit ACE directly in the tissues rather than that present in the blood. However, there is no reported evidence on the relative effectiveness of these different inhibitors.

ACE inhibitors have few interactions with other drugs, but it is advised that they should not be taken with potassium supplements, salt substitutes and other drugs

that may increase the body's potassium levels. Aspirin and other non-steroid anti-inflammatory drugs (NSAIDs) may also weaken the effects of ACE inhibitors. While relatively well-tolerated by most patients, therapy with ACE inhibitors requires careful monitoring as it may cause dizziness in patients who are overdosed or less tolerated with rapid reduction in blood pressure. Other common adverse effects include hyperkalemia, headache, fatigue, hypotension and renal impairment in patients with renal artery stenosis. Some patients may also develop angioedema due to increased bradykinin levels.

4.7.2 Aim

The following decision analysis concerned an anonymous 45-year old Chinese male professional suffering from ischaemic heart disease (IHD) and Type-II diabetes. Unlike Type-I diabetes, which is caused by autoimmune disorders, Type-II diabetes is usually developed in adults above 40 years of age. Although there are exceptions, Type-II patients are usually overweight. The existence of Type-II diabetes has been shown to have brought adverse impact on patients with congestive heart failure (CHF). CHF, also known as congestive cardiac failure (CCF) is a condition in which the heart fails to pump enough blood to the body's other organs. This can be caused by narrow arteries, past myocardial infarction, high blood pressure, heart valve disease, genetic family history, and heart defects present at birth. IHD is a specific heart disease characterised by reduced blood supply to the heart. Typically, there is blockage to the coronary arteries which reduces the blood supply to heart muscles. IHD patients may experience sudden heart attack, which results in long-term damage to heart muscle and structural damage to the organ. There are evidences showing that diabetes was

associated with ICH and eventual heart failure mortality [249-251]. Suffering from diabetes for 5 years, the patient was not on insulin.

The proposed analysis could help the decision maker to decide whether he should consider ACE inhibitors, while taking into consideration its relative effectiveness with other drugs such as beta-blockers and spironolactone. The end-points were hospital re-admission and survival.

Over the past few decades, preventive and therapeutic measures have substantially improved the prognosis of IHD patients. Nevertheless, the risk of cardiovascular complications remains high and progression can be halted in few patients despite regular treatment with established drugs like beta-blockers, aspirin and statins. The analysis also helps to show whether ACE inhibitors could significantly reduce IHD patients' risk of hospital re-admission and mortality, a topic of paramount interest to cardiologists.

4.7.3 Data

The data were obtained from 1,668 patients who were enrolled to the National Healthcare Group (NHG) Multidisciplinary Heart Failure Disease Management Programme, Republic of Singapore, from October 2003 to September 2006. There were definite evidence of CHF in these subjects, on the basis of clinical findings and/or the Boston Criteria, with documented LV systolic dysfunction. The patients had LV ejection fraction (EF) below 40%. However, patients with advanced malignancy, severe renal failure, severe pulmonary disease, psychiatric or cognitive disorders, and who were on cardiovascular interventional procedures were excluded.

All patients enrolled to the programme were educated on dietary and fluid management, and were followed-up by CHF specialists at regular intervals. They

were advised to exercise regularly, quit smoking, and reduce excess body weight and cholesterol. In addition, they were also placed on telephonic case management by CHF-trained nurses, and contacts were made on a monthly basis.

This was essentially a prospective cohort study, where patients were followed-up for 24 to 48 months (mean duration: 36 months). On entry to the programme, the patients' baseline demographic and clinical characteristics were recorded. Other information included quality of life (Minnesota Living with Heart Failure score), medication use and physical functionality (New York Heart Association Classification). Information on hospitalisation for any cause at the emergency departments was collected prospectively by checking the in-hospital admission list and the attendance charts on a daily basis. Evidence of death was collected from the Singapore Registry of Births and Deaths. The primary end-points of interest were hospital re-admission due to heart failure and all-cause mortality.

Collated and managed by a full-time executive, the data were further validated by two consultant cardiologists and a principal medical statistician during analysis. To ensure that the analyses were reliable, the investigators met regularly to fill up the missing data.

The proposed decision analysis involved 1 decision node (use of ACE inhibitors vs. other drugs), 4 deterministic nodes (gender, ethnicity, age and kidney damage), 2 chance nodes (hospital re-admission and survival) and 1 value node (patient's utilities). The decision node concerned comparison between ACE inhibitors and other drugs such as beta-blockers and spironolactone. Beta-blockers, which work on the heart and circulatory system, has been downgraded as 4^{th}-line treatment by the United Kingdom in 2006 as there was reported evidence that frequent users at usual dose could carry an unacceptable risk of provoking Type-II diabetes [252]. However,

it is associated with clinically meaningful reductions in mortality [253]. Both ACE inhibitors and beta-blockers have shown to be helpful in reducing mortality in patients with or without renal insufficiency [254]. Suspected to be associated with stomach bleeding, spironolactone is a synthetic 17-lactone steroid used primarily for treating liver disease, low-renin hypertension and hypokalemia.

Glomerular filtration rate (GFR), based on plasma creatinine, is a measure of patients' kidney function. Diabetes and high blood pressure are among the most notable risk factors for kidney disease. Patients are deemed to be suffering from kidney disease if his GFR falls below 90.

Three demographic variables, namely gender, age and ethnicity, were included as either risk factors or confounders in the proposed analysis. Ageing is expected to affect the patient's prognosis. Gender is also suspected of associating with congestive heart failure mortality [251]. Generally, females have a longer life span than males. There is also evidence showing ethnic differences in acute myocardial infarction events in Singapore, with Malays having the highest case-fatality [255].

Last but not least, the patients' functionality was quantified by the New York Heart Association (NYHA) Classification, a well-regarded functional and therapeutic classification for prescription of physical activity for cardiac patients. A patient with no limitations of activities is graded as Class I. At Class II, the patient has slight limitation of activities, but is comfortable with rest or mild exertion. At Classes III and IV, however, the patient concerned is either limited with daily activities or require complete rest (confined to bed or wheel-chair). Since the decision maker was not functionally restricted, the proposed analysis excluded all data from patients with NHYA>2.

Analysed with Bayesian logistic regression (3.16) and Weibull survival model (3.50), all statistical tests were conducted at 5% level of significance. The priors for the effect of ACE inhibitors and beta-blockers in the Weibull model were taken as 0.7 and 0.8, respectively. These are interpreted as hazard ratios (HR), with values below unity as beneficial effects on survival. The associated variance-covariance matrix was chosen to be $\boldsymbol{\Sigma}_0$:=diagonal$_{3\times3}$(10), thus reflecting a conservative stand. There were either dubious or short of direct evidence concerning the effects of the drugs in published references [251, 256-259]. In reference [256] the investigators failed to recognise the multicollinearity effect revealed in the analysis and reported an adverse impact of ACE inhibitors on mortality (HR=1.53, 95% C.I.: 0.86—2.75). In fact, the investigators wrongly expressed the HR as odds ratio (OR). It is suspected that ACE inhibitors' beneficial effect was "masked" by that of beta-blockers (HR=0.77, 95% C.I.: 0.54—1.09), whose effect was "expected" but non-significant. This is a common problem encountered in real-life data analysis and investigators ought to pay special attention to unexpected signs. One prospective study reported that survival rates after 1, 3 and 5 years, as determined by the Kaplan-Meier curves, were found to be around 78.9%, 57.2% and 39.0% [257]. Another reference presented figures on 6-month mortality only [259]. There was also no direct and confirmative result concerning hospital readmission due to heart failure [249]. Estimated to be 40%, the logit analysis of hospital readmission used a prior Beta[2, 3]. Though restricted to patients enrolled at NHG, this was the largest hospital-based CHF cohort study in Asia.

4.7.4 Results

The patients' profiles are depicted in Table 4.21. Altogether, 411 NYHA≤2 patients with IHD and diabetes were included for analysis. No patient was prescribed

with insulin. Since there were relatively few patients in the "other" ethnic group (2.2%), they were discarded from all analyses below.

Table 4.22 shows that the use of beta-blockers was beneficial in reducing hospital readmission due to heart failure. However, the effect of ACE inhibitors was dubious. Careful analysis showed that there was indeed no significant result observed. The non-significant result was not caused by multicollinearity among the predictors. Demographics (ethnicity, gender and age) and occurrence of kidney disease) were not significantly associated with hospital re-admission.

On the other hand, Table 4.23 suggests that there were some protective effects against death by the use of ACE inhibitors, beta-blockers and other drugs, according to the Weibull Bayesian survival analysis. However, their effects were non-significant in explaining the time to death. Their individual effects on mortality might be more directly explained by hospital re-admission. Sensitivity analysis with several informative priors did not change the result drastically (details not shown).

The patients' probability of death was well below 25% throughout the 4 years (approximately 1,500 days) of follow-up (Figure 4.8). The 6-month (180 days) mortality was similar to published results in reference [259]. The prognosis was substantially better than most of the reported results in references. This is not surprising because the local sample excluded patients who were physically dependent (NHYA>2).

Table 4.21: Sample characteristics of diabetic IHD patients

	n (%)
Ethnicity:	
Chinese	257 (62.5%)
Malay	81 (19.7%)
Indian	64 (15.6%)
Others	9 (2.2%)
Gender:	
Male	269 (65.5%)
Female	142 (34.5%)
Age:	
Mean (range)	65.8 (30-94)
Use of ACE inhibitors	
No	171 (41.6%)
Yes	240 (58.4%)
Use of beta-blockers and/or spironolactone	
No	327 (79.6%)
Yes	84 (20.4%)
GFR	
Normal (≥90)	171 (41.6%)
Kidney damage (15-89)	235 (57.2%)
Kidney failure (<15)	5 (1.2%)
Hospital re-admission due to heart failure	
No	376 (91.5%)
Yes	35 (8.5%)
All-cause death within 4 years	
No	337 (82.0%)
Yes	74 (18.0%)
≤6 months	28 (6.8%)
6 months to 2 years	44 (10.7%)
>2 years	2 (0.5%)
Total	411 (100.0%)

Table 4.22: Informative Bayesian logit analysis of hospital re-admission in diabetic ICD patients (NYHA≤2)

Covariates	Coefficient	95% P.I.
Ethnicity:		
Chinese	Reference	—
Malay	-4.54×10^{-3}	-0.07 — 0.06
Indian	-0.02	-0.09 — 0.05
Gender:		
Male	Reference	—
Female	-0.02	-0.07 — 0.04
Age	2.33×10^{-3}	-3.69×10^{-4} — 5.04×10^{-3}
Use of ACE inhibitors		
No	Reference	—
Yes	0.02	-0.05 — 0.09
Use of beta-blockers and/or spironolactone *		
No	Reference	—
Yes	-0.10	-0.17 — -0.03
GFR		
Normal (≥90)	Reference	—
Kidney damage (<90)	0.05	-0.02 — 0.12
Intercept *	-0.77	-0.95 — -0.58

* Statistically significant at 5%

Table 4.23: Non-informative Weibull Bayesian survival analysis of ICD patients with diabetes (NYHA≤2)

Covariates	Coefficient (hazard ratio)	95% P.I.
Ethnicity:		
Chinese	Reference	—
Malay	0.98	0.52—1.83
Indian	0.81	0.41—1.60
Gender:		
Male	Reference	—
Female	0.89	0.54—1.45
Age	1.02	0.99—1.04
Use of ACE inhibitors		
No	Reference	—
Yes	0.76	0.39—1.48
Use of beta-blockers and/or spironolactone		
No	Reference	—
Yes	0.86	0.39—1.91
GFR		
Normal (≥90)	Reference	—
Kidney damage (<90)	1.51	0.78—2.95
Hospital readmission *		
No	Reference	—
Yes	2.95	1.60—5.42

* Statistically significant at 5%

4.7.5 Decisions

Based on a close discussion with the patient the following exponential utility function was derived (see subsection 3.5.5 for details):

$$u(x) = 1.30(1 - e^{-x/67.70}) \tag{4.2}$$

His most preferred situation was no relapse and stayed alive. The worst scenario was readmitted to hospital and died within 6 months.

The Bayesian logit model (Table 4.22) suggests that his probability of re-admission (due to heart failure) after being treated with ACE inhibitors and beta-blockers were 0.12 and 0.11, respectively. There was little practical difference between the two drugs in terms of re-admission.

Moreover, if he did not re-admit to hospital his probability of survival was 0.86 with ACE inhibitors (Table 4.23). His chance of death was 0.05 within 6 months and 0.09 beyond 6 months. On the other hand, his chance of survival with beta-blockers and without readmission was slightly higher at 0.9. The probability of death was 0.05 for both within 6 months and beyond 6 months.

Based on Figure 4.9, there was no significant difference in utilities in choosing ACE inhibitors or beta-blockers. However, beta-blockers seemed to be a more effective drug for reducing hospital re-admission due to heart failure for diabetic ICH patients.

4.7.6 Future Study

Future analysis may include glycemic control (indexed by HbA1c). A recent study suggested that poor glycemic control might be associated with increased risk of

cardiovascular events (hospital readmission and death) and new onset heart failure in patients with diabetes [260]. The impact of HbA1c on prognosis in patients with established systolic heart failure has not been previously investigated. This presents a new direction for clinical research on diabetic ICH.

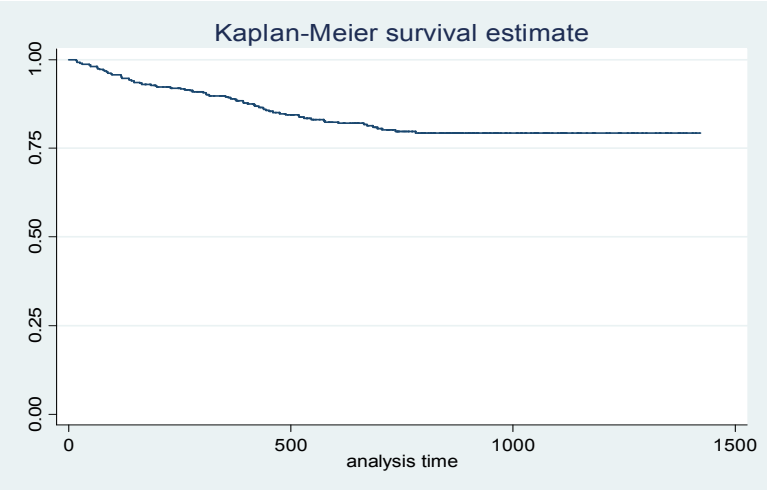


Fig 4.8: Kaplan-Meier analysis of ICH patients with diabetes (NYHA≤2)

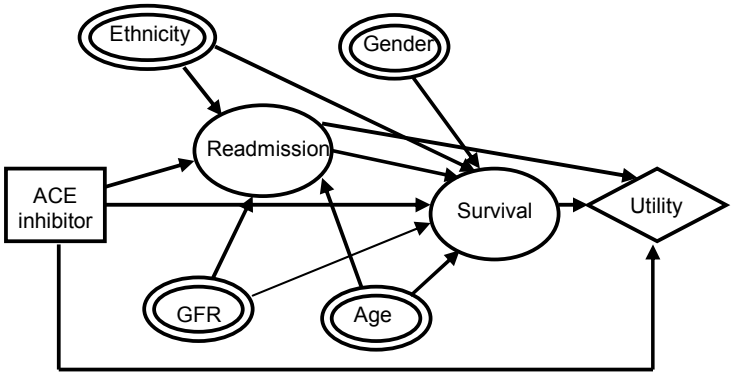


Figure 4.9: Re-admission and Mortality with IHD

4.8 Peritoneal Dialysis for treating End-Stage Renal Disease

4.8.1 Background

Kidney failure is the inability of the kidneys to filter metabolic wastes (creatinine and blood urea nitrogen) from the blood and regulate the salt/water content of the body adequately. Common causes include diabetes mellitus, high blood pressure, autoimmune disorders and other abnormalities (such as polycystic kidney disease and glomerulonephritis).

The incidence and prevalence of end-stage renal disease (ESRD) are expected to grow throughout the world [261]. There are two forms of treatment for ESRD: dialysis and transplantation. In the latter, a kidney from a living or brain-dead donor (cadaveric) is removed and implanted to the patient. Known as the best form of treatment, the average waiting time for receiving a kidney transplant from a cadaveric donor is 7 years or more in Singapore. It is understood that the transplantation rate is influenced by socio-economic, religious, and cultural attitudes [261]. As such, the more common treatment is dialysis, a process by which the patient's blood is cleansed artificially so that metabolic wastes and excess fluids are removed from the body. Nephrologists recommend dialysis when the patient's kidney failure is causing abnormal brain function, inflammation of the sac around the heart, high level of acid and potassium in the blood, and total body fluid overloaded, etc.

There are two types of dialysis, namely haemodialysis (HD) and peritoneal dialysis (PD). In the former, blood is removed from the body and pumped by a machine outside the body into a dialyser, which helps to filter metabolic wastes and then returns the purified blood to the patient. Complications of HD include fever, infection, low blood pressure, abnormal heart rhythms, bleeding in the intestine and life-threatening allergic reactions. In PD, a membrane that lines the abdomen and

covers the abdominal organs acts as the filter. It creates a space within the abdomen, where fluid and waste products are drained. Complications of PD include inflammation of the abdominal cavity (peritonitis), bleeding, leakage of fluid, low level of albumin, constipation and hernias of the abdomen and groin.

Compared to HD, PD is a more convenient dialysis as it can be performed at home, thus eliminating the need to travel to a dialysis centre. There is mixed evidence suggesting that HD has a higher survival advantage over PD [262-267], and it is believed that survival differences vary substantially according to the underlying causes of ESRD instead [268]. Various techniques are used for PD. In automated peritoneal dialysis (APD), a machine is needed to fill and drain patient's abdomen. With continuous ambulatory peritoneal dialysis (CAPD), a patient does not require a machine and may even walk around with the dialysis solution in his abdomen.

4.8.2 Aim

The primary aim of this analysis was to enable a newly-diagnosed 36-year old Chinese female patient with diabetic ESRD to decide the mode of PD treatment, i.e., CAPD vs. APD. She was recommended by her nephrologist to receive PD in view of her lifestyle, co-morbid conditions and financial status. The incidence and prevalence of ESRD are well-known to be linked to the funding of dialysis [269]. Moreover, PD is better tolerated than HD, and she did not have recent abdominal wounds or surgery. Like the majority of new PD patients in Singapore, she was scheduled to dialyse regularly at a restructured hospital. To manage her condition and lifestyle better, she also participated in a detailed and systematic PD training.

The decision node was the use of a specific mode of dialysis (1. APD; 2. CAPD). The chance nodes were occurrence of peritonitis and mortality. PD

frequently results in peritonitis, an inflammation or infection of the abdominal cavity. This is usually caused by an infection gaining access through the drains placed in the abdominal cavity. Unless contamination persists, peritonitis usually does not progress and can be healed with proper treatment. However, it might cause a lot of discomfort to the patient on dialysis and if not attended to properly, death may be the potential outcome.

4.8.3 Data

There were relevant local data concerning the occurrence of peritonitis among patients on CAPD and APD. An on-going prospective observational study based on 100 ESRD patients with diabetes (mean age: 63.1; gender: 64.3%; mean dialysis duration: 20.7 months) dialysed with APD or CAPD from 1 March 2001 to 31 July 2006, was conducted. It collected detailed information on patients' demographics, albumin levels, comorbid conditions and development of peritonitis.

4.8.4 Results & Decision

Based on the recorded prospective data, the Bayesian Poisson regression model with prior Beta[1, 1] (3.22) showed that there was no significant difference in developing peritonitis among ESRD patients with DM between APD and CAPD, after adjusting for age, gender and conditions like IHD and cerebrovascular disease (Table 4.24). The exposure variable was time on PD.

The patient's probability of developing peritonitis with APD was 0.04 in 6 months, 0.09 in 12 months and 0.18 in 24 months. On the other hand, her probability of developing peritonitis was slightly higher with CAPD, i.e., 0.05 in 6 months, 0.11 in 12 months and 0.21 in 24 months. Gender was found to be the only significant

predictor. As expected, there was no significant difference in developing peritonitis at different time frames between APD and CAPD.

The patient's influence diagram is depicted in Figure 4.10. Her utilities were determined from an exponential function (risk tolerance: 52,200). The most favourable and unfavourable outcome were free from peritonitis and occurrence of the complication in 2 years, respectively. Computations show that she might prefer to dialyse with APD (expected utility: 0.48).

4.8.5 Discussion

Future decision analysis may involve information on mortality due to PD dialysis, especially for patients who are severely diseased and need to change the mode of dialysis or seriously in need of transplantation. Based on published evidence, the survival rate of patients on dialysis was 90.6% at 1 year, 78.8% at 2 years, 62.2% at 4 years, and 40% at 8 years [266, 270]. However, the most relevant reference [264] suggested that diabetic patients on PD faced a 13.5% chance of death after 1 year, 48% at 2 years, 66% at 3 years and 67% after 5 years. Another article also found that diabetic patients' chance of survival was lowered to 29% after 4 years [271]. Unfortunately, there was no such confirmative and readily available evidence in Singapore.

Reference also suggested that there could be an increase in the chance of survival should the patient switched to haemodiaylsis if her condition worsens [272]. Kidney transplant is a life-saving alternative to dialysis. It is expected that 90% of kidneys obtained from living donors are functioning properly 1 year after operation, and 3 to 5% of these kidneys stop functioning during each year that follows. About 70 to 90% of the kidneys from donors who has just died are functioning after 1 year,

and 5% to 8% stop functioning during each year that follows. Some references showed that the survival rate at 2 years was as high as 95% [270]. Like some other countries, the main causes of death of ESRD patients in Singapore were cardiovascular disease and infection.

From published records, about 58% of the total transplantation carried out in Singapore were cadaveric [272]. Kidney transplantation is a major operation where the donated kidneys is placed in the pelvis through an incision and is attached to the recipient's blood vessels and bladder. Rejection usually happen within 3 to 4 months after operation and the recipient must continue to consume immunodepressants throughout her life. Compared to the general population, kidney transplant recipients are about 10 to 15 times more likely to develop cancer.

The above results and discussion may form the basis for future studies, which may in turn help to formulate the medical guideline(s) concerning the treatment of end-stage renal failure.

Table 4.24: Non-informative Bayesian Poisson regression analysis on the occurrence of peritonitis for ESRD patients with diabetes

Covariates	Coefficient	95% P.I.
Gender *		
Male	Reference	—
Female	1.67	0.19—3.16
Age	-0.03	-0.07—0.02
Cerebrovascular disease		
No	Reference	—
Yes	0.74	-0.25—1.72
Ischemic heart disease		
No	Reference	—
Yes	1.26	0.03—2.50
System		
1. APD	Reference	—
2. CAPD	0.18	-0.90—1.26
Intercept	-7.48	-12.09— -2.86

* Statistically significant at 5%

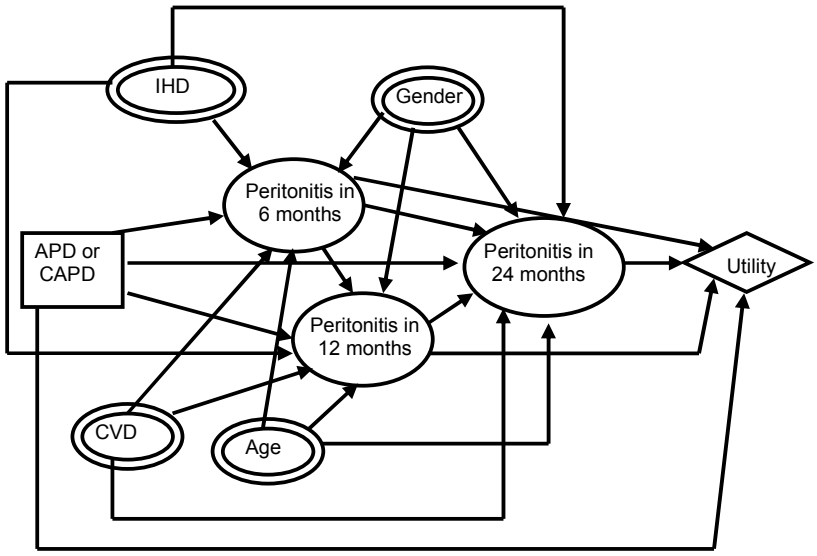


Figure 4.10: APD versus CAPD

4.9 Treatment of Asthma Patients at Special Centre

4.9.1 Aim

A 15-year old female acute asthma patient seen at a local hospital wanted to decide if she required special attention at the Emergency Department Treatment Centre (EDTC), a short-stay observation unit designed for serving patients who require more evaluation during an emergency visit. The EDTC was opened in 2001 to ease the overcrowding problem at the hospital.

The proposed analysis involved one decision node (admission to EDTC vs. admission to ordinary ward) and two chance nodes (unplanned admission in 24 hours vs. discharge, and unplanned readmission within 4 weeks vs. no relapse). Although the EDTC promises to deliver more specialised and high-quality care, the patient was more familiar with the ordinary ward where she would receive treatment from her familiar clinician.

4.9.2 Data

The data for encoding the necessary probabilities for the decision problem included 248 patients seen at EDTC from January to December 2006 (Table 4.25). Only aggregate data were available as the study was not initially designed for medical research. There were also other sketchy but relevant published evidences [273-277] useful for setting up the priors for the proposed Bayesian model (Table 4.25).

4.9.3 Results & Decision

Based on hospital record, a total of 779 asthmatic patients required admission during the study period. Of these, 248 were admitted to EDTC. The Bayesian model (3.36), with 100 burn-ins and 500 iterations thereafter, was employed to combine the

prior evidences for encoding the probability for 24-hour discharge from EDTC. Based on the following Bayesian equation, the combined prior evidence was then updated with the collected hospital data:

$$\pi(p) \propto p^{y}(1-p)^{n-y} \times p^{a}(1-p)^{b} \tag{4.3}$$

where n is the total number of patients admitted to the hospital's EDTC during the study, y the number of patients discharged within 24 hours and a and b are the beta parameters derived from the combination of prior published evidences. The same scheme works for generating the probability for readmission within 4 weeks. In the case of discharge within 24 hours at the common ward there was only one set of evidence available. As such, the figures were updated directly with the hospital's record with (4.3).

The results are shown in Table 4.26. Given the patients' preferences and utilities, she should choose to be transferred to the common ward as her expected utility was 1.14. This was higher than that with EDTC at 0.86. As described in Figure 4.11, the patient had more preferences for the common ward.

Table 4.25: Evidences concerning emergency treatment centre

	EDTC		Common Ward	
Study	Discharge within 24 hours n/N	Readmission within 4 weeks n/N	Discharge within 24 hours n/N	Readmission within 4 weeks n/N
Prior:				
O'Brien 1980	328/434	4/434		
Willert 1985	35/52	5/52	16/51	11/51
Miescier 2005	40/161	—	—	—
Levett 2006	3379/4446	—	—	—
Arendts 2006	—	29/211	—	—
Collected hospital data 2005-2006	181/248	12/248	425/531	30/531

Table 4.26: Results of Bayesian analyses

	Recurrence p	95% P.I.
EDTC		
Discharge within 24 hours	0.73	0.68—0.79
Readmission within 4 weeks	0.05	0.04—0.06
Common ward		
Discharge within 24 hours	0.77	0.72—0.82
Readmission within 4 weeks	0.07	0.06—0.08

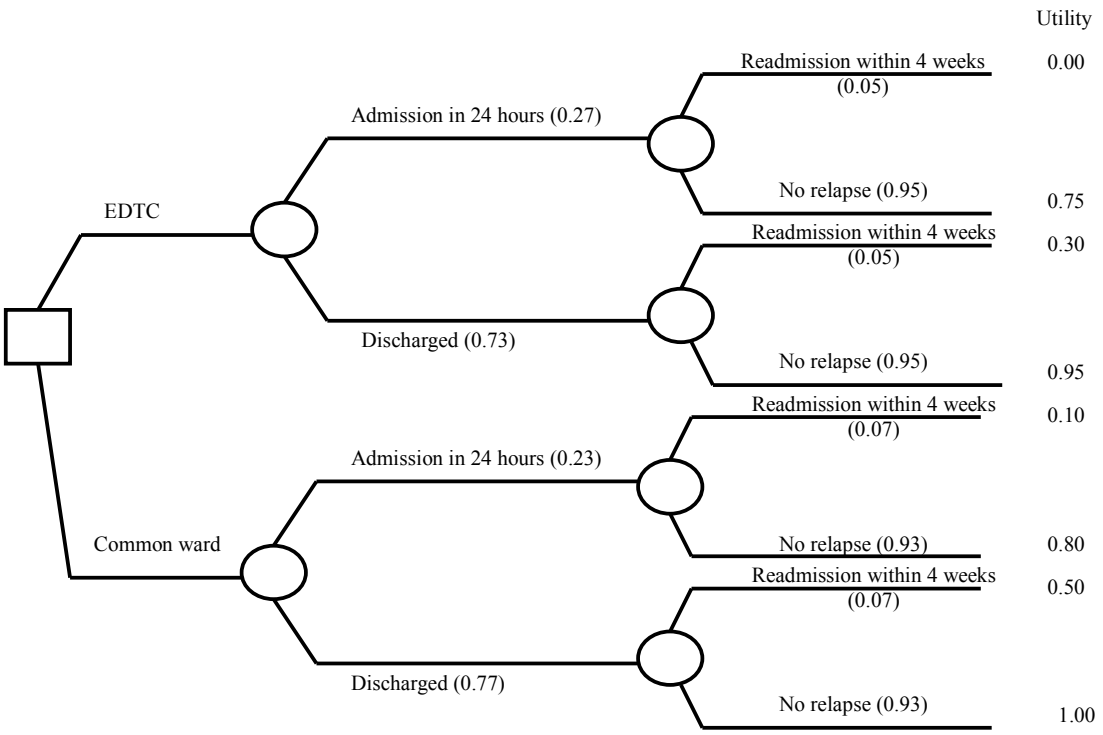


Figure 4.11: EDTC versus common ward

4.10 Polychemotherapy for Treating Early Breast Cancer

4.10.1 Background

Breast cancer is notoriously known as the most common form of cancer among women. The risk of breast cancer arises from a combination of genetic susceptibility and environmental factors, including old age, young age at puberty, family history, prolonged use of oral contraceptives or estrogen therapy, obesity after menopause, exposure to radiation, and presence of fibrocystic breast disease and certain genes (BRCA1 and BRCA2).

In Singapore, about 55 per 100,000 Singaporean women were diagnosed with breast cancer from 1998 to 2002, and the incidence rates have increased by 3 times since 1968. The pattern of increase over time is also consistent across all ethnic groups. Interestingly, the age pattern for 1998-2002 suggests that the highest age-specific incidence rate is occurring progressively later in life (35-60 years old) and 54% of all cases occurred in women 50 years and above.

However, fewer women actually die from the disease now, thanks to early detection and major advancements in chemotherapeutic drugs. Early breast cancer or carcinoma in situ (non-invasive cancer) is usually removed by surgery and chemotherapy (anti-cancer drugs) is used to prevent recurrence. Currently, polychemotherapy (multiple cancer drugs) are at the forefront of therapy, and several estrogen-regulating drugs have been developed as well. New drugs are also under development, including those derived from natural products.

4.10.2 Aim

While the potential for a new drug is thrilling, it may be more practical to work with what is currently available. It is, therefore, timely to compare the

effectiveness of prolonged polychemotherapy (6-24 months) and the shorter treatment (3-6 months) in terms of recurrence and mortality. The encoded probabilities would help a 52 year-old obese woman, who was trained as a clinician, to decide if she needed prolonged polychemotherapy after surgery. She was diagnosed to have early breast cancer and the tumour was less than 2 cm in diameter and had not invaded the surrounding tissue or spread to other parts of the body before surgery. The decision problem was simple in nature, involving 1 decision node (prolonged polychemotherapy vs. short polychemotherapy) and 2 chance nodes, namely first recurrence and all-cause mortality.

4.10.3 Data

A thorough literature search through MEDLINE and the Cochrane Library up to February 2007 [278-280] identified 11 published studies comparing the effectiveness of prolonged polychemotherapy and the shorter treatment (Table 4.27).

Bayesian model (3.36-3.37) was applied to encode the probabilities for first recurrence and mortality for the two treatments. For the analysis of recurrence, the prior for overall or combined proportions of prolonged polychemotherapy and shorter polychemotherapy were Beta[5, 5] and Beta[5, 4], respectively. This reflected the belief that prolonged polychemotherapy was slightly more effective. The prior for between-study precision was fixed as Exponential[1] for both cases. The burn-ins and updates of the MCMC procedure were set at 500 and 1,000, respectively. On the other hand, the probability of death was expected to be 0.7 if the patient suffered relapse. This was substantially reduced to 0.1 if the patient did not relapse. The data was entered into Stata 9.0 for analysis.

4.10.4 Decision

The published data are shown in Table 4.28. Prolonged chemotherapy is shown to be more effective in reducing recurrence than the shorter treatment. Due to the small number of studies involved, the 95% P.I.s are fairly wide.

Given her preferences and utilities the breast cancer patient should choose the shorter treatment (Figure 4.12). The expected utility with prolonged chemotherapy was 0.47, which is lower than that of the shorter treatment at 0.52.

While prolonged chemotherapy is more effective, it is less tolerant than the shorter treatment. This was reflected in the patients' self-assigned utilities. This example also illustrates the fact that a "better" treatment from the clinician's point of view may not necessarily provides patients with the highest satisfaction.

4.10.5 Discussion

Breast cancer is not a disease of modern society as it was recognised by the ancient Egyptians as early as 1600 BC. Unfortunately, many centuries have passed and still no acceptable cure has been discovered. Worst of all, breast cancer is now affecting as many as one in eight women during their lifetime. Not only is the diagnosis of breast cancer frightening, the therapies used to treat the disease are just as daunting—such as surgery or chemotherapy.

As such, clinicians are also looking for ways to make current drugs more effective and less toxic. Patients with early stage breast cancer exhibit promising prognoses to polychemotherapy and have also shown to significantly reduce clinical recurrence among women of all age groups. The above analyses, however, suggest no significant differences between prolonged polychemotherapy and shorter polychemotherapy in terms of recurrence and mortality.

Even though the exact cause of breast cancer has not been fully identified, diagnosis and treatment have improved dramatically in recent years, and many clinicians believe a cure is within reach.

Table 4.27: Published evidences concerning first recurrence with polychemotherapy for treating early breast cancer

	Recurrence	
Study	Prolonged Polychemotherapy n/N	Shorter Polychemotherapy n/N
IBCSG VI-VII	129/394	139/393
INT Milan 7502	134/215	133/219
SAKK 27/76	141/211	148/208
SEC SG 1	98/208	101/221
SWOG 7827C	115/225	129/220
Boston	95/148	109/151
FASG GFEA 01	76/207	87/193
GBAG 3 Germany	153/396	134/370
GBAG 2 Germany	111/239	120/242
IBC SG VI-VII	225/700	240/696
Metaxas Athens	27/106	33/142

Table 4.28: Results of Bayesian analyses

	Recurrence p	95% P.I.
Prolong chemotherapy (6-24 months)	0.48	0.30—0.80
Short chemotherapy (3-6 months)	0.54	0.32—0.85

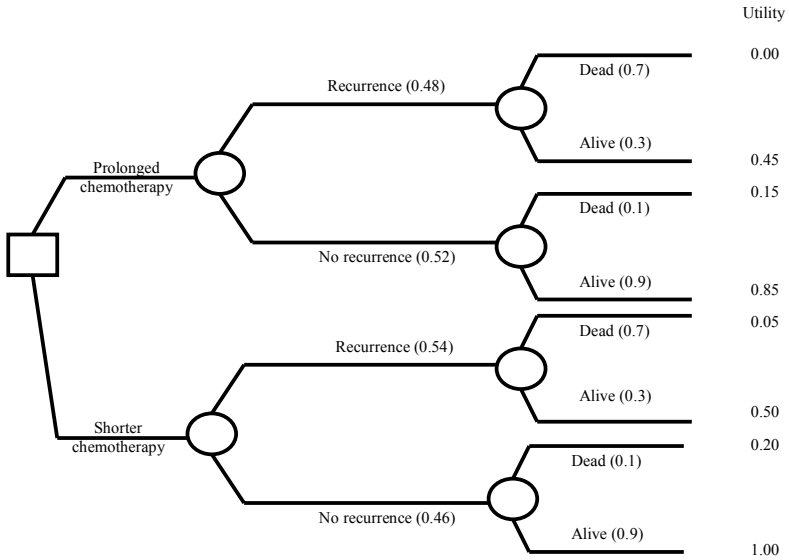


Figure 4.12: Prolonged chemotherapy versus shorter chemotherapy

CHAPTER 5

DISCUSSION & CONCLUSION

In medicine and in all other health sciences, there is a constant search for the "best" evidence [281]. The Bayesian models developed and applied in this dissertation serve to fulfil this aim. In fact, the application of the proposed Bayesian probability-encoding models will bring considerable impact on the way EBM and medical science is practiced. The contribution of Bayesian ideas is not limited to the solving of a specific medical decision problem. The Bayesian framework changes our views about current medical paradigm, scientific methods, inductive logic, nature of medical evidence, systematic review of medical evidences and the roles of patients, clinicians and EBM practitioners.

The following discussion revolves around a common theme, that is, what would clinical practice and EBM become if the Bayesian framework is adopted. The discussion will also touch on the future directions of methodological research, in view of the latest developments in Bayesian statistics and computer-intensive techniques. As mentioned in Chapter 1, EBM—fast becoming an encompassing field that integrates clinical practice with decision analysis—will serve as a good testing ground for new developments in decision analysis.

5.1 The Scope of Medical Practice

Modern medicine, better known as scientific medicine, western medicine or biomedicine, is based upon whatever its practitioners regard as "scientific knowledge", which is

usually loosely defined as that which is objective, demonstrable, measurable, observable, reproducible, or technologically advanced. While the concept of "what is scientific" changes over time and varies across the globe, scientific medicine adopts a more or less single-minded, materialistic approach which values accurate observations and unambiguous measurements. It reduces all bodily functions and dysfunctions to mechanical and biochemical reactions, knowable material causes and structural flaws that can be studied in isolation from the sufferers. Clinicians' judgements and decisions are the results of such unambiguous, if not entirely flawless, rational deductions and empirical investigations.

Such attitude has brought numerous achievements to mankind. As long as the patient's disease can be examined and accommodated within the boundary of current "scientific knowledge", it has a good chance of providing a successful cure or at least some alleviation. If it does not, scientific medicine may have little to offer and may even cause more harm to the sufferers. The merits and demerits of contemporary medicine can be clearly demonstrated in the different remedial demands of acute and chronic illnesses. Scientific medicine is, on the whole, more efficient in treating acute illnesses. But it offers less successful solutions to the chronic diseases (e.g., renal failure, diabetes, osteoarthritis, geriatric-related illnesses, immunological disorders, pituitary insufficiency and allergy) which require not only therapeutic interventions, but also long-term medical care for addressing the whole of patient's prognosis and quality of life. What are characterised and believed as effective cures may save or transform the lives of individuals, but their effects on long-term mortality and quality of life are less impressive and there may be devastating side-effects in using them. Responsible surgeons are also

aware that many operations, once considered vital and valuable, turn out to be unhelpful, redundant or even detrimental to lives, let alone improving the patients' general well-being. Uncertainties pervade all aspects of healthcare, despite the awesome rapid advancement of medical science.

Contemporary medical practice, at first glance, is moving from triumphs to triumphs. It is now possible to examine at every cell in the body and to operate instruments by remote control, often far inside the body, perhaps on a heart or blood vessel. Yet, there are, paradoxically, also times of growing dissatisfaction and the common public complaints are largely related to the fact that it has become highly institutionalised, excessively technological, unaffordable, alien and impersonal. Unintelligible jargon, expensive modus operandi, unfriendly medical hardware and ineffective communication with the clinicians make it more forbidden to the paying patients. Most medical decisions concerning patient care are usually made on behalf by the clinicians as the details involved are highly technical and this leaves little room for participation by the sufferers themselves. Somehow the more scientific medicine achieves, the less it satisfies. We are not only living in times preoccupied by the fear of unknown fatal diseases but also in fear of medical treatments. It is fair to say that scientific medicine has many proven means of saving and improving life and these are constantly increasing in number and efficacy, given the current technological advancement. However, clinicians must also accept the fact that, while their diagnosis and therapeutic methods are successful in general, they do not always provide answers in individual cases.

Contemporary medicine prides itself on being "scientific", but is necessarily based on the availability of evidence, and clinicians' judgement, interventionist skills and healing expertise. This immediately creates a number of paradoxes. Can personal and clinical experience be "scientific"? What is the relationship between medicine and science and how does it influence the way medicine should be developed? More specifically, why do clinicians vary so much in their practice? The differences in their judgements and decisions—both about the general guidelines to adopt and what specific therapeutic actions to be taken in the course of treatment—are critical to their patients. One may question whether standardised guidelines of medical care, which breeds consistency among clinicians, can be established in principle, and whether strict adherence to such guidelines, if enforced, would ensure successful outcomes. These issues offer much food for thought for the medical community to chew over.

The root of these questions may be traced to the way medical decisions are made and our current understanding of the scope of scientific knowledge and evidence. Clinicians make numerous decisions related to their practice on a daily basis and these directly or indirectly impact the health and welfare of their patients. The validity of their actions is based on evidential merits. Evidence is not a decision itself; however, good evidence is required for making sound decisions.

Such notion may bring considerable impact on the way EBM is practiced. Founded upon the principle that evidence should guide clinical practice, EBM has brought about numerous illuminating contributions to medical practice. Common arguments in support of EBM include improvement in clinicians' knowledge, better communication between clinicians and patients, and more effective use of scarce

resources. Abandoning the use of uncritically and unsystematically evaluated clinical research, EBM also provides a useful framework for gathering, evaluating and disseminating medical evidence. However, it is somewhat surprise to learn that EBM has not developed a new concept of evidence [28] despite making tremendous contributions to medical practice. We need to have a broad vision of evidence that embraces the inherent complexity of EBM. It is, therefore, suggested in this dissertation that evidence may refer to any explicit warranted reference for supporting or rejecting a hypothesis, claim or belief. Evidence may be tangible or intangible and could exist in both objective and subjective states. With this in mind, the established definition that "evidence is a fact or datum which is used, or could be used, in making a decision or judgement or in solving a problem" [25] is somewhat limited and should be duly modified.

The conventional notion of empirical observations only skims the surface of evidence. Through the better understanding of medical evidence and the illustration of the Bayesian methodology, in which subjective opinions could be combined with objective evidence, this dissertation aims to provide an alternative approach to the current practice of EBM. The application of various specific Bayesian models in decision analysis, as illustrated in the case studies in Chapter 4, is a small step towards the goal.

However, a series of epistemological questions remains and there are many gaps of knowledge need to be filled urgently. This requires a proper dissection of the unique nature of Bayesian EBM.

5.2 Bayesian Evidence-Based Medicine

The application of Bayesian thinking in medical decision analysis and EBM is a key to unlock many of the questions outlined above. It emphasises the working with "best" evidence, although the ideal concept of "best" poses a formidable challenge to the minds of most EBM practitioners. The traditional EBM practice suggests that the best evidence should be sought from confirmed scientific knowledge, laboratory research, randomised controlled clinical trials and large-scale observational studies. It places opinions of respected authorities, based mainly on experience, or reports of expert committees, way below evidence obtained from "objective" and "scientific" studies.

In reality, however, the scope of EBM is wider than what is perceived. Results from case reports, quasi-experiments, qualitative research and descriptive and analytical studies concerning screening, diagnosis or prognosis have also contributed immensely to our knowledge. This effectively points out that we have not fully understood the scope of scientific knowledge. As such, the most fundamental issue concerning EBM is none but the reconstruction of a complete and unambiguous knowledge of methodology and evidence. This, however, hinges on our understanding of what "science" is and should be.

Science at large has been immensely influenced, if not completely dominated, by the empiricist school of thought—an approach that is primarily concerned with observable facts and eschew moral or metaphysical speculations. Such stance qualifies observable phenomena as the only source of knowledge and clinicians are consistently reminded to respect observed evidence and warrantable factual records unconditionally. Empiricists use the criterion of verifiability to distinguish between medical conjectures, theories and decisions. In the strictest sense, however, such criterion is not verifiable in

itself and with this spirit, the proposed Bayesian framework justifies that observations may not constitute the only source of knowledge. Consequently, the empiricists' position may not be tenable. The Bayesian framework maintains that clinicians' empirical knowledge must be organised according to a prior principle and the subjective nature of observation must not be neglected. In essence, observation depends upon the clinician's preconceptions. This offers an alternative and supplementary approach in knowledge acquisition and our picture of the world reflects both a priori organisation of perceptions, rational deductions, beliefs, opinions, past knowledge and observed evidence.

Undoubtedly, the empiricist approach has served the scientific community well and will continue to do so. It provides us with an objective framework for which clinicians formulate their decisions and defend their chosen actions. However, one must begin to recognise the underlying limitations of this approach. Contingent on the availability of observable evidence, it is often difficult, if not impossible, to formulate sound decision(s) in real practice. In short, our current understanding of science, very much limited by the empiricists' definitions, is far from perfect. We need to broaden its scope.

This dissertation offers an alternative approach to EBM by revising the current concept of medical evidence. Defining evidence as "an explicit warranted reference given in an appropriate and specific context for supporting or rejecting a hypothesis, claim or belief", it encompasses any facts, data or information, whether weak or solid, obtained through experience, published results and observational and experimental research. A reference qualifies as evidence so long as it is relevant either to the understanding of the problem or to the clinical decisions (diagnostic or therapeutic) made about the case. In

Bayesian EBM, the "best" evidence should be combined with all components of decision making, such as expert opinion, knowledge from clinical studies, and patients' preferences and values. A primary factor affecting the decision-making process is prior clinical experience, self-evident intuition, published evidence and testimony from fellow clinicians. These in turn form the basis for formulating expert opinion and generating objective evidence from research studies. As such, the next burning question facing EBM practitioners is how to allow subjective expert opinions be combined with objective evidence.

The proposed Bayesian framework offers a wide range of models for probability encoding and data analysis useful for medical decision analysis. It unambiguously asserts that one always forms an incomplete picture of a phenomenon with his subjective horizon of understanding (prior). Through observed evidence, the EBM practitioner develops more insights to the phenomenon encountered and the final interpretation is achieved with the fusing of the subjective and objective knowledge horizons. The posterior understanding incorporates the subject's pre-knowledge and his revised understanding of the phenomenon. During the process, the EBM practitioner tries to better understand the phenomenon and correct his "prejudice" caused by his prior beliefs and opinions.

Through the Bayesian framework the synergism between subjective and objective evidence come into play, with the EBM practitioners and subject experts actively giving valid testimony and searching for relevant evidence useful for decision making. Bayesian EBM fulfils the primary and noble objective of the early advocates of EBM by making use of the most complete evidence available on diagnosis and treatment.

Unlike the conventional scientific approach popularised by the empiricists, Bayesian EBM recognises the special contributions of expert opinions in all aspects of data analysis. In fact, expert opinions may be viewed as the first-line evidence in decision making. EBM practitioners should give utmost care to such opinions, recognising that instincts and independent thinking are invaluable assets of an experienced clinician. When presented with a decision problem, the subject expert should illustrate his unique prior understanding, which may be based on experience or previous evidence collected for a similar purpose.

This dissertation supports Bayesian EBM not only because of its philosophical and pedagogical validity but also its appealing features of data analysis. The Bayes' Theorem [26] shows how inverse probability could be used to encode probability of antecedent events from the occurrence of the consequent event. Because of this, Bayesian models are "optimal" in terms of post-data evaluation, given the evidence that actually occurred. As a consequence, Bayesian models usually outperform the conventional quantitative models in the post-data setting.

In particular, several of the specific models introduced in this dissertation are useful for evidential review and literature critique, which is the core business of EBM. These models may be applied for analysing existing individual data obtained from experimental or observational studies (historical as well as prospective), or for shedding light on the obscure meanings of aggregate published data.

Like all academic disciplines, Bayesian EBM seeks to understand the unknown horizon and attempts to make predictions about it, in the hope of controlling the uncertainties and providing useful hints for decision making. Unknown quantity is a

generic term referring to any value not known to the investigator in this instance. The ideas that form the basis of the Bayesian approach are as follows:

- since one is uncertain about the "true" value of the unknown quantities (commonly known as parameters in the statistical literature) one should consider them as random variables
- the opinion-based priors, mainly elicited from widely-regarded experts, are subjective in nature
- on the other hand, priors based on published evidence or recorded information are essentially objective
- both sets of priors effectively measure how plausible the EBM practitioner considers the unknown values should be before observing/analysing the objective evidence
- the EBM practitioner revises his beliefs, opinions, pre-understanding, prior knowledge after getting the evidences through the Bayesian models and this gives rise to a posterior distribution
- the posterior forms the basis for evidentiary analysis and probability encoding for decision making

Allowing the unknown parameters to be random quantities, one makes probabilistic statements about them conditional on the sample and prior evidence. This is a very unique and attractive feature of Bayesian analysis, where decision making utilises probability statements as the basis for inference. In fact, all probability statements about

the unknown quantities should be more appropriately interpreted as "degree of belief". This contrasts significantly with the conventional approach in data analysis where inference probabilities are solely generated from observed data that believed to have occurred for the fixed parameters.

The Bayesian framework coheres with the current notion of medical care, with provision of sound clinical decisions useful for patient healing as the primary concern. Medicine may be a body of scientific knowledge, but healing is a personal skill. The term "healer" has implied many different meanings to cultures across the world throughout history. In modern day, clinicians of different disciplines are responsible for the physical care of the ill or diseased. Although their precise roles vary in part due to their type of training, they are responsible for the physical and spiritual care of the patients. As such, healing is more an art than a science. A good healer should demonstrate knowledge and competency about the sufferers' conditions. All decisions concerning healing require personal skills and the quality of decisions depend largely on the immense experience possessed by the clinicians.

One philosophical question that puzzles EBM practitioners remains. It is whether one can attain knowledge through the proposed Bayesian framework and whether it corresponds to scientific truth. This dissertation demonstrates that knowledge production becomes possible only when evidence is interpreted by the subject. It remains non-informative if it is not interpreted. The Bayesian framework also suggests that it is "legitimate" for the EBM practitioner to possess a pre-understanding horizon and allows the prior evidence be fused with objective data. The differences in opinion among clinicians may be attributed to the inadequacy in current biomedical knowledge and

patients' peculiar conditions, or simply due to their personal preferences and opinions that govern their judgement. However, the end product (posterior) should be a richer and deeper understanding of the problem investigated, with meanings of the evidence adequately elucidated. This in turn helps the subject to become a knower.

Following the above argument, scientific medicine is a conjectural discipline. Such acknowledgement has a profound influence on how decisions should be made in clinical practice. First, recorded data is never the sole basis of medical decision-making. Second, clinicians should not view diseases merely as biological dysfunction and patients' personal preconceptions must be incorporated in the decision-making process. Third, to generalise the above assertions, there is no pure objective evidence in practice. While many clinicians proudly claim that their decisions are based on scientific merits, the much-publicised variations in medical practice may nullify their claim. Realistically, all individuals experience different knowledge acquisition processes and their preconceptions would, to a large extent, influence how they interpret observable evidence. Depending on personal background and training, a clinician may rely on experience, intuition and subjective judgement alone, or may enhance these peculiar attributes with objective measures for identifying the decision(s) that would lead to the most desired results for his patients.

While the Bayesian framework acknowledges that knowledge may be developed from subjective means (priors involving expert opinions), there are categorical differences between knowledge and opinions, which merely indicates an attitude or belief towards a phenomenon. Knowledge implies having evidential justification. It is through the revision process, as succinctly described by the Bayesian framework, that knowledge

evolves from opinions. Moreover, knowledge possesses a certain property which opinion lacks; that is, the property of generalisability. However, the attainment of knowledge helps to reshape or strengthen one's expert opinions, which in turn gives rise to warranted knowledge. In the event where published evidence is used as a prior, the Bayesian framework reflects that the advance of knowledge (posterior) consists in the modification of earlier knowledge (prior). In a nutshell, the proposed Bayesian framework coheres with the way we learn.

So what are the implications if the Bayesian framework is correct? How does it challenge the traditional views of scientific theories? Traditionally, there are two major schools of thought with differing views concerning the nature of truth. First, the realists maintain that truth is an agreement between theory and evidence. The role of science is to identify and discover the entities that surround us and establish their relationships to one another. A theory is deemed "true" when the entities it refers to and the relationship it describes correspond to real entities that exist in the world and their real relationships. The instrumentalists, on the other hand, believe that a theory is just an abstraction and representation of truth and it is meaningless to ask whether it corresponds to reality. What really matters is how much its predictions agree with observations.

Like instrumentalism and realism, the Bayesian framework maintains that truth is important to us when we interpret evidence. However, it argues that the realists and instrumentalists are fundamentally wrong in holding the belief that one can somehow observe the real world independently of their beliefs and theories. In reality, our beliefs and pre-understanding often shape the way we interpret and gather evidence. This implies that observations are value-laden and there is no absolute truth. Different people

may interpret evidence differently. More importantly, truth is relative and ever-changing and is open to more than one interpretation. That does not mean there is no truth. Rather, it does not exist independently of the perception of the researchers who try to understand it through evidentiary analysis.

This dissertation asserts that "true interpretation" is value-laden, relative, transient and ever-changing, as it does not exist independently of the perception of the EBM practitioner. Under the Bayesian framework, the true interpretation is the one that best coheres with both the prior evidence and objective evidence. Truth may come to light from the union of these horizons, albeit an uncertain or a transitional one.

Lastly, this dissertation does not intend to suggest that all problems concerning medical decision analysis are solved with the acceptance of the Bayesian framework. To provide sound solutions to medical decision problems, clinicians must have solid information about the consequences of different choices and must be able to process that information accurately [282]. Unfortunately, many clinicians are unable to make consistent decisions with the use of medical evidence. Previous research has also found that clinicians asked to consider an individual patient generally make different decisions than those asked to consider a group of comparable patients [283]. In fact, a recent article even suggests that trained statisticians may make fallacious judgements about evidence [284].

However, one should not be over-pessimistic about the future of EBM. EBM practitioners should take up the challenge to look for better evidence in support of their practice and communicate their ideas with their fellow clinicians to upgrade the current medical practice.

5.3 The Future Development of EBM

A competent EBM practitioner should be an informed knower, an effective communicator with his patients and fellow clinicians, a devoted and passionate healer, a diligent evidence-seeker, an avid and adroit user of evidence, an alert critique and analyst, an insightful and rational decision-maker, a brave explorer, an imaginative methodologist, and a keen learner ready to embrace state-of-the-art quantitative techniques in his practice. An immediate implication of adopting the Bayesian framework to EBM practice is that practitioners need to upgrade themselves constantly. An EBM practitioner must not only be competent in his field of specialisation and applications, but also able to guide his chosen experts to reveal their opinions useful for decision making. With such valuable attributes, EBM investigators will be able to apply the "best" evidence available and practice in a setting where the conventional framework does not permit. Recognising the importance and usefulness of incorporating priors in decision analysis, Bayesian EBM encourages practitioners to be more proactive in seeking evidential support from fellow clinicians. Through more collaboration, EBM practitioners with different skills and expertise will be available for co-operation and discussion.

5.3.1 Broaden Sources of Evidence

In future, EBM will rely on an increasingly complex surveillance of evidence. It will integrate elements from a much broader perspective. The pragmatic benefits of integrating results from other health-related fields and complimentary and alternative medicine (nutrition, medical herbalism, Chinese traditional medicine, chiropractic therapy, homeopathy, osteopathy with naturopathy, etc.) is readily seen. Acupuncture

has now been recognised by scientific medicine as a safer alternative to treat acute back sprain and conventional anaesthesia for frail patients to undergo the trauma of minor surgery. A large number of studies involving human subjects and animals have also demonstrated that green tea polyphenols possess cardioprotective, neuroprotective and antimicrobial properties [285-290], provide protection against untraviolet light-induced DNA damage, ageing, obesity and dental caries [291-293], and may even help to improve the effectiveness of chemotherapy in treating cancers as a synergistic agent. In order to identify the "best" evidence, EBM practitioners need to consider more sources of evidence.

Following a similar line of argument, clinical case reports will play a more important role in the future of Bayesian EBM. Clinical experience—the first and foremost source of medical expert opinions—depends heavily on the observation of clinical case reports. Long been recognised as the "special cases that advance the knowledge, research and practice of medicine" [294], case reports have been familiar elements of medical journals. Unlike extensive research such as randomised controlled trial, cohort study and case-control study, case reports published in medical journals, in one form or another, usually comprise much fewer subjects. Although the results are not generalisable, they are the only source of evidence about rare diseases. Where else can a clinician turn to for guidance and inspiration while treating such unusual and unfamiliar diseases? There are no existing guidelines which the clinicians may follow. Case reports, however limited, may also help EBM practitioners to develop hypotheses for further studies. This also means that case reports may serve to formulate the objective priors for Bayesian analyses. Therefore, case reports should form an integral part of a

comprehensive database together with evidences generated from full-scale research studies.

Being able to draw on more than one system of knowledge means that EBM practitioners are exposed to more useful and relevant evidence for identifying alternatives for medical decision making. In addition, it also enriches the knowledge base of EBM practitioners. Undoubtedly, this offers more alternative treating plans for patients in the future.

In fact, patients will become a more integrated part of decision making. Disadvantaged patients, like the rest of the healthy citizens and as equals in their humanity, have their claim to engage in their care. EBM will be viewed as "medicine with a human face", making a place for itself somewhere between bedside clinical work with the patient and decision making [295]. Favoured by health policy-makers [296], EBM decisions will be increasingly accountable in courts of law [297-298].

5.3.2 Power Priors

Being evidence-based, EBM's success will be evaluable and critically evaluated. This in turn propels continual methodological advancement. To excel in his endeavours, a methodologist must continue to sharpen and upgrade his analytical tools. One expects to witness more exciting and ground-breaking activities to take place in the applications and development of Bayesian methodology in the near future.

Recognising that prior elicitation plays a crucial role in Bayesian analysis, the power prior model [299-300] enables analysts to construct informative prior based on historical data. It takes the following general form:

$$g(\alpha \mid \text{evidence}, a_0) \propto h(\alpha) \times L[\alpha \mid \text{evidence}]^{a_0} \tag{5.1}$$

where $h(\alpha)$ is the initial prior for α before observing historical evidences. The parameter a_0, usually ranges from 0 to 1, serves to weight the historical evidences and therefore controls the heaviness of the tail of the prior for α. By setting $a_0=1$, one suggests that the historical evidences are crucial in determining the current evidences and the model (5.1) corresponds to the posterior distribution of α from previous studies. If $a_0=0$, then (5.1) does not depend on historical evidences. The model may be completed by specifying a prior for a_0, which usually assumes a beta distribution or a truncated normal or gamma distribution. Since its development, several attempts were made to facilitate relational analysis based on historical evidences. These include the applications in generalised linear mixed model, logistic regression and survival analysis [101].

5.3.3 Beta Regression

A more recent methodological development which offers good opportunity for path-breaking discoveries lies with the application of beta distribution as the likelihood for observed evidences. Beta distribution, well-known for its versatility as it can model data of all shapes, has a rather unjustifiable limited use in contemporary conventional and Bayesian analyses. As described in Chapter 3, most of the Bayesian models make use of beta distribution as a default conjugate pdf. As a bounded distribution, however, it has many potential applications in medical decision analysis. Many medical outcomes are

bounded in nature (proportions, percentages, etc.), so it may be inappropriate to consider unbounded distributions as a close approximation. In fact, even time to event data are bounded as a life cannot go on indefinitely.

Sadly, beta distribution is only beginning to gain popularity among conventional statisticians in recent years. At its infant stage, most statisticians focus on the application of standard beta distributions ($0 \le y \le 1$; a, b>0) [301-302]. However, this dissertation advocates the use of the more general 4-parameter beta distribution for handling bounded outcome (y) via a relational model [303]:

$$f(y_i; a, b, c, d) = \frac{(y_i - c)^{a-1}(d - y_i)^{b-1}}{B(a,b)(d-c)^{a+b-1}}, \qquad c \le y_i \le d;\ a, b>0;\ b>a \tag{5.2}$$

where a and b are shape parameters, c and d the bounds of y (location parameters), and B(a, b) the beta function. The mean and variance of the distribution are $E[Y] = c+(d-c)a/(a+b)$ and $V[Y] = (d-c)^2ab/[(a+b)^2(a+b+1)] = (E[Y]-c)(d-E[Y])/(a+b+1)$, respectively. Note that the variance is a function of the mean and a dispersion parameter (a+b). With the standard form (i.e., c=0, d=1) as its special case, the proposed distribution is extremely flexible as it allows y be bounded on any interval. It is applicable for modelling all bounded outcomes, with proportion as its special case.

To estimate the effects of predictors on y, one specifies the following link function:

$$E[Y] = \mu = c + (d-c)\frac{e^{x'\boldsymbol{\beta}}}{1+e^{x'\boldsymbol{\beta}}} \tag{5.3}$$

where $\boldsymbol{\beta}$ is a vector of unknown parameters and **x** the observations of the predictors. The dispersion is $\phi = e^{z'\boldsymbol{\alpha}}$ and **x** and **z**, may be distinct. The formulation offers a direct interpretation of the predictors' effects upon E[Y] and dispersion. As readily seen, one advantage with the proposed model is that it handles dispersion explicitly. Following the above discussion, the variance function can be easily derived:

$$V[Y] = \frac{(\mu - c)(d - \mu)}{\phi + 1} \tag{5.4}$$

For a fixed E[Y], a large dispersion results in a small V[Y]. After some algebraic manipulations, the shape parameters can be specified as follows:

$$\begin{aligned} a &\equiv \frac{a}{a+b}(a+b) = \frac{\mu - c}{d - c}\phi \\ b &\equiv \frac{b}{a+b}(a+b) = \frac{d - \mu}{d - c}\phi \end{aligned} \tag{5.5}$$

Therefore,

$$f(y_i;\, c, d, a, b) = \frac{\Gamma(\phi)}{\Gamma(\frac{\mu - c}{d - c}\phi)\Gamma(\frac{d - \mu}{d - c}\phi)} \frac{(y_i - c)^{\frac{\mu - c}{d - c}\phi - 1}(d - y_i)^{\frac{d - \mu}{d - c}\phi - 1}}{(d - c)^{\phi - 1}} \tag{5.6}$$

where $\Gamma(\bullet)$ is gamma function. Following conventional wisdom, analysts would estimate $\boldsymbol{\beta}$ by maximising the log likelihood, namely $\ln \prod_{i=1}^{n} f(y_i; c, d, a, b)$. But the Bayesian

analysts must identify the appropriate priors for the parameters in analysis. Though daunting, this promises to be an exciting field for exploration in methodological research. Other bounded distributions [304] may also be considered.

5.3.4 Generalised Linear Latent and Mixed Model

Another potential area of research in probability encoding with advanced statistical techniques lies with the enhancement of the hugely popular generalised linear latent and mixed model (GLLAMM) [305-309]. As a class of multilevel latent variable models for outcomes of mixed type, GLLAMM (www.gllamm.org) seeks to unify all multivariate statistical models, which include GLM [91], survival analysis [97], hierarchical models [107], latent class analysis [310] and structural equation models [311], etc.

The GLLAMM is specified with:

- the conditional expectation(s) of the outcome given the latent and observed covariates
- structural equations for the latent variables on covariates
- the distribution of the latent variables

As in the Bayesian models for synthesising evidences from various sources, GLLAMM includes latent or unobserved variables that are interpreted as random effects. Moreover, the latent variables can very at different levels. The current version of GLLAMM is run with Stata's maximum likelihood commands, augmented with the adaptive quadrature.

One may view the bounded regression model with beta distribution (5.6), discussed within the context of GLM, as a special case of GLLAMM. It will also be a great achievement if a Bayesian version of GLLAMM is developed.

5.3.5 Bayesian Belief Network

The idea of applying networks to represent probabilistic information was conceived in the 1960s [312] and evolved rapidly in 1980s, thanks largely to the advances derived from the seminal works of Pearl (1986, 1988, 1995) [313-315]. The Bayesian belief networks (BBN) have developed at the interface between statistics, artificial intelligence and expert systems.

A BNN consists of a network of nodes connected by direct links, with a probability function attached to each node [20, 316], which represents a variable. Each node is in turn made up of states, or a set of probable values for each variable. Beliefs are the probability that a variable will be in a certain state based on the addition of evidence in a current situation. Supported by the Bayes' rule, every node also has a conditional probability table associated with it. The nodes are connected to show causality with an arrow indicating the direction of influence. As such, BNNs are graphical models that encode probabilistic relationships among variables of interest [317].

The BNNs are extremely useful when the information is vague, incomplete, conflicting and uncertain. For example, the consulting experts for a particular medical decision problem may be uncertain about his knowledge. By providing a structured combination of diverse lines of evidences, BNNs are able to address many real-life

medical decision problems and serve as decision-support tools helpful for combining expert knowledge with available empirical data [318]. It is now possible for both the structure and parameters of a BNN be learnt directly from a data set, and for this reason BNNs are being increasingly applied in a wide variety of medical domains where automated reasoning is required. The most popular expert system using BNN in pathology is PATHFINDER [319]. A good introduction to BNN and a brief history of its development is found in reference [320].

From the statistical point of view, BNNs are indispensable tools for dealing with high dimensional data problems as they allow a reduction in the complexity of the phenomenon under study by representing joint relationships between a set of variables through conditional relationships. Besides applying for making decisions, statisticians have also found another area of application in BNN [321-322], thanks to its unique hierarchical ordering structure. It is now possible to develop an automated data imputation method whose main goal is to preserve as much as possible the joint distribution specified in the BNN. This may provide great help to the data missing problems which pervade most EBM analyses. Most data imputation techniques are satisfactory only in the univariate scenario, but BNNs are able to preserve multivariate statistical relationships and logical constraints in the data (logical consistency) concurrently. As such, BNN is viewed as a consistent data imputation technique.

5.3.6 Data Mining

More recently, an alternative approach to statistics in constructing predictive models useful for probability encoding have been rapidly developed. The emerging field of data mining is a blend of statistics, artificial intelligence and database research [323]. To be specific, data mining is a technology that blends conventional data analysis methods with sophisticated algorithms for processing large volumes of data. It has also opened up new and exciting opportunities for data exploration and analysing old data in new ways. In fact, BNNs are recognized as a data mining tool by the relevant community.

The two disciplines of statistics and data mining have common aims in that both are concerned with discovering structure in data [324]. Most statisticians, however, are concerned with primary data analysis, that is, the data are collected with a particular question or a set of questions in mind [325]. On the other hand, data mining is entirely concerned with secondary data analysis, which aimed at finding unsuspected relationships that are of interest or value to the database owners [325]. Data mining techniques are able to cope with large data bases, which may contain contaminated information. Moreover, superabundance of data might render all statistical tests meaningless as they will lead to a significant result even with a minute effect. Very large data sets are unlikely to conform to the i.i.d assumption which underlies most statistical models. It is much more likely that some regions of the data space will be sampled more heavily than others at different times. Last but not least, very large data sets are likely to have been subjected to selection bias of various kinds [325] and may not conform to the statisticians' idealised modelling assumptions. The goal of applying data mining

techniques in biomedical research is to promote the optimum use of diverse data sets by enabling EBM researchers to make sound decisions.

As a growing area of research interest, the intersection and interaction between data mining and statistics is inevitable. Many useful data mining techniques are now available for dealing with the prediction problems [326-329], which are of paramount interest to medical decision analysts. Practical guides on solving data analytic problems with both disciplines can be found in references [329-332]. The applications of data mining will help to enrich the field of statistics.

5.4 A Final Word

The field of medicine has been significantly refined and restructured in recent years. Most of the new challenges are related to its direction and methodology. With emphasis on evidential utilisation, EBM opens the gate to further refined and more complete medical research. Accumulated from one generation to the next, the wealth of medical evidence contained in published journals will continue to grow with time. The immediate task is then to make good use of the existing evidences and apply them in decision making, while setting sight on validating these evidences under more stringent conditions and seeking new ones in unexplored fields. This sets the overall future direction for EBM practitioners.

The next challenge is to break new grounds in methodology relevant to EBM practice. As discussed earlier, methodology is required to establish how new knowledge may be gained. A methodology is a system of principles and general ways of organising and structuring theoretical and practical activities. Scientific progress is not limited to the accumulation of knowledge. It is also a process of evolving new means of seeking and acquiring knowledge. The Bayesian methodology, which captures the essence of knowledge acquisition [333], should therefore be seen as an invaluable asset to EBM.

The objectivity of the conventional scientific approach has been obtained by disregarding any prior knowledge about the phenomenon under investigation. However, some form of expert opinion can be quantified and applied in research. It may also provide an angle in which one may adopt for interpreting the collected evidence. While a clinician may remain as objective as possible, he is entitled to have a personal stance. In fact, clinical instincts and independent thinking—developed through experience and

coloured by personal values—should be valued as the prized attributes of a capable clinical decision-maker. It is also a waste of information if such expert opinions are ignored in analysis. Therefore, a competent methodologist should recognise the importance of such opinions and utilise them to the fullest in analysis. The current framework of EBM—devoted to downsize the malpractice of authoritarianism and consequently gives an utmost respect for evidence—places an extremely low value on expert opinion. Moreover, Bayesian analyses may be "objectified" with the use of non-informative priors. The Bayesian framework is thus strongly advocated for EBM practice as a unified framework for actions in the face of uncertainty.

Since its inception EBM seeks to revolutionise the organisation and structure to medical decisions. While EBM does not replace clinical skills and experience, it organises, expands and completes them. It also sets the path for future medical practice and calls for a systematic and integrated approach in searching for the relevant medical evidence for improving our current diagnosis and treatment. EBM reflects a probabilistic shift in today's paradigm of medicine dealing with a myriad of uncertainties. While some clinicians may consider EBM as evolutionary rather than revolutionary, no one can dispute the fact that it has brought about a new thinking in medicine.

The Bayesian framework promises to provide a new dimension to this revolution. The new millennium has already witnessed a burst of research activities in applying Bayesian methods to solve medical problems [14, 102-103, 113, 334-338]. Illuminating the present and pointing to the future, Bayesian models will continue to excite EBM researchers in all areas of research and decision making.

BIBLIOGRAPHY

1. Howard, R.A. (1988). Decision analysis: practice and promise. *Management Science*, 34, 679-695.

2. Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behaviour*. Princeton: Princeton University Press.

3. Stirling, W.C. (2003). *Satisficing games and decision making: with applications to engineering and computer science*. Cambridge: Cambridge University Press.

4. LaValle I.H. (1978). *Fundamentals of decision analysis*. New York: Holt, Rinehart and Winston.

5. Shachter, R. (1986). Evaluating influence diagrams. *Operations Research*, 34, 871-882.

6. Clement, R.T., & Reilly, T. (2001). *Making hard decisions with DecisionTools*. Australia: Duxbury.

7. Sox, H.C., Blatt, M.A., Higgins, M.C., & Marton, K.I. (1990). *Medical decision making*. Singapore: PG Publishing Pte Ltd.

8. Tversky, A., & Kahneman, D. (1974). Judgement under uncertainties: heuristics and biases. *Science*, 185(4157), 1124-1131.

9. Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47, 263-291.

10. Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453-458.

11. Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioural research*. Cambridge: Cambridge University Press.

12. Hogarth, R. (1987). *Judgement and choice*. New York: Wiley.

13. Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., & Richardson, W.S. (1996). Evidence based medicine: what is and what it isn't. *British Medical Journal*, 312(7023), 71-72.

14. Prevost, T.C., Abrams, K.R., & Jones, D.R. (2000). Hierarchical models in generalised synthesis of evidence: an example based on studies on breast cancer screening. *Statistics in Medicine*, 19, 3359-3376.

15. Paffenbarger, R.S., Hyde, R.T., Wing, A-L., & Hsieh, C-C. (1986). Physical activity, all cause mortality, and longevity of college alumni. *New England Journal of Medicine*, 314(10), 605-613.

16. Lissner, L., Bengtsson, C., Björkelund, C., & Wedel, H. (1996). Physical activity levels and change in relation to longevity: a prospective study of Swedish women. *American Journal of Epidemiology*, 143(1), 54-62.

17. Van Saase, J.L.C.M., Noteboom, W.M.P., & Vandenbroucke, J.P. (1990). Longevity of men capable of prolonged vigorous physical exercise: a 32 year follow up of 2259 participants in the Dutch eleven cities ice skating tour. *British Medical Journal*, 301(6790), 1409-1411.

18. Lifson, M.W. (1972). *Decision and risk analysis for practicing engineers*. Boston: Cahners Books.

19. Silverman, M.E., Murray, T.J., & Bryan, C.S. (2002). *The quotable Osler*. Philadelphia: American College of Physicians.

20. Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press.

21. Neyman, J. (1950). *First course in probability and statistics*. New York: Henry Holt.

22. Breiman, L. (1992). *Probability*. Philadelphia: Society for Industrial and Applied Mathematics.

23. Billingsley, P. (1995). *Probability and measure*. New York: Wiley.

24. Rosenthal, J.S. (2003). *A first look at rigorous probability theory*. Singapore: World Scientific.

25. McQueen, D. (2003). Strengthening the evidence base for health promotion. *Health Promotion International*, 16(3), 261-268.

26. Bayes, T. (2002). An essay towards solving a problem in the doctrine of chance. In: Swinburne, R., editor. *Bayes's Theorem*. Oxford: Oxford University Press.

27. Miller, D.W., & Miller, C.G. (2005). On evidence, medical and legal. *Journal of the American Physicians and Surgeons*, 10(3), 70-75.

28. Straus, S.E., & McAlister, F.A. (2000). Evidence-based medicine: a commentary on common criticisms. *Canadian Medical Association Journal*, 153(7), 837-841.

29. Tiwari, L., & Puliyel, J.M. (2004). Truth and evidence based medicine: spin is everything. *British Medical Journal*, 329(7473), 1043.

30. Silverman, W.A. (1999). *Where's the evidence—debates in modern medicine*. Oxford: Oxford University Press.

31. Cramér, H. (1963). *Mathematical methods of statistics*. Princeton: Princeton University Press.

32. Rao, C.R. (1973). *Linear statistical inference and its applications*. New York: Wiley.

33. Lehmann, E. (1986). *Testing statistical hypothesis*. New York: Wiley.

34. Neter, J., Kutner, M.H., Nachtsheim, C.J., & Wasserman, W. (1996). *Applied linear statistical models*. Boston: McGraw-Hill.

35. Gadamer, H.G. (2006). *Truth and methods*. London: Continuum.

36. Bontekoe, R. (1996). *Dimensions of the hermeneutic circle*. New Jersey: Humanities Press International.

37. Norton, J.D. (2007). Probability disassembled. *The British Journal of the Philosophy of Science*, 58(2), 141-171.

38. Nau, R.F. (2001). De Finetti was right: probability does not exist. *Theory and Decision*, 51(2-4), 89-124.

39. De Finetti, B. (1974). *Theory of probability*. New York: Wiley.

40. Ramsey, F. (1950). *Foundations: essays in philosophy, logic, mathematics and economics*. London: Routledge.

41. Savage, L.J. (1954). *The foundation of statistics*. New York: Wiley.

42. Anscombe, F.J., & Aumann, R.J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics*, 34, 199-205.

43. Kadane, J.B., & Winkler, R.L. (1988). Separating probability elicitation from utilities. *Journal of the American Statistical Association*, 83, 357-363.

44. Chaloner, K. (1996). Elicitation of prior distributions. In: Berry, D.A., & Stangl, D.K., editors. *Bayesian Biostatistics*, 141-156. New York: Marcel Dekker.

45. Kadane, J.B., & Wolfson, L.J. (1996). Priors for the design and analysis of clinical trials. In: Berry, D.A., & Stangl, D.K., editors. *Bayesian Biostatistics*, 157-184. New York: Marcel Dekker.

46. Kadane, J.B., & Wolfson, L.J. (1998). Experiences in elicitation. *The Statistician*, 47(1), 3-19.

47. Karni, E., & Safra, Z. (1995). The impossibility of experimental elicitation of subjective probabilities. *Theory and Decision*, 38(3), 313-320.

48. O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *The Statistician*, 47(1), 21-35.

49. Huber, F. (2005). Subjective probabilities as basis for scientific reasoning? *Theory and Decision*, 56(1), 101-116.

50. Dongen, S.V. (2006). Prior specification in Bayesian statistics: three cautionary tales. *Journal of Theoretical Biology*, 242(1), 90-100.

51. Chan, S-P., & Poh, K-L. (2005). Application of Bayesian linear regression in biomedical research. *Singapore General Hospital Proceedings*, 13(3), 154-161.

52. Laplace, P. (1812). *Théorie analytique des probabilitiés*. Courcier, 387.

53. Williamson, J. (2007). Inductive influence. *Theory and Decision*, 58(4), 689-708.

54. Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, 1(3), 385-402

55. Bayarri, M.J., & Berger, J. (2004). The interplay between Bayesian and frequentist analysis. *Statistical Science*, 19, 58-80.

56. Jeffreys, H. (1961). *Theory of probability*. London: Oxford University Press.

57. Jaynes, E.T. (2003). *Probability theory: the logic of science*. Cambridge: Cambridge University Press.

58. Fienberg, S.E. (2006). When did Bayesian inference becomes "Bayesian"? *Bayesian Analysis*, 1(1), 1-40.

59. Goldstein, M. (2006). Subjective Bayesian analysis: principles and practice. *Bayesian Analysis*, 1(3), 403-420.

60. Christen, J.A. (2006). Stop using "subjective" to refer to Bayesian analysis. *Bayesian Analysis*, 1(3), 421-422.

61. Draper, D. (2006). Coherence and calibration: comments on subjectivity and "objectivity". *Bayesian Analysis*, 1(3), 423-428.

62. Feinberg, S.E. (2006). Does it make sense to be an "objective Bayesian"? *Bayesian Analysis*, 1(3), 429-432.

63. Kadane, J.B. (2006). Is "objective Bayesian analysis" objective, Bayesian, or wise? *Bayesian Analysis*, 1(3), 433-436.

64. Kass, R.E. (2006). Kinds of Bayesians. *Bayesian Analysis*, 1(3), 437-440.

65. Lad, F. (2006). Objective Bayesian statistics…Do you buy it? Should we sell it? *Bayesian Analysis*, 1(3), 441-444.

66. O'Hagan, A. (2006). Science, subjectivity and software. *Bayesian Analysis*, 1(3), 445-450.

67. Wasserman, L. (2006). Frequentist Bayes is objective. *Bayesian Analysis*, 1(3), 451-456.

68. Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika*, 74, 604-608.

69. O'Hagan, A. (1994). *Kendall's advanced theory of statistics, Vol. 2B. Bayesian inference*. London: Arnold.

70. Hampel, C.G. (1945). Studies in the logic of confirmation. *Mind*, 54, 1-26.

71. Good, I.J. (1960). The paradox of confirmation. *The British Journal of the Philosophy of Science*, 11, 145-149.

72. Vranas, P.B.M. (2004). Hempel's raven paradox: a lacuna in the standard Bayesian solution. *The British Journal of the Philosophy of Science*, 55, 545-560.

73. Dowe, D.L., Gardner, S., & Oppy, G. (2007). Bayes not bust! Why simplicity is no problem for bayesians. *The British Journal of the Philosophy of Science*, 58(4), 709-754.

74. Bernado, J.M. (1997). Statistical inference as a decision problem: the choice of sample size. *The Statistician*, 46(2), 151-154.

75. Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *The Statistician*, 46(2), 185-192.

76. Joseph, L., & Bélisle, P. (1997). Bayesian sample size determination for normal means and difference between normal means. *The Statistician*, 46(2), 209-226.

77. Rahme, E., Jospeh, L., & Gyorkos, T.W. (2000). Bayesian sample size determination for estimating binomial parameters from data subject to misclassification. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 49(1), 119-128.

78. Wiper, M.P., & Pettit, L.I. (1994). Bayesian estimation of the binomial parameter n. *Applied Statistics*, 43(1), 233-236.

79. Jeevanand, E.S. (1997). Bayes estimation of $P(X_2<X_1)$ for a bivariate Pareto distribution. *The Statistician*, 46(1), 93-99.

80. Bolstad, W.M. (2004). *Introduction to Bayesian statistics*. New Jersey: Wiley.

81. Box, G., & Tiao, G. (1992). *Bayesian Inference in statistical analysis*. New York: Wiley.

82. Basu, S., Banerjee, M., & Sen, A. (2000). Bayesian inference for Kappa from single and multiple studies. *Biometrics*, 56(2), 577-582.

83. Bernado, J.M., & Perez, S. (2007). Comparing normal means: new methods for an old problem. *Bayesian Analysis*, 2(1), 45-58.

84. Bernardo, J.M., & Smith, A.F.M. (1994). *Bayesian theory*. Chichester: Wiley.

85. Kadane, J.B. (1996). *Bayesian methods and ethics in a clinical trial design*. New York: Wiley.

86. Johnson, V.E. (2007). Bayesian model assessment using pivotal quantities. *Bayesian Analysis*, 2(4), 719-734.

87. Bertolino, F., Racugno, W., & Moreno, E. (2000). Bayesian model selection approach analysis of variance under heteroscedasticity. *Journal of the Royal Statistical Society Series D (The Statistician)*, 49(4), 503-517.

88. Maddala, G.S. (1977). *Econometrics*. Singapore: McGraw-Hill.

89. Koop, G. (2003). *Bayesian econometrics*. Chichester: Wiley…

90. Press, S.J. (1989). *Bayesian statistics: principles, models, and applications*. New York: Wiley.

91. McCulloch, P., & Nelder, J.A. (1989). *Generalized linear models*. London: Chapman & Hall, 2nd ed.

92. Dipak, D.K., Ghosh, S.K., & Mallick, B.K. (2000). *Generalized linear models: a Bayesian perspective*. New York: Marcel Dekker, Inc.

93. Dey, D.K., & Rao, C.R. (2005). *Bayesian thinking: modelling and computation*. The Netherlands: Elsevier.

94. Congdon, P. (2001). *Bayesian statistical modelling*. Chichester: Wiley.

95. Berry D.A., & Stangl, D.K. (1996). *Bayesian biostatistics*. New York: Marcel Dekker.

96. Gill, J. (2002). *Bayesian methods: a social and behavioural sciences approach*. Boca Raton: Chapman & Hall.

97. Hosmer, D.W., & Lemeshow, S. (1999). *Applied survival analysis: regression modelling of time to event data*. New York: Wiley.

98. Collette, D. (2003). *Modelling survival data in medical research*. London: Chapman & Hall.

99. Cleves, M.A., Gold, W.W., & Gutierrez, R.G. (2002). *An introduction to survival analysis using Stata*. Texas: Stata Press.

100. Cox, D. (1972). Regression models and life tables. *Journal of the Royal Statistical Society Series B*, 34, 187-220.

101. Ibrahim, J.G. Chen, M-H., & Sinha, D. (2001). *Bayesian survival analysis*. New York: Springer Verlag.

102. Volinsky, C.T., Madigan, D., Raftery, A.E., & Kronmal, R.A. (1997). Bayesian model averaging in proportional hazard models: assessing the risk of a stroke. *Applied Statistics*, 46(4), 433-448.

103. Soliman, A.A. (2000). Bayes prediction in a Pareto lifetime model with random sample size. *Journal of the Royal Statistical Society Series D (The Statistician)*, 49(1), 51-62

104. Sinha, D., & Dey, D.K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 92, 1195-1212.

105. Lindsay, J.K. (1993). *Models for repeated measurements*. Oxford: Oxford University Press.

106. Hardin, J.W., & Hilbe, J.M. (2001). *Generalized linear models and extensions*. Texas: Stata Press.

107. McCulloch, C.E., & Searle, R.S. (2001). *Generalized, linear and mixed models*. New York: Wiley.

108. Hardin, J.W., & Hilbe, J.M. (2003). *Generalized estimating equations*. Boca Raton: Chapman & Hall/CRC Press.

109. Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

110. Rossi, P.E., Allenby, G.M., & McCulloch, R. (2005). *Bayesian statistics and marketing*. Chichester: Wiley.

111. Hobert, J., & Casella, G. (1996). The effect of improper priors on Gibb sampling in hierarchical models. *Journal of the American Statistical Association*, 91, 1461-1473.

112. Lenk, P., & DeSarbo, W. (2000). Bayesian inference for finite mixtures of generalised linear models with random effects. *Psychometrika*, 65, 93-119.

113. Dukić, V., & Dignam, J. (2007). Bayesian hierarchical multiresolution hazard model for the study of time-dependent failure patterns in early stage breast cancer. *Bayesian Analysis*, 2(3), 591-610.

114. Sutton, A.J., Jones, D.R., Abrams, K.R., Sheldon, T.A., & Song, F. (2000). *Methods for Meta-analysis in medical research.* London: Wiley.

115. DerSimonian, R., & Laird, N. (1986). Meta analysis in clinical trials. *Controlled Clinical Trials,* 7, 177-188.

116. Thompson, S. (1994). Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal,* 309(6965), 1351-1355.

117. Smith, G., Egger, M., & Phillips, A. (1997). Meta-analysis: beyond the grand mean? *British Medical Journal*, 315(7122), 1610-1614.

118. Stern, J.M., & Simes, R.J. (1997). Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *British Medical Journal*, 315(7109), 640-645.

119. Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London: Academic Press.

120. Thompson, S., & Higgins, R. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21, 1559-1574.

121. Goldstein, H., Yang, M., Omar, R., Turner, R., & Thompson, S. (2000). Meta-analysis using multi-level models with an application to the study of class size. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 49(3), 399-412.

122. Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.

123. Casella, G., & George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46(3), 167-174.

124. Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22, 1701-1762.

125. Bremaud, P. (1999). *Markov chain: Gibbs fields, monte carlo simulation, and queues*. New York: Springer Verlag.

126. Chen, M-H., Shao, Q-M., & Ibrahim, J.G. (2000). *Monte Carlo methods in Bayesian computation*. New York: Springer Verlag.

127. Liu, J. (2001). *Monte Carlo strategies in scientific computing*. New York: Verlag Springer.

128. Brooks, S.P. (1998). Markov chain Monte Carlo method and its application. *The Statistician*, 47(1), 69-100.

129. Diaconis, P., & Ylvisaker, D. (1979). Conjugate priors for exponential families. *The Annals of Statistics*, 17, 269-281.

130. R.A. Fisher (1921). On the probably error of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3-32.

131. R.A. Fisher (1922). On the mathematical foundation of theoretical statistics. *Philosophical Transactions of the Royal Society*, 222, 309-368.

132. Albert, J., & Chib, S. (1993). Bayesian regression analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88, 657-667.

133. Hosmer, D.W., & Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley.

134. Von Mises, R. (1947). On the asymptotic distribution of differentiable statistical function. *Annals of Mathematical Statistics*, 18, 309-348.

135. Papathanasiou, V. (1993). Some characteristic properties of the Fisher information matrix type inequalities. *Journal of Multivariate Analysis*, 14, 256-265.

136. Rao, B.L.S.P. (1978). Rate of convergence of Berstein-von Mises approximation process. *Serdica*, 4, 36-42.

137. Das, S., & Dey, D.K. (2006). On Bayesian analysis of generalised linear models using the Jacobian technique. *The American Statistician*, 60(3), 36-42.

138. Cameron, A.C., & Trivedi, P.K. (1998). *Regression analysis of count data.* New York: Cambridge University Press.

139. Hilbe, J.M. (2007). *Negative binomial regression*. Cambridge: Cambridge University Press.

140. Geyer, C.J. (1992). Practical Markov chain Monte Carlo. *Statistical Science*, 7(4), 473-511.

141. Kass, R.E., Carlin, B.P., Gelman, A., & Neal, R.M. (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, 52(2), 93-100.

142. Zellner, A., & Min, C-K. (1995). Gibbs sampler: convergence criteria. *Journal of the American Statistical Association*, 90, 921-927.

143. Serfling, R.J. (1980). *Approximation theorems of mathematical statistics*. 143-149. Singapore: Wiley.

144. Carlin, B.P., & Louis, T.A. (2000). *Bayes and empirical Bayes methods for data analysis*. New York: Chapman & Hall.

145. Huber, P.J. (1981). *Robust statistics*. New York: Wiley.

146. Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information & Computation*, 121, 256-285.

147. Metz, C.E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4), 283-98.

148. Hanley, J.A., & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.

149. Hanley, J.A., & McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148(3), 839-43.

150. Sampson, S.M. (2001). Treating depression with selective serotonin reuptake inhibitors: a practical approach. *Mayo Clinic Proceedings*, 76, 739-744.

151. Barbey, J.T., & Roose, S.P. (1998). SSRI safety in overdose. *Journal of Clinical Psychiatry*, 59(Supplement 15), 42-48.

152. Cheeta, S., Schifano, F., Oyefeso, A., Webb, L., & Ghodse, A.H. (2004). Antidepressant-related deaths and antidepressant prescriptions in England and Wales, 1998-2000. *British Journal of Psychiatry*, 184, 41-47.

153. Goldstein, B.J., & Goodnick, P.J. (1998). Selective serotonin reuptake inhibitors in the treatment of affective disorders-III. Tolerability, safety and pharmacoeconomics. *Journal of Psychopharmacology*, 12 (3 Supplement B), S55-87.

154. Pacher, P., Ungvari, Z., Nanasi, P.P., Furst, S., & Kecskemeti, V. (1999). Speculations on difference between tricyclic and selective serotonin reuptake inhibitor antidepressants on their cardiac effects. Is there any? *Current Medical Chemistry*, 6(6), 469-480.

155. Stokes, P.E. (1993). Fluoxetine: a five-year review. *Clinical Therapy*, 15(2), 216-243.

156. Zohar, J., & Westenberg, H.G. (2000). Anxiety disorders: a review of tricyclic antidepressants and selective serotonin reuptake inhibitors. *Acta Psychiatrica Scandinavica Supplement*, 403, 39-49.

157. Katona, C. (2000). Managing depression and anxiety in the elderly patient. *European Neuropsychopharmacology*, 10(Supplement 4), S427-432.

158. Baldwin, D.S. (2001). Unmet needs in the pharmacological management of depression. *Human Psychopharmacology*, 16(Supplement 2), S93-99.

159. Wagstaff, A.J., Cheer, S.M., Matheson, A.J., Ormrod, D., & Goa, K.L. (2002). Paroxetine: an update of its use in psychiatric disorders in adults. *Drugs*, 62(4), 655-703.

160. Vaswani, M., Linda, F.K., & Ramesh, S. (2003). Role of selective serotonin reuptake inhibitors in psychiatric disorders: a comprehensive review. *Progress in Neuropsychopharmacology & Biological Psychiatry*, 27(1), 85-102.

161. Lader, M.H. (1996). Tolerability and safety: essentials in antidepressant pharmacotheraphy. *Journal of Clinical Psychiatry*, 57(Supplement 2), 39-44.

162. Steffans, D.C., Krishnan, K.R., & Helms, M.J. (1997). Are SSRIs better than TCAs? Comparison of SSRIs and TCAs: a meta-analysis. *Depression & Anxiety*, 6(1), 10-18.

163. Thompson, C., & Thompson, C.M. (1989). The prescribing of antidepressants in general practice II: a placebo-controlled trial of low-dose dothiepin. *Human Psychopharmacology*, 4, 191-204.

164. Kernick, D.P. (1997). Which antidepressant? A commentary from general practice on evidence-based medicine and health economics. *British Journal of General Practice*, 47(415), 95-98.

165. Kyle, C.J., Petersen, H.E., & Overo, K.F. (1998). Comparison of the tolerability and efficacy of citalopram and amitriptyline in elderly depressed patients treated in general practice. *Depression & Anxiety*, 8(4), 147-153.

166. Mahendru, R.K., & Mahendru, S. (2001). Selection of antidepressant drugs in general practice. *Journal of Indian Medical Association*, 99(1), 54-55.

167. Montgomery, S.A., Henry, J., McDonald, G., Dinan, T., Lader, M., Hindmarch, I., Clare, A., & Nutt, D. (1994). Selective serontonin reuptake inhibitors: meta-analysis of discontinuation rates. *International Clinical Psychopharmacology*, 9(1), 47-53.

168. Montgomery, S.A., & Kasper, S. (1995). Comparison of compliance between serotonin reuptake inhibitors and tricyclic antidepressants: a meta-analysis. *International Clinical Psychopharmacology*, 9(Supplement 4), 33-40.

169. Anderson, I.M. (2000). Selective serotonin reuptake inhibitors versus tricyclic antidepressants: a meta-analysis of efficacy and tolerability. *Journal of Affecive Disorders*, 58(1), 19-36.

170. MacGillivray, S., Arroll, B., Hatcher, S., Ogston, S., Reid, I., Sullivan, F., William, B., & Crombie, I. (2003). Efficacy and tolerability of selective serotonin reuptake inhibitors compared with tricyclic antidepressants in depression treated in primary care: a systematic review and meta-analysis. *British Medical Journal*, 326(7397), 1014.

171. Yildiz, A., Pauler, D.K., & Sachs, G.S. (2004). Rates of study completion with single versus split daily dosing of antidepressants: a meta-analysis. *Journal of Affective Disorders*, 78(2), 157-162.

172. Barbui, C., Hotopf, M., Freemantle, N., Boynton, J., Churchill, R., & Eccles, M.P. (2004). Treatment discontinuation with selective serotonin reuptake inhibitors (SSRIs) versus tricyclic antidepressants (TCAs). *Cochrane Database of Systematic Reviews*, 2.

173. Yusuf, S., Peto, R., Lewis, J., Collins, R., & Slieght, P. (1985). Beta blockage during and after myocardial infarction: an overview of the randomised trials. *Progress in Cardiovascular Diseases*, 27(5), 335-371.

174. Corne, S.J., & Hall, J.R. (1989). A double-blind comparative study of fluoxetine and dothiepin in the treatment of depression in general practice. *International Clinical Psychopharmacology*, 4(3), 245-254.

175. Stott, P.C., Blagden, M.D., & Aitken, C.A. (1993). Depression and associated anxiety in primary care: a double-blind comparison of paroxetine and amitriptyline. *European Neuropsychopharmacology*, 3, 324-325.

176. Rosenberg, C., Damsbo, N., Fuglum, E., Jacobsen, L.V., & Horsgard, S. (1994). Citalopram and imipramine in the treatment of depressive patients in general practice. A Nordic multicentre clinical study. *International Clinical Psychopharmacology*, 9(Supplement 1), 41-48.

177. Doogan, D.P., & Langdon, C.J. (1994). A double-blind, placebo-controlled comparison of sertraline and dothiepin in the treatment of major depression in general practice. *International Clinical Psychopharmacology*; 9(2): 95-100.

178. Moon, C.A.L., & Vince, M. (1996). Treatment of major depression in general practice: a double-blind comparison of paroxetine and lofepramine. *British Journal of Clinical Practice*, 50(5), 240-244.

179. Christiansen, P.E., Behnke, K., Black, C.H., Ohrstrom, J.K., Bork-Rasmussen, H., & Nilsson, J. (1996). Paroxetine and amitriptyline in the treatment of depression in general practice. *Acta Psychiatrica Scandinavica*, 93(3), 158-163.

180. Ravindran, A.V., Judge, R., Hunter, B.N., Bray, J., & Morton, N.H. (1997). A double-blind, multicentre study in primary care comparing paroxetine and clomipramine in patients with depression and associated anxiety. Paroxetine study group. *Journal of Clinical Psychiatry*, 58(3), 112-118.

181. Thompson, C. (1991). Sertraline in a primary care setting. In: Racagni NB, Fukuda T, editors. *Biological Psychiatry*. Amsterdam: Elservier, 863-865.

182. Hutchinson, D.R., Tong, S., Moon, C.A.L., Vince, M., & Clarke, A. (1992). Paroxetine in the treatment of elderly depressed patients in general practice: a double-blind comparison with amitriptyline. *International Clinical Psychopharmacology*, 6(4 Supplement), 43-51.

183. Rosner, B. (2000). *Fundamentals of biostatistics*. Pacific Grove: Brooks/Cole.

184. Swinkels, J.A., De Jonghe, F. (1995). Safety of antidepressants. *International Clinical Psychopharmacology*, 9(Supplement 4), 19-25.

185. Kasper, S., Hoflich, G., Scholl, H.P., & Moller, H.J. (1994). Safety and antidepressant efficacy of selective serotonin re-uptake inhibitors. *Human Psychopharmacology*, 9, 1-12.

186. Glass, G.V. (1976). Primary, secondary and meta-analysis of research. *Educational Research*, 5, 3-8.

187. Egger, M.G., (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 314(7109), 629-634.

188. Fleiss, J.L. (1993). The statistical basis of meta-analysis. *Statistical Methods in Medical Research*, 2, 121-145.

189. Wong, M-K., Arjandas, Ching, L-K., Lim, S-L., & Lo, N-N. (2002). Osteoporotic hip fractures in Singapore: costs and patient's outcome. *Annals of Academy of Medicine Singapore*, 31(1), 3-7.

190. Cooper, C., Campian, G., & Melton, L.J. (1992). Hip fracture in the elderly: a worldwide projection. *Osteoporosis International*, 2, 285-289.

191. Koh, L.K-H., Saw, S-M., Lee, J.J-M., Leong, K-H., & Lee, J. (2001). Hip fracture incidence rates in Singapore 1991-1998. *Osteoporosis International*, 12(4), 311-318.

192. Scientific Advisory Board, Osteoporosis Society of Canada. (1996). Clinical practice guidelines for the diagnosis and management of osteoporosis. *Canadian Medical Association Journal*, 155, 1113-1133.

193. Council of the National Osteoporosis Foundation. (1996). Guidelines for the early detection of osteoporosis and prediction of fracture risk. *South African Medical Journal*, 86(9), 1113-1116.

194. Lydick, E., Cook, K., Turpin, J., Melton, M., Stine, R., & Byrnes, C. (1998). Development and validation of a simple questionnaire to facilitate identification of women likely to have low bone density. *American Journal of Managed Care*, 4(1), 37-48.

195. Cadarette, S.M., Jaglal, S.B., Kreiger, N., McIssac ,W.J., Darlington, G.A., & Tu, J.V. (2000). Development and validation of the osteoporosis risk assessment instrument to facilitate selection of women for bone densitometry. *Canadian Medical Association Journal*, 162(9), 1289-1294.

196. Weinstein, L., & Ullery, B. (2000). Identification of at-risk women for osteoporosis screening. *American Journal of Obstetrics Gynecology*, 183(3), 547-549.

197. Michaelsson, K., Bergstrom, R., Mallmin, H., Holmberg, L., Wolk, A., & Ljunghall, S. (1996). Screening for osteoporosis: selection by body composition. *Osteoporosis International*, 6(2), 120-126.

198. Koh, L.K-H., Sedrine, W.B., Torralba, T.P., Kung, A., Fujiwara, S., Chan, S-P., Huang, Q.R., Rajatanavin, R., Tsai, K-S., Park, H-M., & Reginster, J.Y. (2001). A

simple tool to identify Asian women at increased risk of osteoporosis. *Osteoporosis International*, 12(8), 699-705.

199. Marshall, D., Johnell, O., & Wedel, H. (1996). Meta-analysis of how well measure of bone mineral density predicts occurrence of osteoporortic fractures. *British Medical Journal*, 312(7041), 1254-1259.

200. Scheaffer, R.L., Mendenhall, W., & Ott, L. (1990). *Elementary Survey Sampling*. 118-119. Belmont: Duxbury.

201. Ministry of Health (2000). *National Health Survey 1998*. Ministry of Health, Singapore.

202. Food and Nutrition Department (1994). *Food Consumption Study 1993*. Ministry of Health, Singapore.

203. Park, H-M., Sedrine, W.B., Reginster, J.Y., Ross, P.D., & OSTA (2003). Korean experience with the OSTA risk index for osteoporosis: a validation study. *Journal of Clinical Densitometry*, 6(3), 247-250.

204. Kung, A-W., Ho, A-Y., Sedrine, W.B., Reginster, J.Y., & Ross, P.D. (2003). Comparison of a simple clinical risk index and quantitative bone ultrasound for identifying women at increased risk of osteoporosis. *Osteoporosis International*, 14(9), 716-721.

205. Ministry of Health, National Medical Research Council, Osteoporosis Society of Singapore (2002). *Clinical Practice Guideline: Osteoporosis*. Ministry of Health, Singapore.

206. Altman, D.G. (1999). *Practical statistics for medical research*. 417-418. Boca Raton: Chapman & Hall.

207. Giardiello, F.M., Hamilton, S.R., Krush, A.J., Piantadosi, S., Hylind, L.M., Celano, P., Booker, S.V., Robinson, CR, & Offerhaus, G.J. (1993). Treatment of colonic and rectal adenomas with sulindac in familial adenomatous polyposis. *New England Journal of Medicine*, 328(18), 1313-1316.

208. Giardiello, F.M., Yang, V.W., Hylind, L.M., Krush, A.J., Petersen, G.M., Trimbath, J.D., Piantadosi, S., Garrett, E., Geiman, D.E., Hubbard, W., Offerhaus, G.J., &. Hamilton, S.R. (2002). Primary chemoprevention of familial adenomatous polyposis with sulindac. *New England Journal of Medicine*, 346(14), 1054-1059.

209. Pasricha, P.J., Bedi, A., O'Connor, K., Rashid, A., Akhtar, A.J., Zahurak, M.L., Piantadosi, S., Hamilton, S.R., & Giardiello, F.M. (1995). The effects of sulindac on colorectal proliferation and apoptosis in familial adenomatous polyposis. *Gastroenterology*, 109(3), 994-998.

210. Piantadosi, S. (1997). *Clinical trials: a methodological perspective*. 479-481. New York: Wiley, 479-481.

211. Ladenheim, J., Garcia, G., Titzer, D., Herzenburg, H., Lavori, P., Edson, P., & Omary, M.B. (1995). Effect of sulindac on sporadic colonic polyps. *Gastroenterology*, 108(4), 1083-1087.

212. Giardiello, F.M., Offerhaus, J.A., Tersmette, A.C., Hylind, L.M., Krush, A.J., Brensinger, J.D., Booker, S.V., & Hamilton, S.R. (1996). Sulindac induced regression of colorectal adenomas in familial adenomatous polyposis: evaluation of predictive factors. *Gut*, 38(4), 578-581.

213. Halstead, S.B. (2002). Dengue. *Current Opinion of Infectious Disease*, 15(5), 471-476.

214. Gibbons, R.V., & Vaughn, D.W. (2000). Dengue: an escalating problem. *British Medical Journal*, 324(7353), 1563-1566.

215. Chow, A., Ye, T., & Ang, L-W. (2005). Dengue epidemiological update. Ministry of Health Information Paper 2005.

216. Chia, A., Luu, C-D., Mathur, R., Cheng, B., & Chee, S-P. (2006). Electrophysiological findings in patients with dengue-related maculopathy. *Archives of Ophthalmology*, 124(10), 1421-1426.

217. Chan, D.P., Teoh, S.C., Tan, C.S., & The Eye Institute Dengue-Related Ophthalmic Complications Workgroup (2006). Ophthalmic complications of dengue. *Emerging Infectious Diseases*, 12(2), 285-289.

218. Chlebicki, M.P., Ang, B., Barkham, T., & Laude, A. (2005). Retinal hemorrhages in 4 patients with dengue fever. *Emerging Infectious Diseases*, 11(5), 770-772.

219. Madsen, P.L., & Thybo, S. (2005). Ocular complications of dengue fever. *Ugeskr Laeger*, 167(43), 4083-4084.

220. Nainiwal, S., Garg, S.P., Prakash, G., & Nainiwal, N. (2005). Bilateral vitreous haemorrhage associated with dengue fever. *Eye*, 19(9), 1012-1013.

221. Siqueira, R.C., Vitral, N.P., Campos, W.R., Orefice, F., & De Moraes Figueiredo, L.T. (2004). Ocular manifestations in dengue fever. *Ocul Immunoogy & Inflammation*, 12(4), 323-327.

222. Lim, W-K., Mathur, R., Koh, A., Yeo, R., & Chee, S-P. (2004). Ocular manifestations of dengue fever. *Ophthalmology*, 111(11), 2057-2064.

223. Cruz-Villegas, V., Berrocal, A.M., Davis, J.L. (2003). Bilateral choroidal effusions associated with dengue fever. *Retina*, 23(4), 576-578.

224. Haritoglou, C., Dotse, S.D., Rudolph, G., Stephan, C.M., Thurau, S.R., & Klauss, V. (2002). A tourist with dengue fever and visual loss. *The Lancet*, 360(9339), 1070.

225. Haritoglou, C., Scholz, F., Bialasiewicz, A., & Lauss, V. (2000). Ocular manifestation in dengue fever. *Ophthalmologe*, 97(6), 433-436.

226. Wen, K-H., Sheu, M-M., Chung, C-B., Wang, H-Z., & Chen, C-W. (1989). The ocular fundus findings in dengue fever. *Gaoxiong Yi Xue Ke Xue Za Zhi (高雄医科杂志)*, 5(1), 24-30.

227. Spitznas, M. (1978). Macular hemorrhage in dengue fever. *Klinische Monatsblatter fur Augenheilkunde*, 172, 105-107.

228. Deutman, A.F., & Bos, P.J. (1979). Macular bleeding in dengue fever. *Klinische Monatsblatter fur Augenheilkunde*, 175(3), 429.

229. Singapore Ministry of Health (2004). *A guide on infectious diseases of public health importance in Singapore*. 6th ed.

230. Kurane, I, Innis, B.L., Nimmannitya, S., Nisalak, A., Meager, A., Janus, J., & Ennis, F.A. (1991). Activation of T lymphocytes in dengue virus infections. High levels of soluble interleukin 2 receptor, soluble CD4, soluble CD8, interleukin2 and interferon-γ in sera of children with dengue. *Journal of Clinical Investigation*, 88(5), 1473-1480.

231. Kurane, I., & Ennis, F.A. (1992). Immunity and immunopathology in dengue virus infections. *Seminars in Immunology*, 4(2), 121-127.

232. Venketasubramanian, N. (1999). Stroke in Singapore—an overview. *Singapore Medical Journal*, 40(1), 1-7.

233. Ariesen, M.J., Algra, A., Van der Worp, H.B., & Rinkel, G.J.E. (2005). Applicability and relevance of models that predict short term outcome after intracerebral haemorrhage. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(6), 839-844.

234. Nilsson, O.G., Lindgren, A., Brandt, L., & Saveland, H. (2002). Prediction of death in patients with primary intracerebral haemorrhage: a prospective study of a defined population. *Journal of Neurosurgery*, 97(3), 531-536.

235. Wood, S.N. (2006). *Generalized additive model: an introduction with R*. Boca Raton: CRC Press.

236. Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1998). *Classification and regression trees*. Boca Raton: Chapman & Hall.

237. Chan, K-Y., & Low, W-Y. (2004). LOTUS: an algorithm for building accurate and comprehensive logistic regression trees. *Journal of Computer Graphical Statistics*, 13, 826-852.

238. Siegel, S., & Castellan, N.J. (2000). *Nonparametric statistics for the behavioural science*. Singapore: McGraw-Hill.

239. Teasdale, G., & Jennett, B. (1974). Assessment of coma and impaired consciousness: a practical scale. *The Lancet*, 2(7872), 81-84.

240. Lemeshow, S., & Hosmer, D.W. (1982). A review of goodness-of-fit statistics for use in the development of logistic regression model. *American Journal of Epidemiology*, 115(1), 92-106.

241. Flegal, K.M., Graubard, B.I., Williamson, D.F., & Gail, M.H. (2005). Excess deaths associated with underweight, overweight and obesity. *Journal of American Medical Association*, 293(15), 1861-1867.

242. Thom, E. (2000). A randomised, double-blind, placebo-controlled trial of a new weight-reducing agent of natural origin. *The Journal of Medical Research*, 28, 229-233.

243. Pedersen, J.O., Zimmermann, E., Stallknecht, B.M., Brunn, J.M., Kroustrup, J.P., Larsen, J.F., & Helge, J.W. (2006). Lifestyle intervention in the treatment of severe obesity. *Ugeskr Laeger*, 168(2), 167-172.

244. Wadden, T.A., & Foster, G.D. (2000). Behavioural treatment of obesity. *Medical Clinics of North America*, 84, 441-462.

245. Kramer, F.M., Jeffrey, R.W., Forster, J.L., & Snell, M. K. (1989). Long term follow-up of behavioural treatment for obesity: patterns of weight regain in men and women. *International Journal of Obesity*, 13(2), 123-136.

246. Phelan, S., & Wadden, T.A. (2002). Combining behavioural and pharmacological treatments for obesity. *Obesity Research*, 10(6), 560-574.

247. Hensrud, D.D., Weinsier, B.E., Darnell, B.E., & Hunter, G.R. (1995). *Relationship of co-morbidities of obesity to weight loss and four-year weight maintenance/rebound*. Obesity research, 3(Supplements 2), s217-s222.

248. Fox, K.M., & EURopean trial ON reduction of cardiovascular events with Perindopril in stable coronary Artery disease Investigators. (2003). Efficacy of perindopril in

reduction of cardiovascular events among patients with stable coronary artery disease: randomised, double-blind, placebo-controlled, multicentre trial (the EUROPA study). *The Lancet*, 362(9386), 782-788.

249. Domanski, M., Krause-Steinrauf, H., Deedwania, P., Follmann, D., Ghali, J.K., Gilbert, E., Haffner, S., Katz, R., Lindenfeld, J., Lowes, B.D., Martin, W., McGrew, F., Bristow, M.R., & BEST investigators. (2003). The effect of diabetes on outcomes of patients with advanced heart failure in the BEST trial. *Journal of American College of Cardiology*, 42(5), 914-922.

250. De Groote, P., Lamblin, N., Mouquet, F., Plichon, D., McFadden, E., Van Belle, E., & Bauters, C. (2004). Impact of diabetes mellitus on long-term survival in patients with congestive heart failure. *European Heart Journal*, 25(8), 656-662.

251. Gustafsson, I., Brendorp, B., Seibaek, M., Burchardt, H., Hildebrandt, P., Kober, L., & Torp-Pedersen, C. (2004). Influence of diabetes and diabetes-gender interaction on the risk of death in patients hospitalized with congestive heart failure. *Journal of American College of Cardiology*, 43(5), 771-777.

252. Hjalmarson, A, Goldstein, S., & Fagerberg, B. (2000). Effects of controlled-release metoprolol on total mortality, hospitalisation, and well-being in patients with heart failure: the Metoprolol CR/XL randomised intervention trial in congestive heart failure (MERIT-HF). *Journal of the American Medical Association*, 283(10), 1295-1302.

253. Brophy, J.M., Joseph, L., & Rouleau, J.L. (2001). β–blockers in congestive heart failure. *Annals of Internal Medicine*, 134, 550-560.

254. McAlister, F.A., Ezekowitz, J., Tonelli, M., & Armstrong, P.W. (2004). Renal insufficiency and heart failure: prognostic and therapeutic implications from a prospective cohort study. *Journal of the American Heart Association*, 109, 1004-1009.

255. Mak, K-H., Chia, K-S., Kark, J-D., Chua, T., Tan, C., Foong, B-H., Lim, Y-L., & Chew, S-K. (2003). Ethnic differences in acute myocardial infarction in Singapore. *European Heart Journal*, 24(2), 151-160.

256. Kamalesh, M., Subramanian, U., Saweda, S., Eckert, G., Temkit, M'H., & Tierney, W. (2006). Decreased survival in diabetic patients with heart failure due to systolic dysfunction. *The European Journal of Heart Failure*, 8(4), 404-408.

257. Varela-Roman, A., Shamagian, L.G., Caballero, E.B., Ramos, P.M., Veloso, P.R., & Gonzalez-Juanatey, J.R. (2005). Influence of diabetes on the survival of patients hospitalised with heart failure: a 12-year study. *The European Journal of Heart Failure*, 7(5), 859-864.

258. Das, S.R., Drazner, M.H., Yancy, C.W., Stevenson, L.W., Gersh, B.J., & Dries, D.L. (2004). Effects of diabetes mellitus and ischemic heart disease on the progression from asymptomatic left ventricular dysfunction to symptomatic heart failure: a retrospective analysis from the Studies of Left Ventricular Dysfunction (SOLVD) Prevention trial. *American Heart Journal*, 148(5), 883-888.

259. Burger, A., Tsao, L., & Aronson, D. (2005). Prognostic impact of diabetes mellitus in patients with acute decompensated heart failure. *The American Journal of Cardiology*, 95(9), 1117-1119.

260. Eshaghian, S., Horwich, T.B., & Fonarow, G.C. (2006). An unexpected inverse relationship between HbA1c levels and mortality in patients with diabetes and advanced systolic heart failure. *American Heart Journal*, 151(1), 91.e1-91.e6.

261. McDonald, S.P., Russ, G.R., Kerr, P.G., Collins, J.F., & Australia and New Zealand Dialysis and Transplant Registry. (2002). ESRD in Australia and new Zealand at the end of millennium: a report from the ANZDATA registry. *American Journal of Kidney Diseases*, 40(6), 1122-1131.

262. Foley, R.N., Parfrey, P.S., Harnett, J.D., Kent, G.M., O'Dea, R., Murray, D.C., & Barre, P.E. (1998). Mode of dialysis therapy and mortality in end-stage renal disease. *Journal of American Society of Nephrology*, 9(2), 267-276.

263. Disney, A.P. (1995). Demography and survival of patients receiving treatment for chronic renal failure in Australia and New Zealand: report on dialysis and renal transplantation treatment from the Australia and New Zealand Dialysis and Transplant Registry. *American Journal of Kidney Diseases*, 25, 165-175.

264. Marcelli, D., Spotti, D., Conte, F., Tagliaferro, A., Limido, A., Lonati, F., Malberti, F., & Locatelli, F. (1996). Survival of diabetic patients on peritoneal dialysis or haemodialysis. *Peritoneal Dialysis International*, 16(Supplement 1), S283-S287.

265. Vonesh, E.F., & Moran, J. (1999). Mortality in end-stage renal disease: a reassessment of differences between patients treated with haemodialysis and peritoneal dialysis. *Journal of the American Society of Nephrology*, 10(2), 354-365.

266. Mircescu, G., Garneata, L., Florea, L., Cepoi, V., Capsa, D., Covic, M., Gherman-Caprioara, M., Gluhovschi, G., Golea, O.S., Barbulescu, C., Rus, E., Santimbrean, C., Mardare, M., & Covic A. (2006). The success story of peritoneal dialysis in Romania: analysis of differences in mortality by dialysis modality and influence of risk factors in a national report. *Peritoneal Dialysis International*, 26(2), 266-275.

267. Held, P.J., Port, F.K., Turenne, M.N., Gaylin, D.S., Hamburger, R.J., & Wolfe, R.A. (1994). Continuous ambulatory peritoneal dialysis and haemodialysis: comparison of patient mortality with adjustment for comorbid conditions. *Kidney International*, 45(4), 1163-1169.

268. Vonesh, E.F., Snyder, J.J., Foley, R.N., & Collins, A.J. (2004). The differential impact of risk factors on mortality in haemodialysis and peritoneal dialysis. *Kidney International*, 66(6), 2389-2401.

269. Lee, G. (2003). End-stage renal disease in the Asian-Pacific region. *Seminars in Nephrology*, 23(1), 107-114.

270. Van Biesen, W., Vanholder, R.C., Veys, N., Dhondt, A., & Lameire, N.H. (2000). An Evaluation of an Integrative Care Approach for End-Stage Renal Disease Patients. *Journal of the American Society of Nephrology*, 11, 116-125.

271. Locatelli, F., Marcelli, D., Conte, F., Limido, A., Lonati, F., Malberti, F., & Spotti, D. (1995). 1983 to 1992: report on regular dialysis and transplantation in Lombardy. *American Journal of Kidney Diseases*, 25, 196-205.

272. Choong, H-L. (2005). *Second report of the Singapore renal registry 1998*. Singapore Health Promotion Board.

273. O'Brien, S.R., Hein, E.W., & Sly, R.M. (1980). Treatment of acute asthmatic attacks in a holding unit of a pediatric emergency room. *Annals of Allergy*, 45(3), 159-162.

274. Willert, C., Davis, A.T., Herman, J.J., Holson, B.B., & Zieserl, E. (1985). Short-term holding room treatment of asthmatic children. *The Journal of Pediatrics*, 106(5), 707-711.

275. Miescier, M.J., Nelson, D.S., Firth, S.D., & Kadish, H.A. (2005). Children with asthma admitted to a pediatric observation unit. *Pediatric Emergency Care*, 21(10), 645-649.

276. Levett, I., Berry, K., & Wacogne, I. (2006). Review of a paediatric emergency department observation unit. *Emergency Medicine Journal*, 23(8), 612-613.

277. Arendts, G., MacKenzie, J., & Lee, J.K. (2006). Discharge planning and patient satisfaction in an emergency short-stay unit. *Emergency Medicine Australasia*, 18(1), 7-14.

278. Abe, O., Abe, R., Enomoto, K., & Kikuchi, K. (1998). Polychemotherapy for early breast cancer: an overview of the randomised trials. *The Lancet*, 352(9132), 930-942.

279. Cole, B.F., Gelber, R.D., Gelber, S., Coates, A.S., & Goldhirsch, A. (2001). Polychemotherapy for early breast cancer: an overview of the randomised clinical trails with quality-adjusted survival analysis. *The Lancet*, 358(9278), 277-286.

280. Clarke, M., Collins, R., Darby, S., Davies, C., Elphinstone, P., Evans, E., Godwin, J., Gray, R., Hicks, C., James, S., MacKinnon, E., McGale, P., McHugh, T., Peto, R.,

Taylor, C., Wang, Y., & Early Breast Cancer Trialists' Collaborative Group. (2005). Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials.. *The Lancet*, 366(9503), 2087-2106.

281. Jenicek, M. (2003). *Foundations of evidence-based medicine*. Boca Raton: The Parthenon Publishing Group.

282. Eddy, D.M. (1996). *Clinical decision making: from theory to practice*. Boston: Jones and Bartlett Publishers

283. Redelmeier, D.A., & Tversky, A. (1990). A discrepancy between decisions for individual patients and for groups. *New England Journal of Medicine*, 322, 1162-1164.

284. Sesardic, N. (2007). Sudden infant death or murder? A royal confusion about probabilities. *The British Journal for the Philosophy of Science*, 58(2), 299-329.

285. Imai, K., Suga, K., & Nakachi, K. (1997). Cancer preventive effects of drinking green tea among a Japanese population. *Preventive Medicine*, 26(6), 769-775.

286. Doss, M.X., Potta, S.P., Hescheler, J., & Sachinidis, A. (2005). Trapping of growth factors by catechins: a possible therapeutic target for prevention of proliferative diseases. *Journal of Nutritional Biochemistry*, 16(5), 259-266.

287. Nagano, J., Kono, S., Preston, D.L., & Mabuchi, K. (2001). A prospective study of green tea consumption and cancer incidence, Hiroshima and Nagasaki (Japan). *Cancer Causes Control*, 12(6), 501-508.

288. Zhong, L., Goldberg, M.S., Gao, Y-T, Hanley, J.A., Parent, M.E., & Jin, F. (2001). A population-based case-control study of lung cancer and green tea consumption among women living in Shanghai, China. *Epidemiology*, 12(6), 695-700.

289. Wu, A-H., Yu, M-C., Tseng, C-C., Hankin J., & Pike, M.C. (2003). Green tea and risk of breast cancer in Asian Americans. *International Journal of Cancer*, 106(4), 574-579.

290. Jian, L., Xie, L-P., Lee, A-H., & Binns, C.W. (2004). Protective effect of green tea against prostate cancer: a case-control study in southeast China. *International Journal of Cancer*, 108(1), 130-135.

291. Ji, B-T, Chow, W-H, Yang, G., McLaughlin, J.K., Gao, R-N, Zheng, W, Shu, X-O, Jin, F., Fraumenti, J.F., & Gao, Y-T. (1996). The influence of cigarette smoking, alcohol and green tea consumption on the risk of carcinoma of the cardia and distal stomach in Shanghai, China. *Cancer*, 77(12), 2449-2457.

292. Katiyar, S.K., Perez, A., & Mukhtar, H. (2000). Green tea polyphenols treatment to human skin prevents formation of ultraviolet light B-induced pyrimidine dimmers in DNA. *Clinical Cancer Research*, 6(10), 3864-3869.

293. Kovacs, E.M., Lejeune, M.P., Nijs, I., Westerterp-Plantenga, M.S. (2004). Effects of green tea on weight maintenance after body-weight loss. *British Journal of Nutrition*, 91(3), 431-7.

294. Jenicek, M. (2001). *Clinical case reporting in evidence-based medicine*. London: Arnold.

295. Lohr, K.N., Eleazer, K., & Mauskopf, J. (1998). Health policy issues and applications for evidence-based medicine and clinical practice guidelines. *Health Policy*, 46(1), 1-19.

296. Straus, S.E., & McAlister, M.D. (1999). Evidence-based medicine: past, present, and future. *Annals of the Royal College of Physicians and Surgeons of Canada*, 32, 260-263.

297. Hurwitz, B. (1995). Clinical guidelines and the law: advice, guidance or regulation? *Journal of Evaluation in Clinical Practice*, 1(1), 49-60.

298. Pelly, J.E., Newby, L., Tito, F., Redman, S., & Adrian, A.M. (1998). Clinical practice guidelines before the law: sword or shield? *Medical Journal of Australia*, 169, 330-333.

299. Ibrahim, J.G., & Chen, M-H. (1998). Power prior distributions and Bayesian computation for proportional hazards model. *Sankhya Series B*, 60, 48-64.

300. Ibrahim, J.G., & Chen, M-H. (2002). Power prior distributions for regression models. *Statistical Science*, 15(1), 46-60.

301. Paolino, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis*, 9(4), 325-346.

302. Ferrari, S.L., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799-815.

303. Chan, S-P. (2006). Beta regression analysis of Singapore general election results. In Proceedings: 2nd IMT-GT 2006 Regional Conference on Mathematics, Statistics and Applications, June 2006, Penang, Malaysia, 201-204.

304. Kumaraswamy, P. (1980). A generalized probability density function for double-bounded ransom processes. *Journal of Hydrology*, 46, 79-88.

305. Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalised linear mixed models using adaptive quadrature. *The Stata Journal*, 2, 1-21.

306. Rabe-Hesketh, S., Skrondal, A, & Pickles, A. (2004). Generalized multilevel structural equation modelling. *Psychometrika*, 69(2), 167-190.

307. Rabe-Hesketh, S., & Skrondal, A. (2005). *Multilevel and longitudinal modeling using Stata*. Texas: Stata Press.

308. Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128, 301-323.

309. Skrondal, A, & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. Boca Raton: Chapman & Hall/CRC Press.

310. McCutcheon, A. L. (1987). *Latent class analysis*. California: Sage Publications.

311. Kaplan, D. (2000) *Structural equation modeling: foundations and extensions.* California: Sage Publications.

312. Good, I.J. (1961). A causal calculus. *British Journal of the Philosophy of Science*, 11, 305-318.

313. Pearl, J. (1986). Fusion, propagation and structuring in belief networks. *Artificial Intelligence*, 29, 241-288.

314. Pearl, J. (1988). *Probabilistic reasoning in intelligence systems: networks of plausible inference*. California: Morgan Kauffmann.

315. Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4), 669-688.

316. Jensen, F.V. (2001). *Bayesian networks and decision graphs*. New York: Springer Verlag.

317. Bøttcher, S.G., & Dethlefsen, C. (2003). *DEAL: a package for learning Bayesian networks* (http://www.math.auc.dk/novo/deal).

318. Marcot, B.G., Holthausen, R.S., Raphael, M.G., Rowland, M., & Wisdom, M. (2001). Using Bayesian belief networks to evaluate fish and wildlife population variability under land management alternatives from an environmental impact statement. *Forest Ecology and Management*, 153(1-3), 29-42.

319. Heckerman, D. (1991). *Probabilistic similarity networks*. Cambridge: MIT Press.

320. Russell, S.J., & Norvig, P. (2003). *Artificial intelligence*. New Jersey: Pearson Education.

321. Thibaudeau, Y., & Winkler, W.E. (2002). Bayesian networks representations, generalized imputation, and synthetic micro-data satisfying analytic constraints. *Technical Report RRS200219, United States Bureau of the Census, Washington DC*.

322. Di Zio, M., Scanu, M., Coppola, L., Luzi, O., & Ponti, A. (2004). Bayesian networks for imputation. *Journal of the Royal Statistical Society (Statistics in Society)*, 167(2), 309-322.

323. Pregibon, D. (1996). Data mining. *Statistical Computing and Graphics Newsletter*, 7, 8.

324. Hand, D.J. (1999). Statistics and data mining: intersecting disciplines. *Proceedings of the 7th Association of Computing Machinery's* (*ACM SIGKDD*) *International Conference on Knowledge Discovery and Data Mining*, 1(1), 16-19.

325. Hand, D.J. (1998). Data mining: statistics and more? *The American Statistician*, 52(2), 112-118.

326. Han, J-W., & Kamber, M. (2006). *Data mining: concepts and techniques*. California: Morgan Kauffmann.

327. Giudici, P. (2003). *Applied data mining: statistical methods for business and industry*. Chichester: Wiley.

328. Cherkassky, V., & Mulier, F. (2007). *Learning from data: concepts, theory, and methods*. New Jersey: Wiley.

329. Myatt, G.J. (2007). *Making sense of data: a practical guide to exploratory data analysis and data mining*. New Jersey: Wiley.

330. Fisher, D., & Lenz, H-J. (1996). *Learning from data: artificial intelligence and statistics Vol. 5*. New York: Springer.

331. Nakhaeizadeh, G., & Taylor, C.C. (1997). *Machine learning and statistics*. New York: Wiley.

332. Kay, J.W., & Titterington, D.M. (1999). *Statistics and neural network: advances at the interface*. New York: Oxford University Press.

333. Lindley, D.V. (2006). *Understanding Uncertainty*. New Jersey: Wiley.

334. Chen, M-H., Dey, D.K., & Sinha, D. (2002). Bayesian analysis of multivariate mortality data with large families. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 49(1), 135-150.

335. Chen, M-H., Harrington, D.P., & Ibrahim, J.G. (2002). Bayesian cure rate models for malignant melanoma: a case-study of Eastern Cooperative Oncology Group trial E1690. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 51(2), 135-150.

336. Basu, S., Sen, A. & Benerjee, M. (2003). Bayesian analysis of competing risks with partially masked cause of failure. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 52(1), 77-93.

337. Mezzetti, M., Ibrahim, J.G., Bois, E.Y., Ryan, L.M., Ngo, L., & Smith, T.J.. (2003). A Bayesian compartmental model for the evaluation of th1.3-butadiene metabolism. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 52(3), 291-305.

338. Rutter, C.M., & Simon, G. (2004). A Bayesian method for estimating the accuracy of recalled depression. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 53(2), 341-353.