Statistical Challenges in the Modelling and Analysis of Critical Care Mortality Data

Richard Leigh Bailey

September 22, 2020

Thesis submitted for the degree of
Doctor of Philosophy
in
Statistics
at The University of Adelaide
Faculty of Engineering, Computer and Mathematical Sciences
School of Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

List of Tables

List of Figures

List of Abbreviations

AIC	Akaike Information Criterion
AMI	acute myocardial infarction
ANZICS	Australian and New Zealand Intensive Care Society
APACHE	Acute Physiology and Chronic Health Evaluation
APD	Adult Patient Database
AQ	adaptive quadrature
AUC	area under the curve
BIC	Bayesian Information Criterion
CABG	coronary artery bypass graft
CI	confidence interval
CIHI	Canadian Institute for Health Information
CMS	Centers for Medicare and Medicaid Services
CORE	Centre for Outcome and Resource Evaluation
DT	default tolerance (always preceded by AQ)
EPV	events per variable
FE	fixed-effects (always succeeded by model)
GC	garbage collection
ICNARC	Intensive Care National Audit and Research Centre
ICU	intensive care unit
LOS	length of stay
LSMR	(natural) logarithm of the standardised mortality ratio
LT	low tolerance (always preceded by AQ)
MV	mechanical ventilation
NHS	National Health Service
NZ	New Zealand
PDF	probability density function
PMF	probability mass function
Q-Q	quantile-quantile (always contained in normal Q-Q plot)
RE	random-effects (always succeeded by model)
ROC	receiver operating characteristic
RSMR	risk-standardised mortality rate
SAPS	Simplified Acute Physiology Score
SD	standard deviation
SE	standard error

SES	socio-economic status
SMO	standardised mortality odds
SMR	standardised mortality ratio
UK	United Kingdom
USA	United States of America

Abstract

Monitoring the performance of healthcare providers assists with maximising patient safety, maintaining quality control, and managing financial resources. However, such monitoring also attracts controversy, particularly when jurisdictions publish performance rankings, due to the effect on providers' reputations, referrals, and reimbursements for treatment. Thus, accurate and fair performance assessment is vital.

The focus of this thesis is the performance of critical care providers in hospital intensive care units (ICUs). Patient mortality in hospital is the outcome that is monitored, and the standardised mortality ratio (SMR) is the performance indicator. Four inter-related methods of calculating SMRs are considered using a general multilevel model that contains patient- and provider-level attributes. Two options for the numerator of the SMR are either the observed mortalities or a smoothed version of the observed mortalities that allows for sampling variation. There are also two options for the denominator of the SMR, which represents the expected mortalities. The denominator adjusts for the patient casemix and either the actual or average provider attributes. This thesis presents the first known implementation of SMRs where the numerator is the smoothed observed mortalities, and both the numerator and denominator are based on a model containing (patient- and) provider-level attributes. The advantages and disadvantages of the four methods are discussed.

The smoothed observed mortalities and the expected mortalities are defined in terms of intractable integrals. These integrals are evaluated numerically, often by adaptive quadrature. A much quicker but less accurate alternative is the Laplace approximation, and extending this technique may lead to a better compromise between speed and accuracy. In this thesis, the Laplace approximation is extended beyond that covered in the literature, and details are provided on how the integrals are expressible as series of any order. Comparisons against published results demonstrate that the newly developed extensions are much faster and almost as accurate as adaptive quadrature.

The challenge of estimating SMR uncertainty is addressed by bootstrapping, and a non-parametric procedure is described for this purpose. Bootstrap procedures tend to be computationally intensive, which highlights a significant advantage of using extended versions of the Laplace approximation.

The four variations of SMRs are calculated for a subset of the Australian and New Zealand Intensive Care Society (ANZICS) Adult Patient Database (APD). Deidentified patient records are voluntarily and retrospectively submitted to the APD by the majority of Australian and New Zealand adult ICUs. This thesis examines the in-hospital mortality data of 370,554 patient admissions to 170 ICUs between 2011 and 2015 inclusive. Risk-adjusted mortality generally decreased over time. Furthermore, risk-adjusted mortality was significantly higher for the ICUs of tertiary hospitals. This was reflected in the estimated SMRs when the expected mortalities were based on average ICU attributes. In this application, the smoothing of observed

mortalities is recommended for the process of identifying unusually performing ICUs. There was also substantial variability in the estimated SMRs by each method. This was predominantly due to the variability within hospital types, and was largest for the ICUs of private hospitals.

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: .. Date: ....22 / 9 / 2020.....

Acknowledgements

Firstly, I would like to thank my principal supervisors, Professor Patricia (Patty) Solomon and Associate Professor Andrew Metcalfe. Patty, thank you for being my principal supervisor from the commencement of my PhD candidature until your retirement. Andrew, thank you for taking on the role of principal supervisor from then onwards. I am extremely grateful to you both for your time, guidance, and encouragement.

I also thank my co-supervisor, Associate Professor John Moran. Your support throughout my candidature is really appreciated.

My sincere thanks also go to the Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcomes and Resource Evaluation (CORE) for granting me access to a subset of the ANZICS Adult Patient Database (APD). My research using data would not have been possible without the contribution of the ICU community to the ICU registries; I gratefully acknowledge their contribution.

Finally, thanks must go to my parents, Lynda and Raymond, and my brother, Darren. I thank you all for your support and encouragement.

Dedication

I dedicate this thesis to my mother, Lynda, and my late maternal grandparents, Barbara and Robert (Jim).

Chapter 1

Introduction

The earliest known system of healthcare management was written (in cuneiform) on a large codex that listed the laws of King Hammurabi of Babylon, circa 1700 B.C. (Spiegel and Springer, 1997). The codex was excavated from the remains of the ancient city of Susa, Persia (now Iran), in early 1902, and translated later that same year (Holmes, 1905). Spiegel and Springer (1997) discussed the contents of Hammurabi's code in relation to healthcare, which included decrees that

- patients needed to be informed and understand their rights;
- physicians needed to maintain medical records of their patients' treatments and outcomes, and governmental regulators could inspect those records;
- for a successful outcome, the patient (or their owner in the case of a slave) would pay their physician a fixed amount according to the treatment category and the patient's class (upper, middle, or slave); and
- for an unsuccessful outcome (patient death or injury), the physician would be penalised according to the treatment category and the patient's class.

Based on Hammurabi's Code, physicians held an esteemed position in Babylonian society and could be well paid for their work (Holmes, 1905). Nevertheless, the quality control in that society could result in a physician's first misfortune ending their career and leaving them mutilated. To illustrate the harshness of some Babylonian laws, one of the physician's hands would be amputated if they unsuccessfully performed a general or ocular surgery on an upper-class citizen (Spiegel and Springer, 1997).

Since the mid-19th century, epidemiological and clinical philosophies have been promoted in the monitoring of healthcare outcomes (Spiegelhalter, 1999). First, the epidemiological approach was pioneered by Florence Nightingale, an epidemiologist and nursing reformer (Bottle and Aylin, 2016, pp. 7 – 8). Nightingale campaigned for the adoption of a uniform system of recording hospital statistics, and her scheme received official approval (with additions proposed) at the International Statistical Congress of 1860, held in London (Farr, 1861, pp. 173 – 174). In accordance with that approval, hospitals were requested to record a set of administrative and clinical details pertaining to each of their patients. The requested details included each patient's age, sex, disease/injury, surgery, in-hospital complications, and mortality outcome (Nightingale, 1863, p. 172). According to Nightingale (1863, p. 176), the objectives of adopting this approach were to improve treatment and resource management by ascertaining each hospital's

- mortality rate, overall, and stratified by age and disease/injury; and
- relative treatment frequency by disease and injury.

Unfortunately, she was rarely successful in acquiring adequate hospital records for comparative purposes and determining whether hospitals were spending money judiciously (Nightingale, 1863, p. 176).

A clinical approach to the monitoring of surgical performance was endorsed by Ernest Amory Codman, one of the founders of the American College of Surgeons (Neuhauser, 2002). Codman began developing his system in 1900, campaigning for its adoption in 1910 before opening his own private hospital at Boston, Massachusetts, in 1911 (Spiegelhalter, 1999). Codman's hospital implemented his End Result System, where surgeons maintained patient records on symptoms; diagnostic and surgical details, including errors; and post-operative progress (Codman, 1918). The objective of surgeons admitting and recording their errors was for failures to be analysed so that steps could be taken to effect possible improvements (Codman, 1918, p. 9). A key difference between Codman's and others' hospitals was that he published his hospital's errors and challenged others to do likewise for their hospitals (Neuhauser, 2002). At that time, Codman's idealistic approach effectively imposed legal liabilities on his own hospital for the types of errors that other hospitals would have evaded any liability for (Codman, 1918, p. 121). Spiegelhalter (1999) noted that Codman's hospital closed in 1918, though his work on hospital standardisation and accreditation continued. However, it was not until 1987 that standardisation was first used for evaluating hospital performance, by the Health Care Financing Administration in the United States of America (USA) (Normand et al., 1997).

In general, monitoring the performance of healthcare providers assists with maximising patient safety, maintaining quality control, and managing financial resources. Concerns for public safety are understandably heightened when healthcare scandals are (appropriately) reported on and inquired into. Since the early 1990s, examples of healthcare scandals in England involved the Bristol Royal Infirmary (Dyer, 2001), the Mid Staffordshire National Health Service (NHS) Trust (Dyer, 2015), and the mass-murdering, general practitioner Harold Shipman (Bottle and Aylin, 2016, p. 9). Meanwhile, several Australian hospitals and health services have also been involved in scandalous misconduct (Duckett et al., 2016). Various quality control procedures are used to monitor, and report on, healthcare performance through regulation and accreditation, and indeed all developed countries have some form of regulatory system (Bottle and Aylin, 2016, p. 16). Regulatory and accreditation requirements undoubtedly ensure some level of transparency in healthcare. As Bottle and Aylin (2016, p. 13) noted, failings in the Mid Staffordshire NHS Trust were originally detected through a monitoring system. Further illustrating the importance of healthcare monitoring is for financial accountability towards funders of those (often expensive) services.

The focus of this thesis is the performance of critical care providers in hospital intensive care units (ICUs). Critical (or intensive) care is the type of care that patients receive for life-threatening conditions which require continual monitoring and treatment (MedicineNet, 2020). Although patients usually receive critical care in an ICU, patients may initially receive such care elsewhere, including the hospital's emergency department if the ICU has reached its maximum capacity (Huang, 2004). In comparison to the general wards of a hospital, ICUs provide more invasive treatments which can temporarily support or supplant the functions of a patient's

vital organs (Kelly et al., 2014). The first ICU in the world was established in 1953 by the Danish anaesthetist Bjørn Ibsen at Kommunehospitalet (municipal hospital) in Copenhagen, Denmark (Berthelsen and Cronqvist, 2003). The Australian and New Zealand Intensive Care Society (ANZICS) stated that the first ICU in New Zealand was founded at Auckland Hospital in 1958, and the first ICU in Australia was founded at St Vincent's Hospital, Melbourne, in 1961 (ANZICS, 2020). ICUs have undergone tremendous advances in their monitoring and treatment of patients, and improvements will doubtlessly continue to occur (Kelly et al., 2014). Further emphasising the importance of ICUs in the healthcare system, the ANZICS Centre for Outcomes and Resource Evaluation (CORE) suggested that the monitoring of ICU data may assist with the early detection of healthcare problems and pandemics (ANZICS CORE, 2017).

Established in 1975, ANZICS is the Australian and New Zealand intensive care community's foremost advocate, dedicated to not only providing medical leadership and continuing professional education, but also conducting and supporting critical care research and analysis (ANZICS, 2020). Research and analysis are conducted or supported by ANZICS CORE, administrator of four critical care registries (ANZICS CORE, 2019): the Adult Patient Database (APD), the Australian and New Zealand Paediatric Intensive Care Registry, the Critical Care Resources Registry, and the Central Line Associated Blood Stream Infection Registry. Hart (2008) explained that ANZICS Databases changed its name to ANZICS CORE in 2008 because the former name only reflects their means of functioning, while the latter name reflects their actual functions. The improved clarity is likely to be important for the funders of ANZICS CORE, which are the governments of New Zealand and every Australian state and territory (ANZICS CORE, 2019).

This thesis uses a subset of the critical care data from the APD. The University of Adelaide guidelines indicate that all research which was conducted for this thesis is exempt from Human Research Ethics Committee review due to its negligible risk. I submitted an application for access to some of the APD from its owners, ANZICS CORE. The CORE Management Committee approved my request in June 2016, and I entered into an ANZICS confidentiality agreement. Pursuant to that agreement, ANZICS CORE retains sole ownership of all data which they disclosed to me, and no party is permitted access to those data without prior official approval from ANZICS. Furthermore, ANZICS CORE have reviewed this thesis before its submission to ensure that there is no inadvertent identification of specific units (ICUs or patients).

The APD was constructed to benchmark and audit adult ICU performance, with the aim of promoting improvements in patient outcomes (ANZICS CORE, 2019). Since its inception in 1992, the APD has collected de-identified administrative and clinical data on patients' admissions, submitted voluntarily by Australian and New Zealand ICUs (Stow et al., 2006). Although voluntary in principle, most ICUs currently submit data. The APD is updated quarterly and retrospectively, now containing details on over two million patient admissions, making it one of the largest databases of its kind in the world (ANZICS, 2019). The outcome is whether the patient survived to hospital discharge or died in hospital. Given the high mortality risks of patients in critical care, mortality-related outcomes are typically used in assessing their healthcare providers. For example, ANZICS CORE uses mortality-related statistics in the initial stage of their detailed process for assessing ICU performance (McClean et al., 2017).

Chapter 2 is a review of standardised mortality statistics for the assessment of healthcare provider performance. General risk-adjustment models are defined for this purpose, and various issues are addressed in regard to what mortality risks may be adjusted for. Most of this chapter is devoted to calculating standardised mortality ratios (SMRs). The SMR is the most prevalent and viable standardised mortality statistic used for evaluating provider performance (Manktelow et al., 2014). There are two considerations: whether to smooth the observed mortalities to allow for sampling variation, and whether the benchmark of expected performance should account for the actual or average provider attributes. The SMR methods obtained from those four possible combinations are considered. Furthermore, one of the methods was suggested by Ash et al. (2012) but has yet to be applied in the literature. There are two main challenges with implementing these SMR methods. Firstly, the underlying SMR expressions are defined in terms of intractable integrals. Secondly, estimating the uncertainty associated with the SMRs is complicated. The second of these challenges is addressed by a bootstrap procedure, though doing so increases the (already considerable) computational intensity which needs to be addressed by the first challenge.

In Chapter 3, the computational intensity of evaluating an enormous number of intractable integrals is addressed. These integrals often have lengthy integrands, which makes their evaluation more time-consuming. One can evaluate these integrals relatively quickly by the Laplace approximation. Pierre-Simon Laplace originally published his eponymous approximation technique in 1774 (Laplace, 1986). Extending the Laplace approximation, as was done, for example, in Tierney and Kadane (1986, Appendix) and Raudenbush et al. (2000), generally yields greater accuracy, but slightly decreases the evaluation speed. However, in this chapter, details are provided on how the Laplace approximation can be extended beyond that covered in the literature, to any desired extent. The speed and accuracy of Laplace-related approximations are compared with the corresponding speed and accuracy of the more commonly used, and readily available, adaptive quadrature option.

In Chapter 4, a five-year subset of the ANZICS APD, from 2011 to 2015, is used to assess ICU performance by the four SMR methods of Chapter 2. This involves approximating the smoothed observed mortalities and the expected mortalities by an extended version of the Laplace approximation developed in Chapter 3. The uncertainty of the SMRs, by each method, is estimated using the bootstrap procedure described in Chapter 2. Finally, the results and future research are discussed. Chapter 5 is a summary of the results and recommendations of this thesis.

Chapter 2

Standardised mortality statistics

The desire to compare provider performance motivates the development of an appropriate summary statistic. Provider comparisons are often based on mortality, and this outcome is typically used in the critical care context (Kasza et al., 2013). However, it is generally inappropriate to assess providers by simply examining their raw (unadjusted) mortality rates because providers that treat a greater proportion of patients with poorer prognoses (and higher mortality risks) are expected to have higher mortality rates (Manktelow et al., 2014). The characteristics of patients which may affect their outcomes can be categorised as pre- or post-admission (Normand et al., 2016). Pre-admission characteristics are beyond a provider's control and comprise the patient casemix. Examples of these characteristics include each patient's age, illness type/severity, surgical status, and time/date of admission. To illustrate, by categorising age into generally decadal ranges, Knaus et al. (1991) quantified the increased mortality risks with each subsequent age category, all other things being equal. Furthermore, general reductions in patient mortality over time across all providers may be attributable to improvements in the quality of care given (Zimmerman et al., 2013). Critical care improvements in recent decades were seldomly due to innovative therapeutics (Gordon and Russell, 2018); these improvements were more likely the result of simplifying the critical care provided, particularly by omitting interventions that were found to be harmful or ineffective (Perner et al., 2018).

The aim of standardising for the patient casemix is to enable the fair evaluation of provider performance. Although post-admission characteristics may also affect patient mortality, they should not be accounted for during the standardisation due to being (at least to some extent) controllable by providers. Examples of these characteristics include in-hospital infections and length of stay (LOS). The allowance for in-hospital infections would create the misleading perception that poorer-quality hospitals performed better than they actually did. The arguments against adjusting for LOS include the ambiguity surrounding how it may be interpreted. Suppose a patient had a remarkably short LOS and survived to hospital discharge. Perhaps he or she received efficient, better-quality treatment, or instead received substandard treatment and was discharged prematurely?

Numerous methods of deriving standardised mortality statistics have been published in the statistical literature, and many of these methods are described in Bottle and Aylin (2016). However, the discussion presented in this thesis will pertain to a range of such statistics that are derived by standardisation using logistic regression. As Normand et al. (2016) indicated, logistic regression is suitable for the task of standardisation because it enables adequate casemix adjustment and accommodates the hierarchical structure of patients being nested within providers.

This chapter explains how various standardised mortality statistics within two main categories, the standardised mortality odds (SMO) and standardised mortality ratio (SMR), may be constructed for evaluating provider performance. Comparisons are also made with the methods used in the existing literature. A review of the literature on evaluating provider performance revealed that SMOs are seldom implemented, while SMRs are far more common. Using multilevel logistic regression, the chapter begins with SMOs in Section 2.1 and builds upon that work with SMRs in Section 2.2. Multilevel models could allow for a variety of hierarchical attributes, including patients, providers, jurisdictions, and time. This thesis will restrict the hierarchy to two levels, namely patients and providers, because even this level of complexity proved highly challenging to implement during the application considered later in the thesis (Chapter 4). The present chapter concludes with Section 2.3, which considers modifications or extensions of the models used in previous sections.

2.1 Standardised mortality odds

Using a two-level logistic model, suppose the providers are numbered $i = 1, \ldots, I$, and the patients treated in provider i are numbered $j = 1, \ldots, n_i$. Denote the outcome for patient j in provider i as

$$Y_{ij} = \begin{cases} 1 & \text{if the patient dies within a fixed time of treatment} \\ 0 & \text{otherwise,} \end{cases}$$

which may differ depending on how long the corresponding patient survived the fixed time from treatment. Take the case of intensive care units (ICUs). Although in-hospital mortality remains the most common outcome, longer-term, fixed points in time (e.g. 30-day) are sometimes preferred due to varying discharge policies between ICUs (Brinkman et al., 2013). However, the use of longer-term mortalities requires adequate linkage between databases, a criterion that can be unachievable in practice (Kasza et al., 2013).

This section accommodates provider effects on their patients' outcomes via a single term per provider in the model. The provider effects may be treated as fixed or random, and each representation is discussed below.

2.1.1 Fixed effects

With a fixed effect, β_i, for provider i, it is assumed that

$$Y_{ij} \,|\, \boldsymbol{x_{ij}}, \beta_i \overset{ind}{\sim} Bern\left(\pi_{ij}\right),$$

where π_{ij} is patient j's fitted mortality probability, and their log odds of mortality is

$$\text{logit}\left(\pi_{ij}\right) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \mu + \boldsymbol{\alpha^T x_{ij}} + \beta_i.$$

The other terms in the model are the vector of patient attributes, $\boldsymbol{x_{ij}}$, the associated risk parameters, $\boldsymbol{\alpha}$, and the base level, μ. Setting up the provider effects to have the zero-sum constraint, $\sum_{i=1}^{I} \beta_i = 0$, is most convenient because a Wald test can be used to assess whether the log odds of survival for patients treated in a specific provider is significantly better or worse than an average provider. Hence, provider i changes the log SMO by β_i, and the SMO will be multiplied by a factor of e^{β_i}.

2.1.2 Random effects

With a random effect, ε_i, for provider i, it is now assumed that

$$Y_{ij} \,|\, \boldsymbol{x_{ij}}, \varepsilon_i \overset{ind}{\sim} Bern\left(\pi_{ij}\right),$$

where

$$\varepsilon_i \overset{iid}{\sim} N\left(0, \tau^2\right)$$

and the log odds of mortality for patient j is now

$$\text{logit}\left(\pi_{ij}\right) = \mu + \boldsymbol{\alpha^T x_{ij}} + \varepsilon_i.$$

As before, the other terms in the model are the vector of patient attributes, $\boldsymbol{x_{ij}}$, the associated risk parameters, $\boldsymbol{\alpha}$, and the base level, μ. The provider-specific information to be estimated by this random-effects model is in the parameter τ^2. However, assessing providers' SMOs necessitates the examination of the ε_i. Now the information about ε_i can be appropriately summarised by its distribution conditional on the data involving provider i (Cox and Solomon, 2002, p. 55). Using an empirical Bayes approach, this proceeds by considering the model as giving the prior distribution of ε_i, then calculating the posterior distribution of ε_i using Bayes' rule. This approach can be thought of as empirical Bayes because τ^2 in the prior distribution is based on the initial logistic regression. It avoids the additional, fully Bayesian issues of defining a suitable prior probability for τ^2, then integrating over its distribution (Cox and Solomon, 2002, p. 54). Addressing such issues would be too computationally intensive for the applications in this thesis. Moreover, given the number of data, the influence of such a prior is likely to be negligible.

The probability density function (PDF) for the posterior distribution of ε_i will now be derived using Rabe-Hesketh and Skrondal (2012). Since Y_{ij} is a Bernoulli-distributed random variable with mortality probability π_{ij}, the probability mass function (PMF) of Y_{ij} is

$$\begin{aligned}
p\left(y_{ij} \,|\, \boldsymbol{x_{ij}}, \varepsilon_i\right) &= \pi_{ij}^{y_{ij}} \left(1 - \pi_{ij}\right)^{1-y_{ij}} \\
&= \left[\frac{e^{\mu+\boldsymbol{\alpha^T x_{ij}}+\varepsilon_i}}{1 + e^{\mu+\boldsymbol{\alpha^T x_{ij}}+\varepsilon_i}}\right]^{y_{ij}} \left[1 - \frac{e^{\mu+\boldsymbol{\alpha^T x_{ij}}+\varepsilon_i}}{1 + e^{\mu+\boldsymbol{\alpha^T x_{ij}}+\varepsilon_i}}\right]^{1-y_{ij}} \\
&= \left[\frac{e^{\mu+\boldsymbol{\alpha^T x_{ij}}+\varepsilon_i}}{1 + e^{\mu+\boldsymbol{\alpha^T x_{ij}}+\varepsilon_i}}\right]^{y_{ij}} \left[\frac{1}{1 + e^{\mu+\boldsymbol{\alpha^T x_{ij}}+\varepsilon_i}}\right]^{1-y_{ij}} \\
&= \frac{e^{\left(\mu+\boldsymbol{\alpha^T x_{ij}}+\varepsilon_i\right)y_{ij}}}{1 + e^{\mu+\boldsymbol{\alpha^T x_{ij}}+\varepsilon_i}}.
\end{aligned}$$

By independence, the joint PMF of $\boldsymbol{Y_i} = \left[Y_{i1}, \ldots, Y_{in_i}\right]^T$ is

$$\begin{aligned}
p\left(\boldsymbol{y_i} \,|\, X_i, \varepsilon_i\right) &= \prod_{l=1}^{n_i} p\left(y_{il} \,|\, \boldsymbol{x_{il}}, \varepsilon_i\right) \\
&= \prod_{l=1}^{n_i} \frac{e^{\left(\mu+\boldsymbol{\alpha^T x_{il}}+\varepsilon_i\right)y_{il}}}{1 + e^{\mu+\boldsymbol{\alpha^T x_{il}}+\varepsilon_i}} \\
&= \exp\left\{\sum_{l=1}^{n_i} \left[\left(\mu + \boldsymbol{\alpha^T x_{il}} + \varepsilon_i\right) y_{il} - \log\left(1 + e^{\mu+\boldsymbol{\alpha^T x_{il}}+\varepsilon_i}\right)\right]\right\},
\end{aligned}$$

where

$$X_i \;=\; \begin{bmatrix} \boldsymbol{x_{i1}^T} \\ \vdots \\ \boldsymbol{x_{in_i}^T} \end{bmatrix}.$$

The prior distribution of ε_i has the PDF

$$f\left(\varepsilon_i\right) \;=\; \frac{1}{\sqrt{2\pi\tau^2}}\,\exp\left\{-\frac{\varepsilon_i^2}{2\tau^2}\right\}.$$

Using Bayes' rule, the posterior distribution of ε_i has the PDF

$$f\left(\varepsilon_i \,|\, \boldsymbol{y_i}, X_i\right) \;=\; \frac{p\left(\boldsymbol{y_i} \,|\, X_i, \varepsilon_i\right)\; f\left(\varepsilon_i\right)}{p\left(\boldsymbol{y_i} \,|\, X_i\right)}$$

since the errors are assumed independent of the covariates and $f\left(\varepsilon_i \,|\, X_i\right) = f\left(\varepsilon_i\right)$. Noting that $p\left(\boldsymbol{y_i} \,|\, X_i\right)$ does not depend on ε_i,

$$\begin{aligned}
&f\left(\varepsilon_i \,|\, \boldsymbol{y_i}, X_i\right) \\
&\propto\; p\left(\boldsymbol{y_i} \,|\, X_i, \varepsilon_i\right)\; f\left(\varepsilon_i\right) \\
&=\; \frac{1}{\sqrt{2\pi\tau^2}}\,\exp\left\{-\frac{\varepsilon_i^2}{2\tau^2} + \sum_{l=1}^{n_i}\left[\left(\mu + \boldsymbol{\alpha^T x_{il}} + \varepsilon_i\right) y_{il} - \log\left(1 + e^{\mu + \boldsymbol{\alpha^T x_{il}} + \varepsilon_i}\right)\right]\right\} \\
&\propto\; \exp\left\{-\frac{\varepsilon_i^2}{2\tau^2} - \sum_{l=1}^{n_i}\left[\log\left(1 + e^{\mu + \boldsymbol{\alpha^T x_{il}} + \varepsilon_i}\right) - \varepsilon_i y_{il}\right]\right\}.
\end{aligned}$$

As a valid PDF,

$$\int_{-\infty}^{\infty} f\left(\varepsilon_i \,|\, \boldsymbol{y_i}, X_i\right)\,\mathrm{d}\varepsilon_i \;=\; 1,$$

so

$$f\left(\varepsilon_i \,|\, \boldsymbol{y_i}, X_i\right) \;=\; k_i\,\exp\left\{-\frac{\varepsilon_i^2}{2\tau^2} - \sum_{l=1}^{n_i}\left[\log\left(1 + e^{\mu + \boldsymbol{\alpha^T x_{il}} + \varepsilon_i}\right) - \varepsilon_i y_{il}\right]\right\},$$

where the expression for the normalising constant,

$$k_i \;=\; \left[\int_{-\infty}^{\infty} \exp\left\{-\frac{\varepsilon_i^2}{2\tau^2} - \sum_{l=1}^{n_i}\left[\log\left(1 + e^{\mu + \boldsymbol{\alpha^T x_{il}} + \varepsilon_i}\right) - \varepsilon_i y_{il}\right]\right\}\,\mathrm{d}\varepsilon_i\right]^{-1},$$

contains an intractable integral.

The posterior distribution of ε_i is unrecognisable from the kernel of its PDF, though is roughly "bell shaped" (close to normal). Although this distribution can be approximated by using the Metropolis-Hastings algorithm (see, e.g., Chib and Greenberg, 1995), the normalising constant will still need to be calculated later in this chapter. The intractable integral in the expression for k_i can be evaluated numerically, by adaptive quadrature, or approximately, by the Laplace approximation, as will be discussed in Chapter 3. Indeed, the Laplace approximation is recommended in the present context because it is a numerically efficient technique, which is particularly accurate for bell-shaped distributions.

The prediction of the log SMO for provider i is usually taken as the mean of the posterior distribution of ε_i. An alternative, more computationally efficient approximation to the log SMO, which does not require integration, uses the mode of

the posterior distribution instead (Skrondal and Rabe-Hesketh, 2009). The strength of evidence for whether that provider performed significantly better or worse than average is assessed by the probability that ε_i is negative or positive, respectively, calculated from the posterior distribution.

2.1.3 Comparison between fixed and random effects

In the fixed-effects (FE) model, the random variation models the difference within providers. The overall, null hypothesis is that there is no difference in the provider effect across the population of the providers. If there is evidence against this hypothesis, given the context of identifying over- or under-performing (outlier) providers, it is appropriate to allow for multiple comparisons. In the random-effects (RE) model, there are two components of variation: between providers and within providers. The research questions are to estimate the two components of variation, to consider whether the between-provider variation is acceptably low, and to identify outliers using the empirical posterior distribution of provider errors.

The aim of fitting FE models is to incorporate all relevant covariates and so adjust for each provider's effect on mortality. The FE approach may be appropriate for applications that are solely interested in those specific providers, though DeLong et al. (1997) noted that philosophical differences exist on this matter. Applications are likely to be hampered by some providers treating relatively few patients (i.e., small n_i for provider i) during the study period. Normand et al. (2016) indicated that it can be quite common to have such limited information from some providers when the focus is restricted to particular medical conditions. For example, over a three-year period between 2010 and 2013, 50% of the hospitals in the United States of America (USA) admitted fewer than 15 patients due to acute myocardial infarction (AMI, more commonly known as heart attack). In general, providers that treat very few patients are more likely to have unusually low or high mortality rates due to chance alone rather than significantly over- or under-performing, respectively. The resultant model could yield unrealistic FE estimates with large standard errors. However, Ash et al. (2012) argued that the data from every provider should still be retained because they each contribute considerable information towards estimating $\boldsymbol{\alpha}$ (the risk parameters) when building the model. A compromise is to pool the data from providers with low patient numbers. Subsequent fitting of the FE model would use the data from each patient treated, but the providers whose data were pooled could not be compared individually. Another strategy for avoiding unrealistic FE estimates is to use RE models.

The RE alternative accounts for the stochastic effects associated with patients being clustered within providers (Normand et al., 2016). Compared to FE models, the provider effects estimated by RE models are shrunk towards zero. This leads to estimates being biased towards the overall mean but having smaller variability (Kalbfleisch and Wolfe, 2013). Thus, a bias-variance trade-off is introduced. The RE approach shrinks the estimates of smaller (based on patient numbers) providers that performed relatively well or poorly more towards the overall mean than similarly-performing larger providers.

More generally, the potential risks and ramifications associated with misclassifying providers as outliers may influence whether FE or RE approaches are implemented. Austin et al. (2003) noted, through the use of Monte Carlo simulations, that FE models have greater sensitivity (i.e., are better at correctly identifying

outliers), while RE models have greater specificity (i.e., are better at correctly identifying non-outliers). Substantial division exists in the statistical literature surrounding preferences for FE or RE approaches in the task of assessing performance (Bottle and Aylin, 2016, p. 172). However, this thesis will concentrate on the RE approach.

The current section limited the modelling of provider effects on mortality to a single term per provider in both the FE and RE approaches to ensure all comparisons between them were consistent. More general (fixed) provider effects, such as location and type, often cannot be included in the FE model of Section 2.1.1 due to their collinearity with the fixed terms representing each provider (Moran and Solomon, 2014). Although it is possible to include the additional provider attributes in RE models[1], such as that of Section 2.1.2, this would restrict the comparisons to specific provider characteristics. Section 2.2.1 will further explore various issues pertaining to whether it may be reasonable to accommodate some provider attributes in the benchmark.

2.2 Standardised mortality ratio

An alternative summary statistic to the crude death rate and SMO, which can be used for evaluating provider performance, is the standardised mortality ratio (SMR). Manktelow et al. (2014) suggested that the SMR is the most prevalent and viable statistic used for such purposes. In their simplest form, SMRs compare providers' actual and expected performances. Incorporating both sets of information into one statistic is an attractive feature of SMRs (Solomon et al., 2014) that SMOs lack. Still, SMOs may be reasonable statistics to use in various applications, though Mohammed et al. (2016) argued that there is a tendency to misinterpret odds ratios (such as SMOs). Such matters may contribute to the widespread popularity of SMRs.

Applications of SMRs have lead to the construction (and publication) of league tables for comparing provider performance (Goldstein and Spiegelhalter, 1996) in numerous countries, including England, Canada, and the USA (Mohammed et al., 2009). In the healthcare context, league tables are inappropriate due to their over-reliance on the point estimates of mortality statistics while ignoring the potentially large random variation which is attributable to sampling error (Bottle and Aylin, 2016, p. 154). A review of the literature indicated that funnel plots (Spiegelhalter, 2005) are becoming more frequently used for comparing provider performance in a fairer manner: individual providers are not ranked by estimated SMR-related statistics[2] and their estimated uncertainty is incorporated in these plots. For example, English and Canadian governmental entities use funnel plots in reporting the performance of their nation's hospitals, via the National Health Service (NHS) website for English hospitals (NHS Digital, 2019) and the Canadian Institute for Health Information (CIHI) website for Canadian hospitals (CIHI, 2019). Elsewhere, uncertainty appears not to have been accommodated for by the USA's Centers for Medicare and Medicaid Services (CMS); the government-owned Medicare Hospital Quality Chartbook website reports hospitals' standardised mortality rates (not ratios) using deciles of performance (CMS, 2019*b*).

[1] For example, this approach was implemented by Moran and Solomon (2014), with ICU-site random effects from a random coefficient model being examined on the odds scale.

[2] Risk-standardised mortality rates (RSMRs) are sometimes used instead. A provider's RSMR is the product of their SMR and the raw mortality rate throughout the population of interest.

Australian hospitals have undergone increased public scrutiny since the construction of the government-owned MyHospitals website in 2010 (AIHW, 2018). Mortality-related statistics (including SMRs) are not currently reported, however, because "extensive methodological development" is still required (AIHW, 2018). With respect to critical care, Australian ICUs are not scrutinised publicly, though they are investigated with the use of SMRs. For example, McClean et al. (2017) details an in-depth process followed by the Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcome and Resource Evaluation (CORE), who identify outlying ICUs from their SMRs before assessing whether further action is warranted. Few instances could be found where ICUs in other countries have undergone public scrutiny either. One notable example is the detailed, online reporting on the performance of NHS ICUs, including their SMRs, by the Intensive Care National Audit and Research Centre (ICNARC), an independent, United Kingdom (UK) charity (ICNARC, 2019).

The decision to use SMRs raises the question of how the benchmark of expected performance may be calculated, which will be the first issue addressed in this section. An expanded model that can be used in the process of calculating SMRs will then be defined. This will be followed by a discussion of four inter-related methods for calculating the SMRs. Subsequently, this section will describe a bootstrap procedure for estimating the uncertainty associated with SMRs.

2.2.1 Comparator attributes

While the benchmark of expected performance must always account for the patient casemix, it is also necessary to consider whether the provider attributes should be accounted for. Examples of these attributes include the hospital's level/type, region, and volume (i.e., the number of patients treated during a specific period of time). Indeed, it may seem reasonable to consider these attributes since providers generally have little or no control over them (Bottle and Aylin, 2016, p. 132). It can be of interest how providers perform, given their actual attributes; for example, when a clinician, working in a hospital, wants to investigate how that clinician's hospital is performing for that actual type of hospital. The benchmark will be labelled here as an actual comparator if the SMR accounts for the provider's actual attributes. The performance of each provider is essentially compared with what would be expected of a provider with those characteristics. Applications that use actual comparators include those by Kasza et al. (2013) and Solomon et al. (2014).[3]

A substantial reason for excluding provider attributes from the benchmark is that such attributes are potential proxy measurements for quality. Volume is one such example. Since Luft et al. (1979), many studies have observed a variable inverse relationship between volume and mortality (Halm et al., 2002). Furthermore, some professional organisations even advocate volume being used as a performance standard (Normand et al., 2016). In circumstances where high-volume providers are identified as the best performers, adjusting for their volume in the comparator would effectively result in those providers being assessed at a higher standard than low-volume providers would be. Nevertheless, a large study by Moran and Solomon (2012) reported that lower-volume ICUs had slightly better (mortality) outcomes for

[3]Kasza et al. (2013) and Solomon et al. (2014) used actual comparators in the manner discussed here to identify potentially outlying ICUs at the first stage of their respective investigations.

the mechanically-ventilated subset of patients who were treated in Australia or New Zealand.

The benchmark will be labelled here as an average comparator if the SMR does not account for the provider's actual attributes. That is, each specific provider is compared to how an average (or typical) provider based on the overall study population would perform (Silber et al., 2016). A precise definition of "average provider" is given by Eq. 2.7 in Section 2.2.5. Circumstances where the use of such SMRs would be appropriate include when government employees or healthcare providers want to investigate how particular hospitals compare to the overall average. The CMS is one such governmental entity to focus on this approach, with its mandate of assessing hospital performance against a national-level comparator (Ash et al., 2012).

To illustrate the key difference between actual and average comparators, suppose some country has two regions, A and B, with the most common illnesses being cardiovascular related. Further suppose that every one of that nation's hospitals provides the same quality of care for patients suffering all illnesses except cardiovascular-related, which are treated much better in region A's hospitals due to regional training or facilities available. Now by comparing hospitals against their own type, which accounts for the region, every hospital may be classified as performing adequately. Alternatively, by comparing hospitals against an average one (nationally), it is likely that region A's hospitals would be classified as over-performers, whereas region B's hospitals would be classified as under-performers. Which comparator is relevant depends on the context. Actual comparators could identify over- or under-performing hospitals relative to their own region. Alternatively, average comparators could highlight performance differences between regions, identifying that region B needs improved training or resources.

2.2.2 Augmented model

The RE model from Section 2.1.2, also referred to as a random intercept model, will now be augmented by introducing the vector of provider attributes, $\boldsymbol{z_i}$, and associated unknown risk parameters, $\boldsymbol{\gamma}$. It is important to note that underlying, significant cross-level interactions between patient and provider characteristics may also be present. For example, Kasza et al. (2013) observed that, all other things being equal, mortality differed significantly between hospital types (tertiary/teaching, metropolitan, rural/regional, and private) for each patient surgical status (elective, emergency, and non-surgical). However, further consideration of cross-level interactions is deferred to Section 2.3.

It is now assumed that

$$Y_{ij} \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \varepsilon_i \overset{ind}{\sim} Bern\left(\pi_{ij}\right),$$

where

$$\text{logit}\left(\pi_{ij}\right) \;=\; \mu + \boldsymbol{\alpha^T x_{ij}} + \boldsymbol{\gamma^T z_i} + \varepsilon_i.$$

It follows that the PMF of Y_{ij} is

$$p\left(y_{ij} \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \varepsilon_i\right) \;=\; \frac{e^{\left(\mu + \boldsymbol{\alpha^T x_{ij}} + \boldsymbol{\gamma^T z_i} + \varepsilon_i\right) y_{ij}}}{1 + e^{\mu + \boldsymbol{\alpha^T x_{ij}} + \boldsymbol{\gamma^T z_i} + \varepsilon_i}}, \qquad (y_{ij} = 0, 1)$$

and the posterior distribution of ε_i has the PDF

$$f\left(\varepsilon_i \,|\, \boldsymbol{y_i}, X_i, \boldsymbol{z_i}\right) \;=\; k_i \,\exp\left\{-\frac{\varepsilon_i^2}{2\tau^2} - \sum_{l=1}^{n_i}\left[\log\left(1 + e^{\mu + \boldsymbol{\alpha^T x_{il}} + \boldsymbol{\gamma^T z_i} + \varepsilon_i}\right) - \varepsilon_i y_{il}\right]\right\},$$

where the normalising constant

$$k_i \;=\; \left[\int_{-\infty}^{\infty} \exp\left\{-\frac{\varepsilon_i^2}{2\tau^2} - \sum_{l=1}^{n_i}\left[\log\left(1+e^{\mu+\boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{x_{il}}+\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z_i}+\varepsilon_i}\right) - \varepsilon_i y_{il}\right]\right\} \mathrm{d}\varepsilon_i\right]^{-1}.$$

2.2.3 Standard approach, actual comparator

The standard approach, with actual comparator, for calculating the SMR of provider i is

$$SMR_i \;=\; \frac{\sum_{j=1}^{n_i} Y_{ij}}{\sum_{j=1}^{n_i} E_{ij}},$$

where the numerator is the observed number of deaths and the denominator is the expected number of deaths based on the provider's actual attributes and its patient casemix. The expected outcomes,

$$E_{ij} \;=\; p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}\right),$$

are the marginal probabilities predicted from the fitted model. The E_{ij} are obtained by marginalising $p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \varepsilon_i\right)$ over the prior distribution of ε_i, thus

$$\begin{aligned}
E_{ij} \;&=\; \int_{-\infty}^{\infty} p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \varepsilon_i\right)\, f\left(\varepsilon_i\right)\, \mathrm{d}\varepsilon_i \\
&=\; \int_{-\infty}^{\infty} \frac{e^{\mu+\boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{x_{ij}}+\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z_i}+\varepsilon_i}}{1+e^{\mu+\boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{x_{ij}}+\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z_i}+\varepsilon_i}} \times \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{\varepsilon_i^2}{2\tau^2}\right\} \mathrm{d}\varepsilon_i \\
&=\; \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\tau^2}} \left(1+e^{-\mu-\boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{x_{ij}}-\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z_i}-\varepsilon_i}\right)^{-1} \exp\left\{-\frac{\varepsilon_i^2}{2\tau^2}\right\} \mathrm{d}\varepsilon_i \\
&=\; \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\log\left(2\pi\tau^2\right) - \frac{\varepsilon_i^2}{2\tau^2} - \log\left(1+e^{-\mu-\boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{x_{ij}}-\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z_i}-\varepsilon_i}\right)\right\} \mathrm{d}\varepsilon_i. \quad (2.1)
\end{aligned}$$

For provider i, $SMR_i = 1$ indicates performance as expected given its casemix of patients and actual provider attributes; $SMR_i < 1$ indicates better-than-expected performance; and conversely, $SMR_i > 1$ indicates worse-than-expected performance. If the modelling contains no provider-level fixed effects, then $\boldsymbol{z_i}$ and hence $\boldsymbol{\gamma}$ are omitted from the model. In doing so, the current (augmented) model reduces to the RE model of Section 2.1.2. Furthermore, the provider's attributes no longer enter the SMR comparator (via E_{ij}) so the comparator shifts from actual to average. Thus, provider i's performance would be compared with that of an average provider, given the same patient casemix. A bootstrapping procedure can be used to account for the uncertainty associated with estimating SMRs.

The above formula for the SMR considers, for a given provider, the expected mortalities for its own casemix of patients, based on the outcomes observed throughout the entire study population (Pouw et al., 2013). Such SMRs are essentially calculated by indirect standardisation (Moran and Solomon, 2014). The alternative type of standardisation is direct, where each provider's performance is assessed with respect to a common, reference casemix of patients (Pouw et al., 2013). This presents

the challenge of selecting an appropriate reference casemix. From a practical viewpoint, a common subset of the entire population is generally inappropriate because models that adequately adjust for mortality risks are likely to contain a large number of variables, and some subcategories are likely to have insufficient patient numbers (Pouw et al., 2013). Another option would be to select the entire population as the reference; examples of such implementations are in Silber et al. (2016) and Lindmark et al. (2016). However, it is advantageous to instead restrict each provider's comparison to their own casemix, and this is followed throughout the thesis. It avoids the likely scenario of performance being over-assessed on the types of patients that an individual provider rarely treats due to such patient types being frequently treated by most of the other providers (Heijink et al., 2008).

2.2.3.1 Simple approximation to the expected outcome

The intractable integral contained in Eq. 2.1 encumbers the evaluation of E_{ij}. However, the Eq. 2.1 representation is preferred in this thesis because efficient estimates can be obtained; Chapter 3 will describe a technique for approximating such integrals. The current chapter now describes an alternative, equivalent expression for E_{ij} which uses the direct definition of the expected outcome,

$$E_{ij} \;=\; E\left[\left(1+e^{-\mu-\boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{x_{ij}}-\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z_i}-\varepsilon_i}\right)^{-1}\right], \tag{2.2}$$

where the expectation integrates over the prior distribution of ε_i, as before. This expression could be evaluated by Monte Carlo simulation. However, the expected value representation in Eq. 2.2 is expressible as an infinite series, and this enables a much more efficient method of calculation. By denoting

$$\eta_{ij} \;=\; \mu+\boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{x_{ij}}+\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z_i}$$

and

$$g\left(\eta_{ij}\right) \;=\; \left(1+e^{-\eta_{ij}}\right)^{-1},$$

then

$$E_{ij} \;=\; E\left[\left(1+e^{-\eta_{ij}-\varepsilon_i}\right)^{-1}\right]. \tag{2.3}$$

Using a Taylor series expansion for $(1+e^{-\eta_{ij}-\varepsilon_i})^{-1}$, Eq. 2.3 becomes

$$\begin{aligned}
E_{ij} \;&=\; E\left[\sum_{k=0}^{\infty}\frac{1}{k!}\,g^{(k)}(\eta_{ij})\,\varepsilon_i^k\right] \\
&=\; \sum_{k=0}^{\infty}\left\{\frac{1}{k!}\,g^{(k)}(\eta_{ij})\,E\left[\varepsilon_i^k\right]\right\} \\
&=\; \sum_{k=0}^{\infty}\left\{\frac{1}{(2k)!}\,g^{(2k)}(\eta_{ij})\,E\left[\varepsilon_i^{2k}\right]+\frac{1}{(2k+1)!}\,g^{(2k+1)}(\eta_{ij})\,E\left[\varepsilon_i^{2k+1}\right]\right\} \\
&=\; \sum_{k=0}^{\infty}\left\{\frac{1}{(2k)!}\,g^{(2k)}(\eta_{ij})\times\frac{(2k)!\,\tau^{2k}}{2^k\,k!}+\frac{1}{(2k+1)!}\,g^{(2k+1)}(\eta_{ij})\times 0\right\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(using Appendix A.1)} \\
&=\; \sum_{k=0}^{\infty}\frac{\tau^{2k}}{2^k\,k!}\,g^{(2k)}(\eta_{ij})\,.
\end{aligned}$$

A simple, computationally efficient, alternative approximation to

$$E_{ij} \;=\; \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\log\left(2\pi\tau^2\right) - \frac{\varepsilon_i^2}{2\tau^2} - \log\left(1+e^{-\eta_{ij}-\varepsilon_i}\right)\right\} \mathrm{d}\varepsilon_i \qquad (2.4)$$

follows from considering the fixed effects only. Glance et al. (2003) adopted this approach, effectively setting all random effects to their expected value of zero when calculating the expected deaths, thus approximating E_{ij} by $g\left(\eta_{ij}\right)$. However,

$$E_{ij} \;\neq\; p\left(y_{ij}=1 \,|\, \eta_{ij}, \varepsilon_i = 0\right) \;=\; g\left(\eta_{ij}\right)$$

because the expected value of a nonlinear function generally differs from a nonlinear function of the expected value. With the hierarchical model presented in Glance et al. (2003) containing multiple random effects (an intercept and slope), the accuracy of their approximations will be examined in Section 2.3.2.1 after the necessary expansions to the model have been discussed. The bias of this approximation, using the current, random intercept model, is

$$h\left(\eta_{ij}\right) \;=\; g\left(\eta_{ij}\right) - E_{ij} \;=\; -\sum_{k=1}^{\infty} \frac{\tau^{2k}}{2^k k!}\, g^{(2k)}(\eta_{ij})\,. \qquad (2.5)$$

The above function reveals a direct relationship between the bias and τ^2: the bias amplifies as τ^2 increases, and attenuates as τ^2 decreases. Not surprisingly, the random effect essentially disappears if τ^2 is zero, in which case, the simple approximation is the analytical solution. Irrespective of τ^2, there is no bias when η_{ij} is zero, and the proof is given in Appendix A.2 that h is an odd function.

2.2.3.2 Illustrative example of the accuracy achievable

The random intercept model fitted in Moran and Solomon (2014) will now be used to illustrate the potential accuracy that may be achieved by implementing the simple approximation technique ($\varepsilon = 0$); however, it is important to note that Moran and Solomon (2014) did not actually use this technique in their work. For their fitted, random intercept model, Moran and Solomon (2014, Table S2) reported that the ICU site random effect had an estimated standard deviation (i.e. $\hat{\tau}$) of 0.303. This illustration sets τ^2 to the value of the parameter estimated by Moran and Solomon (2014), that is, $0.303^2 = 0.091809$.

For various η_{ij}, the intractable integrals from Eq. 2.4 will be calculated using adaptive quadrature (AQ) via R's `integrate` function (R Core Team, 2017). In evaluating each of these integrals, the `integrate` function firstly transforms the integrand, mapping the region $(-\infty, \infty)$ onto $(-1, 1)$. The accuracy of an AQ estimate to a (usually unknown) solution may be assessed in terms of absolute and relative errors. These errors indicate the (approximate) accuracy in terms of decimal places and significant digits, respectively. The `integrate` function uses a combination of the absolute and relative tolerances (`abs.tol` and `rel.tol`, respectively), that is, the maximum of `abs.tol` and the product of the `rel.tol` and the estimated magnitude of the integral. Both `abs.tol` and `rel.tol` are set to 10^{-12} for every AQ computation, and it is assumed that the AQ estimates yielded from enforcing such stringent tolerances are essentially the true values. Since each integral represents a probability, its value ranges between zero and one. Hence, the accuracy requested of the `integrate` function is approximately twelve decimal places.

Figure 2.1a displays the plot of the expected outcomes (mortality probabilities) associated with a rather large range of values for η. Figure 2.1b displays the plot of the bias function corresponding to the use of the simple technique to approximate the expected outcomes. This thesis concentrates on actual rather than relative differences (ratio of actual differences to expected outcomes) when reporting on biases because the use of actual differences is better suited to the task of estimating SMRs. Calculating such bias from estimating the total expected outcomes at a provider (i.e., their expected deaths) involves summing the associated biases from estimating the expected outcome for each of their patients. The simplified approximation technique would be reasonable if the total bias was sufficiently close to zero to negligibly impact the estimated SMR for each provider. Noting the direct relationship between bias and τ^2, every bias approaches zero as τ^2 approaches zero. For larger values of τ^2, the biases associated with estimating each provider's expected deaths may roughly cancel out, though this seems unlikely. Each provider's casemix would need to represent some group of patients whose actual mortality probabilities are spread reasonably well across the valid range.

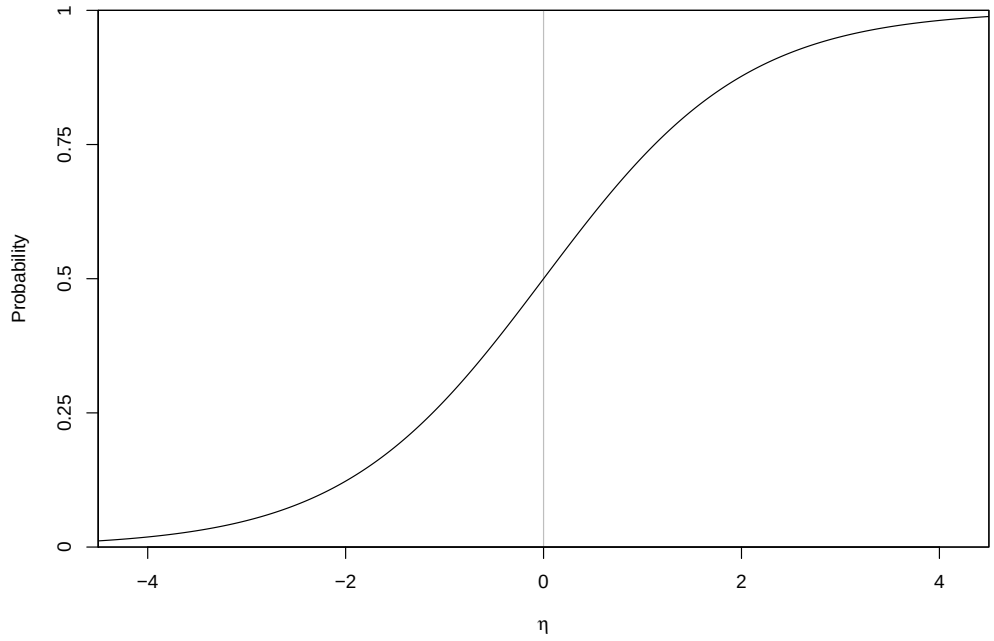
Figure 2.1b verifies that the bias is an odd function of η and passes through the origin. The approximations are unbiased when η is zero; this corresponds to the mortality probability of one half, as shown in Figure 2.1a. When η is positive, the approximations consistently overestimate the mortality risks attributed to patients; this corresponds to mortality probabilities above one half. Conversely, when η is negative, the approximations consistently underestimate the mortality risks attributed to patients; this corresponds to mortality probabilities below one half. Inspection of Figure 2.1b reveals that the bias function has two stationary points, that is, where the bias has largest magnitude. The stationary points are at $(-1.3425, -0.0043)$ and $(1.3425, 0.0043)$, which correspond to the expected outcomes of 0.2114 and 0.7886, respectively, in Figure 2.1a. Since the biases are, overall, quite small, use of the simplified approximation technique may be reasonable here.

2.2.4 Smoothed approach, actual comparator

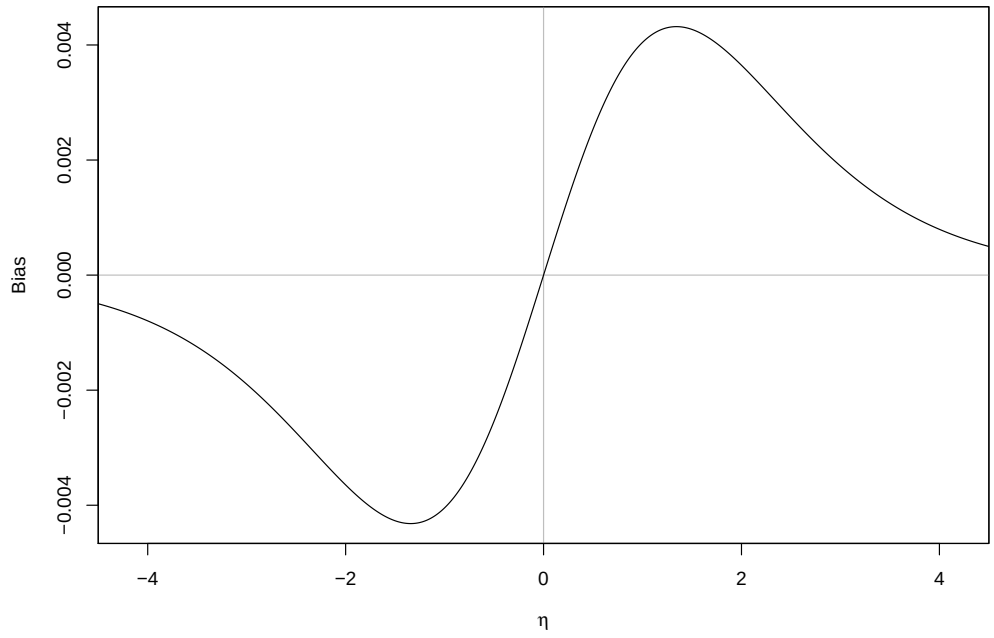
An alternative approach of constructing SMRs involves replacing the observed outcomes by "smoothed", model-based predictions (Shahian et al., 2005). The rationale behind this replacement is the attempt to address the sampling error in the observed deaths (Bottle and Aylin, 2016, p. 128), creating more meaningful estimates, particularly for low-volume providers. Estimates are shrunken towards a provider-attribute specific mean, given the patient casemix, by an amount that is dependent on the sampling variability for specific parameters (Normand et al., 2016). As a result, the estimates obtained will have smaller variances. In advocating this approach, Normand et al. (2016) suggested that, based on an approximate 15% mortality rate for patients admitted to hospital due to AMI, it would be inappropriate to report a hospital's mortality rate as 0%, 50%, or 100% if they only admitted two patients with that particular condition.

Retaining use of the actual comparator from Section 2.2.3, the smoothed approach, as it will be referred to in this thesis, for calculating the SMR of provider i is

$$SMR_i = \frac{\sum_{j=1}^{n_i} S_{ij}}{\sum_{j=1}^{n_i} E_{ij}},$$



(a) Expected outcome (mortality probability)



(b) Bias of the simple approximation to the expected outcome

Figure 2.1: Evaluation of, and approximation to, the expected outcomes based on the fitted, random intercept model in Moran and Solomon (2014).

where the numerator becomes the "smoothed" observed number of deaths based on that provider's actual attributes and patient casemix. The smoothed outcomes,

$$S_{ij} \;=\; p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}\right),$$

are the posterior mean probabilities predicted using the knowledge about ε_i for provider i, which is represented by the posterior distribution (Rabe-Hesketh and Skrondal, 2012, p. 551). The S_{ij} are obtained by marginalising $p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \varepsilon_i\right)$ over the posterior distribution of ε_i, thus

$$\begin{aligned}
S_{ij} \;&=\; \int_{-\infty}^{\infty} p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \varepsilon_i\right)\; f\left(\varepsilon_i \,|\, \boldsymbol{y_i}, X_i, \boldsymbol{z_i}\right)\, \mathrm{d}\varepsilon_i \\
&=\; \int_{-\infty}^{\infty} \left(1+e^{-\mu-\boldsymbol{\alpha^T x_{ij}}-\boldsymbol{\gamma^T z_i}-\varepsilon_i}\right)^{-1} \\
&\qquad\qquad \times k_i \,\exp\left\{-\frac{\varepsilon_i^2}{2\tau^2}-\sum_{l=1}^{n_i}\left[\log\left(1+e^{\mu+\boldsymbol{\alpha^T x_{il}}+\boldsymbol{\gamma^T z_i}+\varepsilon_i}\right)-\varepsilon_i y_{il}\right]\right\}\,\mathrm{d}\varepsilon_i \\
&=\; k_i \int_{-\infty}^{\infty} \exp\left\{-\frac{\varepsilon_i^2}{2\tau^2}-\sum_{l=1}^{n_i}\left[\log\left(1+e^{\mu+\boldsymbol{\alpha^T x_{il}}+\boldsymbol{\gamma^T z_i}+\varepsilon_i}\right)-\varepsilon_i y_{il}\right]\right. \\
&\qquad\qquad \left. -\log\left(1+e^{-\mu-\boldsymbol{\alpha^T x_{ij}}-\boldsymbol{\gamma^T z_i}-\varepsilon_i}\right)\right\}\,\mathrm{d}\varepsilon_i,
\end{aligned}$$

where

$$k_i \;=\; \left[\int_{-\infty}^{\infty} \exp\left\{-\frac{\varepsilon_i^2}{2\tau^2}-\sum_{l=1}^{n_i}\left[\log\left(1+e^{\mu+\boldsymbol{\alpha^T x_{il}}+\boldsymbol{\gamma^T z_i}+\varepsilon_i}\right)-\varepsilon_i y_{il}\right]\right\}\,\mathrm{d}\varepsilon_i\right]^{-1}.$$

The interpretation of the SMRs is unchanged from Section 2.2.3. As before, if the modelling contains no provider-level fixed effects, then the comparator is an average provider instead. This is essentially the method applied by the CMS (Ash et al., 2012), which will be discussed further in Section 2.2.5.

Sections 2.2.3.1 and 2.2.3.2 showed that it may not be sufficiently accurate to approximate the E_{ij} by simply evaluating the mortality probabilities at the mean of the prior distribution of the random effects (i.e., setting the ε_i to zero). Similar inaccuracies could also affect the S_{ij} if each mortality probability is only evaluated at the mean of the posterior distribution of ε_i (Rabe-Hesketh and Skrondal, 2012, pp. 551 – 552). Nevertheless, it seems typical for statistical software to restrict the reported probabilities to having random effects set to the mean or mode of their posterior distributions. Perhaps the computational intensity that would otherwise be involved with integration contributes to the preference for simpler evaluations.

2.2.5 Smoothed approach, average comparator

Silber et al. (2016) criticised the current CMS method, arguing that it holds the "common misconception" that the inclusion of a random effect for hospitals in the model will adjust sufficiently for all hospital characteristics. In a report commissioned by the CMS, Ash et al. (2012) suggested generalising the current CMS method to include provider-level (fixed) variables in such a way that it would still uphold its mandate towards the use of a national-level comparator. Much of the content in

Ash et al. (2012) was subsequently published by the same authors in Normand et al. (2016). However, both sources recommended that further evaluation is necessary before deciding which, and even if, hospital-level attributes ought to be included in the model. Furthermore, the CMS outcome measure methodology still excludes provider-level variables (CMS, 2019*a*), and the generalisation described below has not been applied elsewhere in the literature.

The general formula of the SMR for provider i is

$$SMR_i \;=\; \frac{\sum_{j=1}^{n_i} S_{ij}}{\sum_{j=1}^{n_i} E_{ij}},$$

as before, though the interpretation of the SMRs differs insofar as an individual provider's performance is now being compared to that of an average provider with the identical patient casemix. The model fit, and hence the smoothed outcomes, are likely to be improved by the inclusion of provider attributes; calculation of the S_{ij} is influenced by provider i's attributes, through $\boldsymbol{z_i}$. Calculating the E_{ij} is considerably more complicated than before because it now requires prior identification of the attributes of an average provider, denoted $\boldsymbol{z^*}$ (Normand et al., 2016). It is important to note, however, that provider i's attributes do not appear in the calculation of the comparator (Normand et al., 2016); the E_{ij} are produced without influence from the $\boldsymbol{z_i}$. The only difference from Section 2.2.3 in the expression for E_{ij} is that $\boldsymbol{z_i}$ has been changed to $\boldsymbol{z^*}$. Now

$$E_{ij} \;=\; p\left(y_{ij} = 1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z^*}\right),$$

which is obtained by marginalising $p\left(y_{ij} = 1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z^*}, \varepsilon_i\right)$ over the prior distribution of ε_i. It follows that

$$E_{ij} \;=\; \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\log\left(2\pi\tau^2\right) - \frac{\varepsilon_i^2}{2\tau^2} - \log\left(1 + e^{-\mu - \boldsymbol{\alpha^T x_{ij}} - \boldsymbol{\gamma^T z^*} - \varepsilon_i}\right)\right\} \mathrm{d}\varepsilon_i. \quad (2.6)$$

The vector representing an average provider, $\boldsymbol{z^*}$, is obtained[4] by equating

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i} Y_{ij} \qquad \text{to} \qquad \sum_{i=1}^{I}\sum_{j=1}^{n_i} E_{ij}. \quad (2.7)$$

Thus, the total number of mortalities observed at all providers is equated with the expected number of mortalities, on the supposition that every patient had instead been treated at an average provider. In other words, the average expected mortality from all

$$N \;=\; \sum_{i=1}^{I} n_i$$

patients is the raw mortality rate. If the model contains multiple provider-level variables, then an infinite number of average providers exist. However, the provider-level contribution to $\mathrm{logit}(\pi_{ij})$ by any two average providers, denoted $\boldsymbol{z_a^*}$ and $\boldsymbol{z_b^*}$, is

[4]This allows for the missing denominator, $\sum_{i=1}^{I} n_i$, on the right-hand side of Eq. (6) in both Ash et al. (2012, p. 57) and Normand et al. (2016, p. 37).

the same: $\boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z}_{\boldsymbol{a}}^{*} = \boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z}_{\boldsymbol{b}}^{*} = \zeta$. It is therefore much more efficient to find the one-dimensional ζ by substituting $\zeta = \boldsymbol{\gamma}^{\boldsymbol{T}}\boldsymbol{z}^{*}$ into Eq. 2.6 then using it to solve Eq. 2.7. Whichever $\boldsymbol{z}^{*}$ is used, the log odds of mortality for patient j in provider i, $\text{logit}(\pi_{ij})$, and hence that patient's expected outcome, E_{ij}, does not change. Thus, the total expected outcomes, $\sum_{j=1}^{n_i} E_{ij}$, and therefore the corresponding SMR for provider i cannot vary between particular $\boldsymbol{z}^{*}$ either.

Constraints on average provider attributes can render many $\boldsymbol{z}^{*}$ which satisfy Eq. 2.7 as infeasible. Proportionality constraints exemplify this restriction. Take the case of the model containing only one hospital-level predictor, size, with three levels: small, medium, and large. Suppose that size is categorical, with small as the reference level, and the indicator variables z_1 and z_2 represent whether hospital i is medium and large, respectively. Thus, the characteristics of hospital i are indicated by the vector $\boldsymbol{z_i} = (z_{1i}, z_{2i})^{T}$. Furthermore, the elements of $\boldsymbol{z}^{*}$ will indicate what proportion of each size hospital comprises an average one, with the proportion of each hospital's size ranging between zero and one. Feasible $\boldsymbol{z}^{*} = [z_1^{*}, z_2^{*}]^{T}$ are constrained by the conditions $z_1^{*} + z_2^{*} \leq 1$ and $z_1^{*}, z_2^{*} \geq 0$; the proportion of the reference level comprising an average hospital is $1 - z_1^{*} - z_2^{*}$. If a valid $\boldsymbol{z}^{*}$ was $[0.5, 0.3]^{T}$, then it would indicate that an average hospital was 20% small, 50% medium, and 30% large.

It was noted with justification in Section 2.2.3 that indirect standardisation will be used in the calculation of all SMRs in this thesis. The alternative, direct standardisation was advocated and utilised by Silber et al. (2016) in their investigation of hospital quality. Although hospital-level variables were included in the modelling by Silber et al. (2016), they avoided approximating SMRs and average hospital characteristics, instead approximating the posterior distribution of each hospital's average mortality rate as if that hospital had treated every patient in the study population. Based on the notation and terminology of this thesis, Silber et al. (2016) approximated the distribution of the directly standardised mortality rate for hospital h by evaluating

$$\frac{1}{N}\sum_{i=1}^{I}\sum_{j=1}^{n_i} p\left(y_{ij} \mid \boldsymbol{x_{ij}}, \boldsymbol{z_h}, \varepsilon_h\right)$$

using 10,000 random samples from the posterior distribution of ε_h. This approach contrasts with what was described earlier in this section. An average provider was effectively defined as one whose patient casemix and raw mortality rate were identical to those of the entire study population. This can be verified by dividing both sides of Eq. 2.7 by N. Through the use of indirect standardisation, it is necessary to define an average provider because the evaluation of each provider's smoothed and expected outcomes (the SMR's components) is restricted to that provider's own patient casemix.

2.2.6 Standard approach, average comparator

The remaining SMR combination of the two approaches, standard and smoothed, and the two comparators, actual and average, will now be briefly discussed. Using an average comparator, the standard approach for calculating the SMR of provider i is

$$SMR_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{\sum_{j=1}^{n_i} E_{ij}},$$

where the numerator is obtained as in Section 2.2.3, but there are multiple ways of obtaining the denominator. If the model contained provider-level effects, then the denominator can be calculated as in Section 2.2.5. This would be appropriate if one wanted to compare the SMRs obtained by this method with those obtained by the other three methods because no further calculations are required. However, solely using this SMR method in an application would justify the implementation of a simpler, alternative procedure. Since the provider's specific attributes do not enter the SMR formula, it would be more efficient to fit the reduced model containing no provider-level effects. SMRs would then be calculated as in Section 2.2.3, after having omitted all $\boldsymbol{\gamma}$ and $\boldsymbol{z_i}$ terms beforehand. This is essentially the method applied by the ICNARC (2019).

2.2.7 Estimating SMR uncertainty

Thus far, four inter-related methods for calculating the point estimates of SMRs have been discussed, though the uncertainty associated with them has yet to be addressed. Chance alone may explain why some providers' estimated SMRs appear to differ remarkably from the benchmark of one (indicative of expected performance). The statistical literature contains various approaches to estimating SMR uncertainty, especially when SMRs are calculated by a standard approach (ratio of the observed to expected deaths). Bottle and Aylin (2016, p. 125) noted that the random variability in the expected deaths has typically been ignored, and the observed deaths have often been assumed to follow a Poisson distribution. It was demonstrated by Kasza et al. (2013), however, that assuming the expected deaths were fixed could lead to substantial variability being ignored.

Attempts to estimate SMR uncertainty are further complicated when SMRs have been calculated by a smoothed approach (ratio of the smoothed observed to expected deaths). The numerators for those SMRs are (almost surely) non-integers, and the expressions contain intractable integrals. The bootstrap is suitable for the task of estimating SMR uncertainty, including for the four SMR methods that have been presented in this section. As Efron and Tibshirani (1993, p. 45) stated, the bootstrap is available irrespective of the complicated expressions involved.

This thesis recommends the use of a non-parametric bootstrap procedure, a "cases bootstrap" (van der Leeden et al., 2008, pp. 413 – 414). Before discussing the procedure, some assumptions and notational simplifications are stated here. For convenience, it will be assumed that the data were fitted to a model with the general structure described in Section 2.2.2. Thus, the estimated parameters of the fitted model ($\widehat{\mu}$, $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\gamma}}$, and $\widehat{\tau}^2$) have already been obtained, and the SMRs were then estimated (the $\widehat{SMR_i}$) by one of the methods described in Sections 2.2.3 to 2.2.6. The (natural) logarithm of the SMRs will be taken as being approximately normally distributed (Hosmer and Lemeshow, 1995), consistent with numerous studies in the medical and statistical literature, including Kasza et al. (2013, 2018) and Solomon et al. (2014). Indeed, the use of a logarithmic scale seems intuitively reasonable due

to the inverse scale of SMR formulae.[5] The logarithm of the SMR will be denoted LSMR, so the SMR of provider i on the logarithmic scale is $LSMR_i = \log(SMR_i)$. For LSMRs, the benchmark of zero is indicative of expected performance; negative and positive LSMRs are indicative of better-than-expected and worse-than-expected performance, respectively.

A sufficiently large number of bootstrap iterations, B, must be selected (a priori) in order to obtain reliable estimates of the standard errors. Efron and Tibshirani (1993, p. 52) indicated that $B = 200$ would usually suffice for estimating standard errors; however, the application in Chapter 4 will set $B = 1000$. For each bootstrap iteration, $b = 1, \ldots, B$, repeat the following steps of the bootstrap procedure.

1. For each provider, $i = 1, \ldots, I$, sample n_i cases (patient admission records) with replacement, and merge them into bootstrap dataset S_b. There is no resampling at the provider level, and dataset S_b has exactly the same size as the original dataset.

2. Fit the selected model to dataset S_b, and obtain bootstrap estimated parameters $\widehat{\mu}^{(b)}$, $\widehat{\boldsymbol{\alpha}}^{(\boldsymbol{b})}$, $\widehat{\boldsymbol{\gamma}}^{(\boldsymbol{b})}$, and $\widehat{\tau}^{2(b)}$.

3. Use the bootstrap dataset and estimated parameters to calculate, for each provider i, the bootstrap estimated SMR on the logarithmic scale,

$$\widehat{LSMR_i}^{(b)} = \log\left(\widehat{SMR_i}^{(b)}\right).$$

Hence, an approximate $(1-\alpha)\times 100\%$ confidence interval (CI) for the SMR of provider i on the logarithmic scale, $LSMR_i$, is

$$\widehat{LSMR_i} \;\pm\; z_{\alpha/2} \times \widehat{SE}_{boot}\left[\widehat{LSMR_i}\right],$$

where the bootstrap estimated standard error (SE) is

$$\widehat{SE}_{boot}\left[\widehat{LSMR_i}\right] \;=\; \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left[\widehat{LSMR_i}^{(b)} - \widehat{LSMR_i}^{(\bullet)}\right]^2},$$

the bootstrap estimated mean is

$$\widehat{LSMR_i}^{(\bullet)} \;=\; \frac{1}{B}\sum_{b=1}^{B}\widehat{LSMR_i}^{(b)},$$

and the upper $100\times\alpha/2$ percentile of the standard normal distribution is $z_{\alpha/2}$. It follows that an approximate $(1-\alpha)\times 100\%$ CI for SMR_i is

$$\left(e^{\widehat{LSMR_i} - z_{\alpha/2}\times\widehat{SE}_{boot}\left[\widehat{LSMR_i}\right]},\;\; e^{\widehat{LSMR_i} + z_{\alpha/2}\times\widehat{SE}_{boot}\left[\widehat{LSMR_i}\right]}\right).$$

[5] For example, a provider had twice as many deaths as expected if $SMR = 2$, but one half as many deaths as expected if $SMR = 1/2$.

2.3 Further expansions

Until now, this chapter has restricted the (two-level) models to containing no cross-level terms and (a maximum of) one random component, the intercept. This section will consider issues surrounding possible model expansions in the context of assessing provider performance. A more general, random coefficient model will also be detailed. Subsequently, an example of its application is discussed, along with many of the issues surrounding the implementation of these expanded models.

2.3.1 General considerations

The model used for estimating SMRs needs to be fit for purpose. It must adequately and appropriately adjust for patient risk. For example, adjusting for in-hospital infections may improve the model fit, but (as mentioned in this chapter's preamble) is certainly inappropriate. Still, the residuals from the fitted model must be compatible with the modelling assumptions. This can be checked by examining the posterior means of the random effects (Snijders and Bosker, 1999, p. 132).

Allowing for significant cross-level interactions may improve the model fit; for example, the interaction between hospital level and patient surgical status discussed in Section 2.2.2. How cross-level interactions are dealt with depends on the type of SMR required from a study. If actual comparators were required, then cross-level interactions could be included in the model without causing any problems. In the case of average comparators, provider type is not a predictor, nor are any interactions between provider type and patient characteristics. However, provider type may be used in smoothing the observed outcomes and defining an average provider. As a result, cross-level interactions cannot be included in SMR studies that involve average comparators. Now suppose that it is necessary to include cross-level interactions for an appropriately fitted model to be found. If a study required average comparators, then separate models could be used. This may be based on specific types of patients by illness category (e.g. restrict one model to patients with cardiovascular illnesses, another to patients with respiratory illnesses, etc.) or specific hospital types (e.g. restrict one model to tertiary hospitals, another to private hospitals, etc.). With certain investigations, it may even be a good starting point to limit the focus to one particular illness.

Further modelling continues to exclude cross-level interactions to avoid excessive algebraic details. However, the principles could still be applied as necessary in the case of actual comparators.

2.3.2 A general random coefficient model

Letting $x_1, \ldots, x_p$ be the p patient-level variables in the general model of Section 2.2.2, then $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^T$ is the corresponding vector of fixed coefficients. Thus, $\boldsymbol{x_{ij}} = [x_{1ij}, \ldots, x_{pij}]^T$ is the p-dimensional vector of patient attributes associated with patient j in provider i. The model is now expanded to allow q ($\leq p$) patient-level variables to have random coefficients. Without loss of generality, the first q patient-level variables, from x_1 to x_q, are assigned both fixed and random coefficients, while the remaining $p - q$ patient-level variables, from x_{q+1} to x_p, are assigned fixed coefficients only. For notational convenience, let $\boldsymbol{x^r_{ij}} = [1, x_{1ij}, \ldots, x_{qij}]^T$, a $(q+1)$-dimensional vector of patient attributes, where the first q elements of $\boldsymbol{x_{ij}}$ are the last

q elements (from 2 to $q+1$) of $\boldsymbol{x}^{\boldsymbol{r}}_{\boldsymbol{ij}}$. As will be shown below, the first element of $\boldsymbol{x}^{\boldsymbol{r}}_{\boldsymbol{ij}}$ accommodates the random intercept for provider i.

It is now assumed that

$$Y_{ij} \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \boldsymbol{\varepsilon_i} \overset{ind}{\sim} Bern\left(\pi_{ij}\right),$$

where

$$\boldsymbol{\varepsilon_i} = \left[\varepsilon_{0i},\, \varepsilon_{1i},\, \ldots,\, \varepsilon_{qi}\right]^T \overset{iid}{\sim} N_{q+1}\left(\mathbf{0}, \Upsilon\right),$$

with the variance-covariance matrix

$$\Upsilon = \begin{bmatrix} \tau_0^2 & \tau_{01} & \cdots & \tau_{0q} \\ \tau_{01} & \tau_1^2 & \cdots & \tau_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{0q} & \tau_{1q} & \cdots & \tau_q^2 \end{bmatrix},$$

and the log odds of mortality for patient j in provider i is now

$$\operatorname{logit}\left(\pi_{ij}\right) \;=\; \mu + \boldsymbol{\alpha^T x_{ij}} + \boldsymbol{\gamma^T z_i} + \boldsymbol{\varepsilon_i^T x^r_{ij}}.$$

The random intercept term from the previous model (ε_i) is represented by ε_{0i} in the expanded model. Similarly to before, the PMF of Y_{ij} is now

$$p\left(y_{ij} \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \boldsymbol{\varepsilon_i}\right) \;=\; \frac{e^{\left(\mu + \boldsymbol{\alpha^T x_{ij}} + \boldsymbol{\gamma^T z_i} + \boldsymbol{\varepsilon_i^T x^r_{ij}}\right) y_{ij}}}{1 + e^{\mu + \boldsymbol{\alpha^T x_{ij}} + \boldsymbol{\gamma^T z_i} + \boldsymbol{\varepsilon_i^T x^r_{ij}}}}.$$

Furthermore, the prior distribution of $\boldsymbol{\varepsilon_i}$ has the PDF

$$f\left(\boldsymbol{\varepsilon_i}\right) \;=\; \frac{1}{(2\pi)^{\frac{q+1}{2}} \left|\Upsilon\right|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon_i^T}\Upsilon^{-1}\boldsymbol{\varepsilon_i}\right\},$$

so it follows from the previous derivations given in Section 2.1.2 that the posterior distribution of $\boldsymbol{\varepsilon_i}$ has the PDF

$$f\left(\boldsymbol{\varepsilon_i} \,|\, \boldsymbol{y_i}, X_i, \boldsymbol{z_i}\right) \;=\; k_i \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon_i^T}\Upsilon^{-1}\boldsymbol{\varepsilon_i} - \sum_{l=1}^{n_i}\left[\log\left(1 + e^{\mu + \boldsymbol{\alpha^T x_{il}} + \boldsymbol{\gamma^T z_i} + \boldsymbol{\varepsilon_i^T x^r_{il}}}\right) - \boldsymbol{\varepsilon_i^T x^r_{ij}} y_{il}\right]\right\},$$

where the normalising constant

$$k_i \;=\; \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\boldsymbol{\varepsilon_i^T}\Upsilon^{-1}\boldsymbol{\varepsilon_i} - \sum_{l=1}^{n_i}\left[\log\left(1 + e^{\mu + \boldsymbol{\alpha^T x_{il}} + \boldsymbol{\gamma^T z_i} + \boldsymbol{\varepsilon_i^T x^r_{il}}}\right) - \boldsymbol{\varepsilon_i^T x^r_{ij}} y_{il}\right]\right\} \mathrm{d}\varepsilon_{0i} \ldots \mathrm{d}\varepsilon_{qi}\right]^{-1}.$$

Calculations of the SMRs by the four methods discussed in Section 2.2 are quite similar when the expanded model is used. The differences affect how the expected and smoothed outcomes (the E_{ij} and S_{ij}, respectively) are calculated; the interpretations of the SMRs remain unaffected. Expected outcomes based on actual comparators,

$$E_{ij} \;=\; p\left(y_{ij} = 1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}\right),$$

are now obtained by marginalising $p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \boldsymbol{\varepsilon_i}\right)$ over the prior distribution of $\boldsymbol{\varepsilon_i}$; using the work in deriving Eq. 2.1,

$$E_{ij} \;=\; \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2}\log\left[(2\pi)^{q+1}\,|\Upsilon|\right] - \frac{1}{2}\boldsymbol{\varepsilon_i^T}\Upsilon^{-1}\boldsymbol{\varepsilon_i} \right. \\ \left. - \log\left(1+e^{-\mu-\boldsymbol{\alpha^T x_{ij}}-\boldsymbol{\gamma^T z_i}-\boldsymbol{\varepsilon_i^T x_{ij}^r}}\right) \right\} \mathrm{d}\varepsilon_{0i} \ldots \mathrm{d}\varepsilon_{qi}. \tag{2.8}$$

Expected outcomes based alternatively on average comparators,

$$E_{ij} \;=\; p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z^*}\right),$$

are obtained by marginalising $p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z^*}, \boldsymbol{\varepsilon_i}\right)$ over the prior distribution of $\boldsymbol{\varepsilon_i}$; by changing $\boldsymbol{z_i}$ to $\boldsymbol{z^*}$ in Eq. 2.8,

$$E_{ij} \;=\; \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2}\log\left[(2\pi)^{q+1}\,|\Upsilon|\right] - \frac{1}{2}\boldsymbol{\varepsilon_i^T}\Upsilon^{-1}\boldsymbol{\varepsilon_i} \right. \\ \left. - \log\left(1+e^{-\mu-\boldsymbol{\alpha^T x_{ij}}-\boldsymbol{\gamma^T z^*}-\boldsymbol{\varepsilon_i^T x_{ij}^r}}\right) \right\} \mathrm{d}\varepsilon_{0i} \ldots \mathrm{d}\varepsilon_{qi},$$

where the vector representing an average provider, $\boldsymbol{z^*}$, is obtained (as before) by equating

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i} Y_{ij} \qquad \text{to} \qquad \sum_{i=1}^{I}\sum_{j=1}^{n_i} E_{ij}.$$

Meanwhile, smoothed outcomes,

$$S_{ij} \;=\; p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}\right),$$

are obtained by marginalising $p\left(y_{ij}=1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{z_i}, \boldsymbol{\varepsilon_i}\right)$ over the posterior distribution of $\boldsymbol{\varepsilon_i}$, irrespective of whether actual or average comparators are used. It follows that

$$S_{ij} \;=\; k_i \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2}\boldsymbol{\varepsilon_i^T}\Upsilon^{-1}\boldsymbol{\varepsilon_i} - \sum_{l=1}^{n_i}\left[\log\left(1+e^{\mu+\boldsymbol{\alpha^T x_{il}}+\boldsymbol{\gamma^T z_i}+\boldsymbol{\varepsilon_i^T x_{il}^r}}\right) \right.\right. \\ \left.\left. - \boldsymbol{\varepsilon_i^T x_{ij}^r} y_{il}\right] - \log\left(1+e^{-\mu-\boldsymbol{\alpha^T x_{ij}}-\boldsymbol{\gamma^T z_i}-\boldsymbol{\varepsilon_i^T x_{ij}^r}}\right) \right\} \mathrm{d}\varepsilon_{0i} \ldots \mathrm{d}\varepsilon_{qi}.$$

2.3.2.1 Application from Glance et al. (2003)

Section 2.2.3.1 stated that the accuracy of the approximations used by Glance et al. (2003) from their fitted, hierarchical model would be examined after the necessary expansions to the model were discussed. Their random coefficient model uses a function of the single, patient-level, predictor variable, the Simplified Acute Physiology Score (SAPS) II (Le Gall et al., 1993). SAPS II scores are integer predictors of mortality, ranging from 0 to 163, with higher scores indicative of higher mortality risks. However, the scores contained in the dataset used by Glance et al. (2003) reached a maximum of 125, with 30 being the median.

In terms of the notation used in this chapter, Glance et al. (2003) fitted the following hierarchical model. The log odds of mortality for patient j in hospital i is

$$\mathrm{logit}\left(\pi_{ij}\right) \;=\; \mu + \boldsymbol{\alpha^T x_{ij}} + \boldsymbol{\varepsilon_i^T x_{ij}^r},$$

where $p = 2$, $q = 1$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2]^T$, $\boldsymbol{x_{ij}} = [x_{1ij}, x_{2ij}]^T$, $\boldsymbol{x^r_{ij}} = [1, x_{1ij}]^T$, and $\boldsymbol{\varepsilon_i} = [\varepsilon_{0i}, \varepsilon_{1i}]^T \overset{iid}{\sim} N_2(\mathbf{0}, \Upsilon)$, with

$$\Upsilon = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}.$$

The patient-level variables are denoted x_1 for $\log(\text{SAPS II score} + 1)$ and x_2 for the (non-transformed) SAPS II score. The slope of x_1 is permitted to vary randomly between hospitals, while the slope of x_2 is fixed for all hospitals. Noting that $|\Upsilon| = \tau_0^2\tau_1^2 - \tau_{01}^2$ and

$$\Upsilon^{-1} = \frac{1}{|\Upsilon|} \begin{bmatrix} \tau_1^2 & -\tau_{01} \\ -\tau_{01} & \tau_0^2 \end{bmatrix},$$

the expected outcomes (using Eq. 2.8) are

$$E_{ij} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{ -\log\left(2\pi\sqrt{\tau_0^2\tau_1^2 - \tau_{01}^2}\right) - \frac{\varepsilon_{0i}^2\tau_1^2 - 2\varepsilon_{0i}\varepsilon_{1i}\tau_{01} + \varepsilon_{1i}^2\tau_0^2}{2\left(\tau_0^2\tau_1^2 - \tau_{01}^2\right)} - \log\left(1 + e^{-\mu - \alpha_1 x_{1ij} - \alpha_2 x_{2ij} - \varepsilon_{0i} - \varepsilon_{1i}x_{1ij}}\right) \right\} \mathrm{d}\varepsilon_{0i}\,\mathrm{d}\varepsilon_{1i} \qquad (2.9)$$

for particular SAPS II scores. This illustration sets the values of the parameters to the estimates reported by Glance et al. (2003), that is, $\mu = -8.100$, $\boldsymbol{\alpha} = [1.094, 0.063]^T$, and

$$\Upsilon = \begin{bmatrix} 2.412 & -0.680 \\ -0.680 & 0.196 \end{bmatrix}.$$

Similarly to the illustrative example in Section 2.2.3.2, R's `integrate` function was used to calculate the expected outcomes, whose expressions contain intractable, double integrals. The absolute and relative tolerances were again set to 10^{-12}, with the resultant estimates assumed, for convenience, to be essentially the true values. While the `integrate` function was specifically designed for single integrals, the evaluation of double integrals is readily achievable through the use of R's `sapply` function in conjunction with two function calls to `integrate`. In doing so, this process transforms the integrand, mapping the region $(-\infty, \infty) \times (-\infty, \infty)$ onto $(-1, 1) \times (-1, 1)$. Figure 2.2a displays the plot of the expected outcomes (mortality probabilities) associated with the range of SAPS II scores contained in the dataset used to fit the model. Through interpolation of the integer-only SAPS II scores, the plot is smoother than it would otherwise appear.

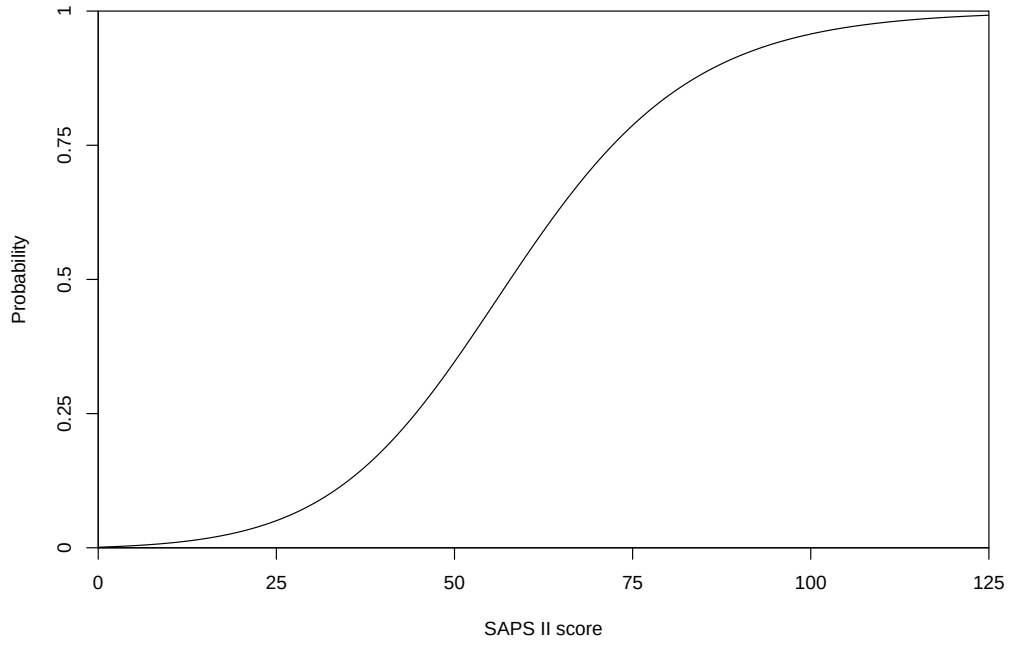
The simple approximation to E_{ij} considers the fixed-effect components only, that is, the E_{ij} were approximated by

$$p(y_{ij} = 1 \,|\, \boldsymbol{x_{ij}}, \boldsymbol{\varepsilon_i} = \mathbf{0}) = g(\boldsymbol{x_{ij}}) = \left(1 + e^{-\mu - \boldsymbol{\alpha}^T \boldsymbol{x_{ij}}}\right)^{-1}.$$

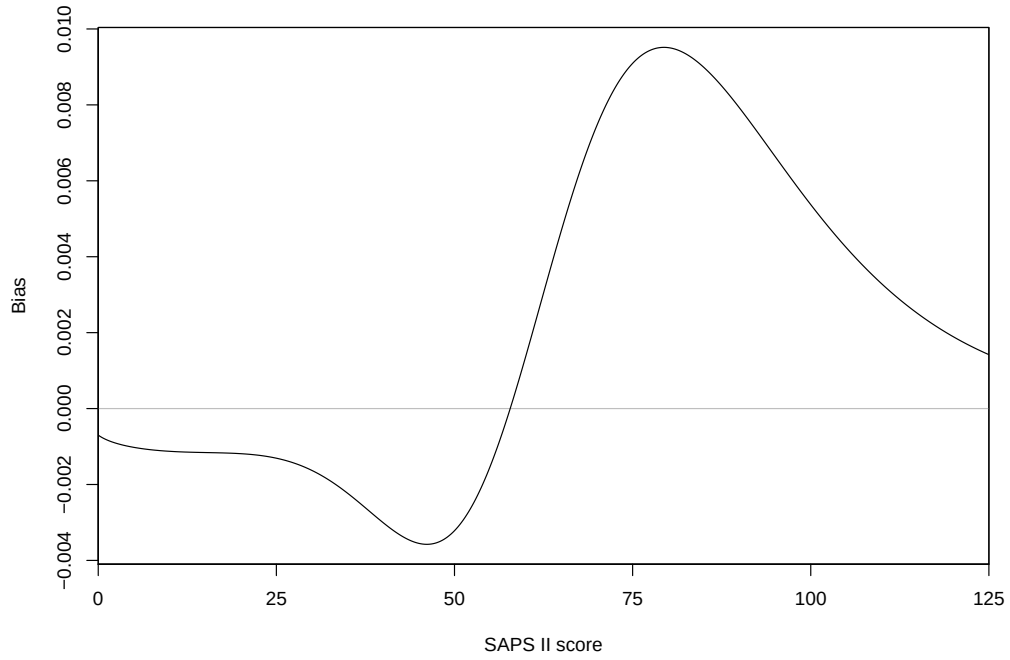
Figure 2.2b displays the plot of the bias function,

$$h(\boldsymbol{x_{ij}}) = g(\boldsymbol{x_{ij}}) - E_{ij},$$

corresponding to the use of the simple technique (considering fixed-effect components only) to approximate the expected outcomes. As before, interpolation of the SAPS II scores facilitated the construction of a smoother plot. Patients with the minimum score of 0 have the expected outcome of 0.0010, while the bias of the simple technique would be -0.0007. From the dataset, patients with the median score of 30 have the



(a) Expected outcome (mortality probability)



(b) Bias of the simple approximation to the expected outcome

Figure 2.2: Evaluation of, and approximation to, the expected outcomes from the fitted, hierarchical model in Glance et al. (2003) across the entire range of SAPS II scores in their data.

expected outcome of 0.0808, while the bias of the simple technique would be −0.0016. Furthermore, patients with the maximum score of 125 have the expected outcome of 0.9923, while the bias of the simple technique would be 0.0014.

Closer scrutiny of the bias function revealed that it would be zero in Figure 2.2b if the SAPS II score could be around 57.82, which corresponds to the expected outcome of one half in Figure 2.2a. Since SAPS II scores are integers, the magnitude of the bias is smallest for scores of 58; this corresponds to the bias of 0.0001 and expected outcome of 0.5036. Inspection of Figure 2.2b reveals that the bias function has two stationary points where the bias has largest magnitude. The stationary points are at $(46.12, -0.0036)$ and $(79.39, 0.0095)$, which correspond to the expected outcomes of 0.2765 and 0.8361, respectively, in Figure 2.2a. Based on integer-only SAPS II scores, the score of 46 is associated with the maximum negative bias of −0.0036 and expected outcome of 0.2745; the score of 79 is associated with the maximum positive bias of 0.0095 and expected outcome of 0.8322.

The simple technique yields consistently biased results in this application, though each bias is fairly slight, ranging between 0.0001 and 0.0095 in absolute value for the SAPS II scores observed in the dataset. However, the median score of 30 lies well below 58, where the bias shifts from negative to positive. Therefore, a considerable majority of the patients' mortality probabilities would have been underestimated in Glance et al. (2003), albeit only slightly for each patient. Perhaps some hospitals' casemixes comprised very few, if any, patients with SAPS II scores of at least 58? If that were the case, then those hospitals would have been attributed (incorrectly) with inflated SMRs, and have an increased risk of being misclassified as under-performers. For other hospitals, the use of the simple approximation technique would not have been problematic if the negative and positive biases roughly cancelled out. Consider a hypothetical example of seven patients, one having the SAPS II score of 1, five having the median score of 30, and the other one having the score of 74. With the SAPS II scores of 1, 30, and 74, the expected outcomes are 0.0015, 0.0808, and 0.7745, respectively, while the simple approximations have biases of −0.0008, −0.0016, and 0.0089, respectively. Approximating the total expected outcome associated with those seven patients yields a bias of only 0.0001 in magnitude (essentially unbiased).

2.3.2.2 Limitations

Applications in critical care studies that extend the model beyond a single random component (the intercept) are quite rare. In many cases, limiting the number of random components may be due to practical constraints. Even if the model-fitting process successfully converges on a solution, it can be very time-consuming. This problem is exacerbated by the large datasets which are often involved, and may only be somewhat mitigated by the use of up-to-date computer hardware. The number of random components is indicative of the dimension of the variance-covariance matrix for that model (beginning with a scalar for the random intercept model). Thus, when estimating the parameters of a model, increasing the number of random components increases the dimension of the integration space involved.

The inclusion of multiple random components also has practical implications. To demonstrate, consider the above-mentioned application from Glance et al. (2003), whose random coefficient model permitted the effect of patient illness severities on mortality to differ between hospitals. It would seem likely that some hospitals performed above average for healthier patients, but below average for sicker patients. Accommodating such scenarios through the inclusion of a random illness-severity

slope may lead to a more realistic model. However, a random slope component absorbs some of the variability in hospital performance that would otherwise be attributed to the random intercept. As Kasza et al. (2013) noted, the inclusion of random coefficients for risks which are beyond a hospital's control may provoke concerns that hospital differences have been absorbed through excessive risk adjustment.

Chapter 3

Laplace approximation

The previous chapter presented various methods of predicting mortality probabilities from the fitted models. Each probability was represented by an expression that involved at least one intractable integral. Furthermore, each (intractable) integral needs to be computed over the entire real number line, while each integrand represents a smooth, unimodal function, with an entirely positive range. These integrals must be evaluated numerically since they cannot be solved analytically. Despite the potential accuracy with quadrature and simulation, these numerical techniques can be too computationally intensive for applications of standardised mortality ratios (SMRs). To illustrate, the application in the next chapter must evaluate in the order of billions of intractable integrals, and the integrals associated with "smoothed" outcomes (Sections 2.2.4 and 2.2.5) involve integrands that are relatively time-consuming to compute at any point. This motivated the pursuit of a more computationally efficient technique which achieves consistently higher accuracy than that of the simple approximation technique discussed in Sections 2.2.3.1 and 2.3.2.1.

The Laplace approximation is a much faster alternative in terms of the average processing times (Raudenbush et al., 2000). Hence, the Laplace approximation could be ideally suited for many applications, including the present endeavour. In the statistical literature, the standard version of the Laplace approximation is frequently referred to as Laplace's method or Laplace's approximation; examples of where these terms are used interchangeably include Davison (2003), Wakefield (2013), and Azevedo-Filho and Shachter (1994). Pierre-Simon Laplace originally published his eponymous approximation technique in 1774 (Laplace, 1986). Laplace's method often achieves remarkable accuracy (Davison, 2003, p. 597), and has been successfully applied in many disciplines (Azevedo-Filho and Shachter, 1994). It is known that Laplace's method yields better approximations when integrands more closely resemble probability density functions (PDFs) of normal distributions; examples of such functions tend to be those associated with posterior distributions because (by the Central Limit Theorem) they are generally asymptotically normal (Hartigan, 1983). However, Laplace's method has been extended in applications which require approximations of even greater accuracy.

This chapter explores the Laplace approximation, beginning with the univariate version, and followed by the more general, multivariate version. The discussion of both the univariate and multivariate versions include derivations of Laplace's method, then extensions to it. The technique for deriving the multivariate version presented here generalises the work of Raudenbush et al. (2000). This chapter concludes by illustrating the accuracy and speed which are potentially achievable from implementing Laplace-related approximations for single and double integrals.

3.1 Univariate version

Consider the positive function $f(x)$, expressible as $e^{nh(x)}$, where h is a unimodal function that is at least twice differentiable, and n is a positive constant. Following Wakefield (2013, p. 106), the exponent includes n solely to highlight the asymptotic arguments that it facilitates. This will also be shown to provide one selection criterion for determining the extent to which an infinite series should be expanded. However, in practice, it can be preferable to use the equivalent and simpler technique of setting the n in f to one by making h absorb the entire contents of the exponent (Davison, 2003, p. 598). The accuracy of subsequent approximations would then depend entirely on the shape of h.

3.1.1 Laplace's method

The single integral

$$I \;=\; \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \;=\; \int_{-\infty}^{\infty} e^{nh(x)}\,\mathrm{d}x$$

is approximated by Laplace's method as follows. Firstly, the exponent has the second-order Taylor series approximation about a general point, m, of

$$h(x) \;\approx\; h(m) + (x-m)\,h'(m) + \frac{1}{2}(x-m)^2\,h''(m)\,,$$

with equality if h is a quadratic function. Hence, Laplace's method becomes increasingly accurate as h more closely resembles a quadratic function, and thus f becomes a more normal-like ("bell-shaped") function. Next, denote $x = m$ as the maximum of h (and therefore f too), so $h'(m) = 0$, $h''(m) < 0$, and

$$h(x) \;\approx\; h(m) + \frac{1}{2}(x-m)^2\,h''(m)\,.$$

It follows that

$$\begin{aligned} I \;&\approx\; \int_{-\infty}^{\infty} \exp\left\{ n\left[h(m) + \frac{1}{2}(x-m)^2\,h''(m) \right] \right\} \mathrm{d}x \\ &=\; \sqrt{\frac{2\pi}{-nh''(m)}}\; e^{nh(m)} \int_{-\infty}^{\infty} \sqrt{\frac{-nh''(m)}{2\pi}}\; \exp\left\{ \frac{n}{2}(x-m)^2\,h''(m) \right\} \mathrm{d}x \\ &=\; \sqrt{\frac{2\pi}{-nh''(m)}}\; e^{nh(m)} \end{aligned}$$

because the integrand is the PDF of $N\left(m,\, [-nh''(m)]^{-1}\right)$.

Suppose instead that the integrand (function f) did not satisfy the previously specified conditions. If f is zero in part of its domain, then expressing the integrand in exponential form requires restrictions to be placed on the limits of integration; taking the case where $f(x) = 0$ for $x \leq 0$, the limits of integration would be $(0, \infty)$. The Laplace approximation will be affected negligibly if the restricted limits of integration contain almost the entire density of the normal probability integral (Davison, 2003, p. 597). Now suppose that function f has multiple maxima. With f being multimodal, use of the Laplace approximation is less likely to yield results of acceptable

accuracy. An individual application of the Laplace approximation would require m to be set to one particular maximum. While the global maximum may seem to be a reasonable choice, a considerable proportion of the density may still surround local maxima. Alternatively, Davison (2003, p. 602) suggested using separate applications of the Laplace approximation for each mode, provided that each one is locatable.

3.1.2 Extending Laplace's method

Laplace's method can be extended by considering higher-order terms in the Taylor series representation of h. Denoting $h_k = h^{(k)}(m)$ for $k \in \mathbb{Z}_{\geq 0}$, the Taylor series of h about its mode is

$$\begin{aligned} h(x) &= h_0 + (x-m)\,h_1 + \frac{1}{2}(x-m)^2 h_2 + \sum_{k=3}^{\infty} \frac{1}{k!}(x-m)^k h_k \\ &= h_0 + \frac{1}{2}(x-m)^2 h_2 + \sum_{k=3}^{\infty} \frac{1}{k!}(x-m)^k h_k, \end{aligned}$$

so

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \exp\left\{ n\left[h_0 + \frac{1}{2}(x-m)^2 h_2 + \sum_{k=3}^{\infty} \frac{1}{k!}(x-m)^k h_k \right] \right\} \mathrm{d}x \\ &= \sqrt{\frac{2\pi}{-nh_2}}\; e^{nh_0} \int_{-\infty}^{\infty} \sqrt{\frac{-nh_2}{2\pi}}\; \exp\left\{ \frac{n}{2}(x-m)^2 h_2 \right\} \exp\left\{ \sum_{k=3}^{\infty} \frac{n}{k!}(x-m)^k h_k \right\} \mathrm{d}x. \end{aligned}$$

Introducing further notational simplification using the substitutions $v = (-h_2)^{-1}$, $w = \sqrt{n}\,(x-m)$ and $\mathrm{d}x = \frac{1}{\sqrt{n}}\mathrm{d}w$,

$$\begin{aligned} I &= \sqrt{2\pi v}\; e^{nh_0} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi v}}\; \exp\left\{ -\frac{w^2}{2v} \right\} \exp\left\{ \sum_{k=3}^{\infty} \frac{1}{k!} n^{1-k/2} h_k w^k \right\} \mathrm{d}w \\ &= \sqrt{2\pi v}\; e^{nh_0}\; E\left[\exp\left\{ \sum_{k=3}^{\infty} \frac{1}{k!} n^{1-k/2} h_k W^k \right\} \right], \end{aligned}$$

where the expectation integrates over the (univariate) normal distribution of W, with mean zero and variance v. By denoting

$$U_k = \frac{1}{k!} h_k W^k,$$

then

$$\begin{aligned} I &= \sqrt{2\pi v}\; e^{nh_0}\; E\left[\exp\left\{ \sum_{k=3}^{\infty} n^{1-k/2} U_k \right\} \right] \\ &= \sqrt{2\pi v}\; e^{nh_0}\; E\left[\sum_{i=0}^{\infty} \frac{1}{i!} \left\{ \sum_{k=3}^{\infty} n^{1-k/2} U_k \right\}^i \right] \end{aligned} \tag{3.1}$$

because

$$e^a = \sum_{i=0}^{\infty} \frac{a^i}{i!}.$$

The infinite series in Eq. 3.1 will now be denoted

$$S = \sum_{i=0}^{\infty} \frac{1}{i!} \left\{ \sum_{k=3}^{\infty} n^{1-k/2} U_k \right\}^i,$$

which is a polynomial in increasing powers of $n^{-1/2}$. This chapter expresses expansions of the summation for S using "big oh" notation as in Davison (2003, p. 35). That is, for $j \in \mathbb{Z}$, $O\left(n^{-j/2}\right)$ contains all terms involving the powers of n up to, and including, that of $-j/2$.[1] It follows that the expansion of S to $O\left(\frac{1}{n^3}\right)$ yields

$$S = 1 + \frac{C_{1/2}}{n^{1/2}} + \frac{C_1}{n} + \frac{C_{3/2}}{n^{3/2}} + \frac{C_2}{n^2} + \frac{C_{5/2}}{n^{5/2}} + O\left(\frac{1}{n^3}\right),$$

where the coefficients can be simplified to

$$\begin{aligned} C_{1/2} &= U_3, \\ C_1 &= \frac{1}{2}U_3^2 + U_4, \\ C_{3/2} &= \frac{1}{6}U_3^3 + U_3U_4 + U_5, \\ C_2 &= \frac{1}{24}U_3^4 + \frac{1}{2}U_3^2U_4 + U_3U_5 + \frac{1}{2}U_4^2 + U_6, \end{aligned}$$

and

$$C_{5/2} = \frac{1}{120}U_3^5 + \frac{1}{6}U_3^3U_4 + \frac{1}{2}U_3U_4^2 + \frac{1}{2}U_3^2U_5 + U_4U_5 + U_3U_6 + U_7.$$

Now the general expression

$$\prod_{i=1}^{l} U_{k_i} = \left[\prod_{i=1}^{l} \frac{h_{k_i}}{k_i!} \right] W^K,$$

where

$$K = \sum_{i=1}^{l} k_i$$

and the k_i need not be distinct. As a result, the coefficients of the polynomial in $n^{-1/2}$ become

$$\begin{aligned} C_{1/2} &= \frac{h_3W^3}{6}, \\ C_1 &= \frac{h_3^2W^6}{72} + \frac{h_4W^4}{24}, \\ C_{3/2} &= \frac{h_3^3W^9}{1296} + \frac{h_3h_4W^7}{144} + \frac{h_5W^5}{120}, \\ C_2 &= \frac{h_3^4W^{12}}{31104} + \frac{h_3^2h_4W^{10}}{1728} + \frac{h_3h_5W^8}{720} + \frac{h_4^2W^8}{1152} + \frac{h_6W^6}{720}, \end{aligned}$$

and

[1] Suppose that $T = \sum_{l=j}^{\infty} a_l n^{-l/2}$, where the a_l are any constants and $j \in \mathbb{Z}$. The largest power of n is when $-l/2$ is maximised, and hence l is minimised. This occurs when $l = j$, so T is $O\left(n^{-j/2}\right)$.

$$C_{5/2} = \frac{h_3^5 W^{15}}{933120} + \frac{h_3^3 h_4 W^{13}}{31104} + \frac{h_3 h_4^2 W^{11}}{6912} + \frac{h_3^2 h_5 W^{11}}{8640} + \frac{h_4 h_5 W^9}{2880} + \frac{h_3 h_6 W^9}{4320} + \frac{h_7 W^7}{5040}.$$

Using Eq. A.1,

$$E\left[W^k\right] = \begin{cases} 0 & \text{if } k \text{ is odd} \\ \frac{k!\, v^{k/2}}{2^{k/2}\left(\frac{k}{2}\right)!} & \text{if } k \text{ is even} \end{cases},$$

so algebraic simplification leads to $E\left[C_{1/2}\right] = E\left[C_{3/2}\right] = E\left[C_{5/2}\right] = 0$,

$$E\left[C_1\right] = \frac{5}{24} h_3^2 v^3 + \frac{1}{8} h_4 v^2,$$

and

$$E\left[C_2\right] = \frac{385}{1152} h_3^4 v^6 + \frac{35}{64} h_3^2 h_4 v^5 + \frac{7}{48} h_3 h_5 v^4 + \frac{35}{384} h_4^2 v^4 + \frac{1}{48} h_6 v^3.$$

Hence

$$\begin{aligned} I &= \sqrt{\frac{2\pi v}{n}}\ e^{n h_0}\ E\left[S\right] \\ &= \sqrt{\frac{2\pi v}{n}}\ e^{n h_0} \left\{1 + \frac{a_1}{n} + \frac{a_2}{n^2} + O\left(\frac{1}{n^3}\right)\right\}, \end{aligned}$$

where $a_1 = E\left[C_1\right]$ and $a_2 = E\left[C_2\right]$. The Laplace approximation essentially omits the $O\left(\frac{1}{n}\right)$ terms from $E\left[S\right]$, which could be adequate when n is sufficiently large. In such cases, the region surrounding the mode of f has the greatest contribution towards the approximation of I (Azevedo-Filho and Shachter, 1994). However, higher-order terms in $E\left[S\right]$ may contribute markedly towards the approximation in certain circumstances, including when n has a relatively small or moderate value.

This chapter introduces new terminology to facilitate more concise referencing of the extensions to the Laplace approximation. Firstly, define the Laplace-I approximation as Laplace's method. The Laplace-II approximation is then defined as the expression derived by relaxing the omissions in $E\left[S\right]$ to $O\left(\frac{1}{n^2}\right)$, so

$$I \approx \sqrt{\frac{2\pi v}{n}}\ e^{n h_0} \left\{1 + \frac{a_1}{n}\right\}.$$

Subsequently, the Laplace-III approximation is defined as the expression derived by further relaxing the omissions in $E\left[S\right]$ to $O\left(\frac{1}{n^3}\right)$, so

$$I \approx \sqrt{\frac{2\pi v}{n}}\ e^{n h_0} \left\{1 + \frac{a_1}{n} + \frac{a_2}{n^2}\right\}.$$

It is noted that this is the extension to the Laplace approximation presented by Tierney and Kadane (1986, Appendix).

3.1.2.1 Further extensions

More accurate approximations to I may be achieved via further expansion of S. To $O\left(\frac{1}{n^4}\right)$, the series is

$$S = 1 + \frac{C_{1/2}}{n^{1/2}} + \frac{C_1}{n} + \frac{C_{3/2}}{n^{3/2}} + \frac{C_2}{n^2} + \frac{C_{5/2}}{n^{5/2}} + \frac{C_3}{n^3} + \frac{C_{7/2}}{n^{7/2}} + O\left(\frac{1}{n^4}\right),$$

where C_3 and $C_{7/2}$ are the only coefficients whose expressions and expectations have yet to be derived in this chapter. Following algebraic simplification, one can obtain the expressions

$$\begin{aligned} C_3 &= \frac{1}{720}U_3^6 + \frac{1}{24}U_3^4U_4 + \frac{1}{4}U_3^2U_4^2 + \frac{1}{6}U_3^3U_5 + U_3U_4U_5 + \frac{1}{2}U_3^2U_6 + \frac{1}{6}U_4^3 \\ &\quad + U_4U_6 + \frac{1}{2}U_5^2 + U_3U_7 + U_8 \\ &= \frac{h_3^6W^{18}}{33592320} + \frac{h_3^4h_4W^{16}}{746496} + \frac{h_3^2h_4^2W^{14}}{82944} + \frac{h_3^3h_5W^{14}}{155520} + \frac{h_3h_4h_5W^{12}}{17280} \\ &\quad + \frac{h_3^2h_6W^{12}}{51840} + \frac{h_4^3W^{12}}{82944} + \frac{h_4h_6W^{10}}{17280} + \frac{h_5^2W^{10}}{28800} + \frac{h_3h_7W^{10}}{30240} + \frac{h_8W^8}{40320} \end{aligned}$$

and

$$\begin{aligned} C_{7/2} &= \frac{1}{5040}U_3^7 + \frac{1}{120}U_3^5U_4 + \frac{1}{12}U_3^3U_4^2 + \frac{1}{24}U_3^4U_5 + \frac{1}{2}U_3^2U_4U_5 + \frac{1}{6}U_3U_4^3 + \frac{1}{6}U_3^3U_6 \\ &\quad + U_3U_4U_6 + \frac{1}{2}U_4^2U_5 + \frac{1}{2}U_3U_5^2 + \frac{1}{2}U_3^2U_7 + U_5U_6 + U_4U_7 + U_3U_8 + U_9 \\ &= \frac{h_3^7W^{21}}{1410877440} + \frac{h_3^5h_4W^{19}}{22394880} + \frac{h_3^3h_4^2W^{17}}{1492992} + \frac{h_3^4h_5W^{17}}{3732480} + \frac{h_3^2h_4h_5W^{15}}{207360} \\ &\quad + \frac{h_3h_4^3W^{15}}{497664} + \frac{h_3^3h_6W^{15}}{933120} + \frac{h_3h_4h_6W^{13}}{103680} + \frac{h_4^2h_5W^{13}}{138240} + \frac{h_3h_5^2W^{13}}{172800} \\ &\quad + \frac{h_3^2h_7W^{13}}{362880} + \frac{h_5h_6W^{11}}{86400} + \frac{h_4h_7W^{11}}{120960} + \frac{h_3h_8W^{11}}{241920} + \frac{h_9W^9}{362880}, \end{aligned}$$

which have the expected values

$$\begin{aligned} E\left[C_3\right] &= \frac{85085}{82944}h_3^6v^9 + \frac{25025}{9216}h_3^4h_4v^8 + \frac{5005}{3072}h_3^2h_4^2v^7 + \frac{1001}{1152}h_3^3h_5v^7 + \frac{77}{128}h_3h_4h_5v^6 \\ &\quad + \frac{77}{384}h_3^2h_6v^6 + \frac{385}{3072}h_4^3v^6 + \frac{7}{128}h_4h_6v^5 + \frac{21}{640}h_5^2v^5 + \frac{1}{32}h_3h_7v^5 + \frac{1}{384}h_8v^4 \end{aligned}$$

and $E\left[C_{7/2}\right] = 0$. Hence

$$I = \sqrt{\frac{2\pi v}{n}}\ e^{nh_0}\left\{1 + \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + O\left(\frac{1}{n^4}\right)\right\},$$

where $a_3 = E\left[C_3\right]$. Continuing with the introduction of new terminology by defining the Laplace-IV approximation as the expression derived by relaxing the omissions in $E\left[S\right]$ even further, to the $O\left(\frac{1}{n^4}\right)$ terms, so

$$I \approx \sqrt{\frac{2\pi v}{n}}\ e^{nh_0}\left\{1 + \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3}\right\}.$$

Section 3.3.2 will present an illustrative example of approximating intractable, single integrals using the techniques described so far. However, this example will also use additional Laplace-related approximation techniques that have yet to be derived, but are obtained after following and extending the multivariate work by Raudenbush et al. (2000), as will be discussed in Section 3.2.2.

3.1.2.2 Limitations

The number of terms which are contained in extensions to the Laplace approximation grows dramatically with higher expansions of S, though the effect of this growth is somewhat eased because $E[C_j] = 0$ when $j \notin \mathbb{N}$. It has been shown that the Laplace-II approximation contains two extra terms over Laplace-I; expanding from the Laplace-II approximation to Laplace-III introduces five extra terms; and expanding from the Laplace-III approximation to Laplace-IV introduces eleven extra terms. Following further algebraic expansion of S, and taking expectations, it can be shown[2] that $E[C_4]$, $E[C_5]$, and $E[C_6]$ introduce, respectively, an extra 22, 42, and 77 terms to the approximation. The terminology introduced in this chapter labels the resultant approximations to I as being the Laplace-V, Laplace-VI, and Laplace-VII approximations, respectively.

3.2 Multivariate version

Section 2.3.2 discussed a general model containing multiple random effects, where the expressions representing mortality probabilities involve intractable, multivariable integrals. This motivates generalising the univariate work presented in the previous section.

Suppose $f(\boldsymbol{x})$ is a positive function, expressible as $e^{nh(\boldsymbol{x})}$, where $\boldsymbol{x} = (x_1, \ldots, x_q)^T$, h is a unimodal function whose first- and second-order partial derivatives all exist, and n is a positive constant. As before, n is included for subsequent consideration of the asymptotic behaviour. The previously discussed effects on the Laplace approximation from the function f instead being either multimodal or zero in part of its domain also apply here. However, these matters will not be explored further because they are unrelated to any integrand considered in this thesis.

3.2.1 Laplace's method

The multiple integral

$$I \;=\; \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(\boldsymbol{x})\, \mathrm{d}x_1 \ldots \mathrm{d}x_q \;=\; \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} e^{nh(\boldsymbol{x})}\, \mathrm{d}x_1 \ldots \mathrm{d}x_q$$

is approximated by Laplace's method as follows. Firstly, the exponent has the second-order Taylor series approximation about $\boldsymbol{m}$ of

$$h(\boldsymbol{x}) \;\approx\; h(\boldsymbol{m}) + h^{(1)}(\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T h^{(2)}(\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m}),$$

where the q-element row vector

$$h^{(1)}(\boldsymbol{m}) \;=\; \left.\frac{\partial h}{\partial \boldsymbol{x}^{\boldsymbol{T}}}\right|_{\boldsymbol{x}=\boldsymbol{m}}$$

and the $q \times q$ matrix

$$h^{(2)}(\boldsymbol{m}) \;=\; \left.\frac{\partial^2 h}{\partial \boldsymbol{x} \partial \boldsymbol{x}^{\boldsymbol{T}}}\right|_{\boldsymbol{x}=\boldsymbol{m}}.$$

[2] e.g. via Mathematica (Wolfram Research Inc., 2017)

As before, equality occurs when h is a quadratic function, indicating that Laplace's method is more ideally suited to integrating normal-like functions. Now by setting $\boldsymbol{x} = \boldsymbol{m}$ as the maximum of h (and therefore of f too), $h^{(1)}(\boldsymbol{m}) = \mathbf{0}^{\boldsymbol{T}}$, $\left|h^{(2)}(\boldsymbol{m})\right| > 0$, and each $\frac{\partial^2 h}{\partial x_k^2}(\boldsymbol{m}) < 0$ for $k = 1, \ldots, q$. Hence

$$h(\boldsymbol{x}) \;\approx\; h(\boldsymbol{m}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T h^{(2)}(\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m}),$$

so

$$\begin{aligned} I \;&\approx\; \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{ n\left[h(\boldsymbol{m}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T h^{(2)}(\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m}) \right] \right\} \mathrm{d}x_1 \ldots \mathrm{d}x_q \\ &=\; \frac{(2\pi)^{q/2}}{|-nh^{(2)}(\boldsymbol{m})|^{1/2}}\; e^{nh(\boldsymbol{m})} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{q/2}\left|\left[-nh^{(2)}(\boldsymbol{m})\right]^{-1}\right|^{1/2}} \\ &\qquad\qquad \times \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T \left[-nh^{(2)}(\boldsymbol{m})\right](\boldsymbol{x}-\boldsymbol{m}) \right\} \mathrm{d}x_1 \ldots \mathrm{d}x_q \\ &=\; \frac{(2\pi/n)^{q/2}}{|-h^{(2)}(\boldsymbol{m})|^{1/2}}\; e^{nh(\boldsymbol{m})} \end{aligned}$$

because the integrand is the PDF of $N_q\Big(\boldsymbol{m},\, \left[-nh^{(2)}(\boldsymbol{m})\right]^{-1}\Big)$.

3.2.2 Extending Laplace's method

Ruli et al. (2016) argue that extensions to Laplace's method can be even more critical in applications involving multidimensional integrals, because the shape of the integrand is more prone to differing vastly from a normal PDF. The standard Laplace approximation can again be extended by considering higher-order terms in the Taylor series representation of h. In doing so, this section follows Raudenbush et al. (2000), who use the Magnus and Neudecker approach to matrix differential calculus. From Raudenbush et al. (2000, Def. 1, p. 145),

$$\begin{aligned} h(\boldsymbol{x}) \;&=\; h(\boldsymbol{m}) + h^{(1)}(\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m}) + \sum_{k=2}^{\infty} \frac{1}{k!} \left[\overset{k-1}{\otimes} (\boldsymbol{x}-\boldsymbol{m})^T \right] h^{(k)}(\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m}) \\ &=\; h(\boldsymbol{m}) + \sum_{k=2}^{\infty} \frac{1}{k!} \left[\overset{k-1}{\otimes} (\boldsymbol{x}-\boldsymbol{m})^T \right] h^{(k)}(\boldsymbol{m})(\boldsymbol{x}-\boldsymbol{m}), \end{aligned}$$

where

$$h^{(k)}(\boldsymbol{m}) \;=\; \left.\frac{\partial\left[\mathrm{vec}\left\{h^{(k-1)}(\boldsymbol{x})\right\}\right]}{\partial \boldsymbol{x}^{\boldsymbol{T}}}\right|_{\boldsymbol{x}=\boldsymbol{m}}$$

and

$$\overset{k}{\otimes}\, \boldsymbol{u} \;=\; \boldsymbol{u} \otimes \boldsymbol{u} \otimes \cdots \otimes \boldsymbol{u}$$

involves $k-1$ Kronecker products.[3] It is important to note that earlier definitions of $h^{(1)}(\boldsymbol{m})$ and $h^{(2)}(\boldsymbol{m})$ conform with the above, being $h^{(k)}(\boldsymbol{m})$ for $k = 1$ and 2.

[3] Appendix A.3 provides general definitions of the "vec" operator and Kronecker product.

After denoting $H_k = h^{(k)}(\boldsymbol{m})$ for $k \in \mathbb{Z}_{\geq 0}$, further expansion of h yields

$$h(\boldsymbol{x}) \;=\; H_0 + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T H_2 (\boldsymbol{x}-\boldsymbol{m}) + \sum_{k=3}^{\infty} \frac{1}{k!} \left[\overset{k-1}{\otimes} (\boldsymbol{x}-\boldsymbol{m})^T \right] H_k (\boldsymbol{x}-\boldsymbol{m}),$$

so

$$\begin{aligned} I \;=\; e^{nH_0} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \Bigg\{ & \frac{n}{2}(\boldsymbol{x}-\boldsymbol{m})^T H_2 (\boldsymbol{x}-\boldsymbol{m}) \\ & + \sum_{k=3}^{\infty} \frac{n}{k!} \left[\overset{k-1}{\otimes} (\boldsymbol{x}-\boldsymbol{m})^T \right] H_k (\boldsymbol{x}-\boldsymbol{m}) \Bigg\} \, \mathrm{d}x_1 \ldots \mathrm{d}x_q. \end{aligned}$$

Following Raudenbush et al. (2000) by denoting

$$t_k \;=\; \frac{1}{k!} \left[\overset{k-1}{\otimes} (\boldsymbol{x}-\boldsymbol{m})^T \right] H_k (\boldsymbol{x}-\boldsymbol{m})$$

and $V = [-H_2]^{-1}$, the multiple integral becomes

$$\begin{aligned} I \;=\; e^{nH_0} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} & \exp\left\{ -\frac{n}{2}(\boldsymbol{x}-\boldsymbol{m})^T V^{-1} (\boldsymbol{x}-\boldsymbol{m}) \right\} \\ & \times \exp\left\{ \sum_{k=3}^{\infty} n t_k \right\} \mathrm{d}x_1 \ldots \mathrm{d}x_q. \end{aligned}$$

By also letting $\boldsymbol{w} = \sqrt{n}(\boldsymbol{x}-\boldsymbol{m})$, then

$$\mathrm{d}x_i \;=\; \frac{1}{\sqrt{n}} \mathrm{d}w_i, \qquad \text{for } i = 1, \ldots, q,$$

and

$$\begin{aligned} t_k \;&=\; \frac{1}{k!} \left[\overset{k-1}{\otimes} \left(\frac{1}{\sqrt{n}} \boldsymbol{w} \right)^T \right] H_k \left(\frac{1}{\sqrt{n}} \boldsymbol{w} \right) \\ &=\; n^{-k/2} \times \frac{1}{k!} \left[\overset{k-1}{\otimes} \boldsymbol{w}^T \right] H_k \boldsymbol{w} \\ &=\; n^{-k/2} \, u_k. \end{aligned}$$

Therefore,

$$\begin{aligned} I \;&=\; n^{-q/2} \, e^{nH_0} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2} \boldsymbol{w}^T V^{-1} \boldsymbol{w} \right\} \exp\left\{ \sum_{k=3}^{\infty} n^{1-k/2} \, u_k \right\} \mathrm{d}w_1 \ldots \mathrm{d}w_q \\ &=\; \left(\frac{2\pi}{n} \right)^{q/2} |V|^{1/2} \, e^{nH_0} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ (2\pi)^{q/2} |V|^{1/2} \right\}^{-1} \\ & \qquad \times \exp\left\{ -\frac{1}{2} \boldsymbol{w}^T V^{-1} \boldsymbol{w} \right\} \exp\left\{ \sum_{k=3}^{\infty} n^{1-k/2} \, u_k \right\} \mathrm{d}w_1 \ldots \mathrm{d}w_q \\ &=\; \left(\frac{2\pi}{n} \right)^{q/2} |V|^{1/2} \, e^{nH_0} \, E\left[\exp\left\{ \sum_{k=3}^{\infty} n^{1-k/2} \, U_k \right\} \right], \end{aligned} \tag{3.2}$$

where

$$U_k \;=\; \frac{1}{k!}\left[\overset{k-1}{\otimes}\; \boldsymbol{W}^T\right] H_k \boldsymbol{W}, \tag{3.3}$$

and the expectation integrates over the q-dimensional multivariate normal distribution of $\boldsymbol{W}$, with mean vector $\mathbf{0}$ and covariance matrix V.

Similar to the process followed in Section 3.1.2, the expression inside the expectation of Eq. 3.2 is denoted

$$S \;=\; \exp\left\{\sum_{k=3}^{\infty} n^{1-k/2}\, U_k\right\} \;=\; \sum_{i=0}^{\infty} \frac{1}{i!}\left\{\sum_{k=3}^{\infty} n^{1-k/2}\, U_k\right\}^i,$$

and when expanded to $O\left(\frac{1}{n^3}\right)$, yields

$$S \;=\; 1 + \frac{C_{1/2}}{n^{1/2}} + \frac{C_1}{n} + \frac{C_{3/2}}{n^{3/2}} + \frac{C_2}{n^2} + \frac{C_{5/2}}{n^{5/2}} + O\left(\frac{1}{n^3}\right),$$

where

$$\begin{aligned}
C_{1/2} &= U_3, \\
C_1 &= \frac{1}{2}U_3^2 + U_4, \\
C_{3/2} &= \frac{1}{6}U_3^3 + U_3U_4 + U_5, \\
C_2 &= \frac{1}{24}U_3^4 + \frac{1}{2}U_3^2U_4 + U_3U_5 + \frac{1}{2}U_4^2 + U_6,
\end{aligned}$$

and

$$C_{5/2} \;=\; \frac{1}{120}U_3^5 + \frac{1}{6}U_3^3U_4 + \frac{1}{2}U_3U_4^2 + \frac{1}{2}U_3^2U_5 + U_4U_5 + U_3U_6 + U_7.$$

The procedure for approximating S can follow (as before) by omitting the $O\left(n^{-j/2}\right)$ terms, and so all terms involving powers of $n^{-1/2}$ for even $\mathbb{Z}_{\geq j}$. As will be discussed later in this section, all terms involving odd-numbered powers of $n^{-1/2}$ are essentially of no consequence because they disappear (equal zero) once expectations are taken.

Raudenbush et al. (2000) opted for a simpler approach to approximating S. Essentially, the expansion of S was firstly limited to the second-order polynomial. Thus, S is initially approximated by

$$\widehat{S} \;=\; \sum_{i=0}^{2} \frac{1}{i!}\left\{\sum_{k=3}^{\infty} n^{1-k/2}\, U_k\right\}^i.$$

Then the expression for $\widehat{S}$ was limited to the sixth order based on the elements of $\boldsymbol{x}$, and $\widehat{S}$ was approximated by

$$\widehat{\widehat{S}} \;=\; 1 + \frac{1}{n^{1/2}}U_3 + \frac{1}{n}U_4 + \frac{1}{n^{3/2}}U_5 + \frac{1}{n^2}U_6 + \frac{1}{2n}U_3^2.$$

It is important to note that the above expression only restricts n to being positive, whereas Raudenbush et al. (2000) implicitly set it equal to one. Omitting the additional terms from S may be justifiable since "the magnitude of higher-order terms diminishes as the factorial denominator rapidly increases" (Raudenbush et al.,

2000). While approximating S by $\hat{\hat{\hat{S}}}$ lead to reasonably accurate results in Raudenbush et al. (2000), they recommended investigating greater expansions of the series. Such expansions are considered later in this section.

Returning to the $O\left(\frac{1}{n^3}\right)$ expansion, consideration will now be given to the more general product of l terms,

$$\prod_{i=1}^{l} U_{k_i},$$

where the $k_i \in \mathbb{Z}_{\geq 3}$ need not be distinct. By denoting

$$K = \sum_{i=1}^{l} k_i,$$

Appendix A.3 proves that Raudenbush et al. (2000, Thm. 2, p. 146) can be extended to accommodate all expressions where $l \geq 3$ (i.e., for any product of more than two U_k terms) with

$$E\left[\prod_{i=1}^{l} U_{k_i}\right] = \begin{cases} 0 & \text{if } K \text{ is odd} \\ \frac{K!}{\left(\prod_{i=1}^{l} k_i!\right) 2^{K/2} \left(\frac{K}{2}\right)!} & \\ \times \operatorname{vec}^T \left\{ \overset{K/2}{\otimes} V \right\} \operatorname{vec}\left\{H_{k_1} \otimes \ldots \otimes H_{k_l}\right\} & \text{if } K \text{ is even} \end{cases}. \quad (3.4)$$

Algebraic simplification leads to $E\left[C_{1/2}\right] = E\left[C_{3/2}\right] = E\left[C_{5/2}\right] = 0$,

$$E\left[C_1\right] = \frac{5}{24} \operatorname{vec}^T \left\{ \overset{3}{\otimes} V \right\} \operatorname{vec}\left\{ \overset{2}{\otimes} H_3 \right\} + \frac{1}{8} \operatorname{vec}^T \left\{ \overset{2}{\otimes} V \right\} \operatorname{vec}\left\{H_4\right\},$$

and

$$\begin{aligned} E\left[C_2\right] = & \frac{385}{1152} \operatorname{vec}^T \left\{ \overset{6}{\otimes} V \right\} \operatorname{vec}\left\{ \overset{4}{\otimes} H_3 \right\} \\ & + \frac{35}{64} \operatorname{vec}^T \left\{ \overset{5}{\otimes} V \right\} \operatorname{vec}\left\{ \left[\overset{2}{\otimes} H_3 \right] \otimes H_4 \right\} \\ & + \operatorname{vec}^T \left\{ \overset{4}{\otimes} V \right\} \left(\frac{7}{48} \operatorname{vec}\left\{H_3 \otimes H_5\right\} + \frac{35}{384} \operatorname{vec}\left\{ \overset{2}{\otimes} H_4 \right\} \right) \\ & + \frac{1}{48} \operatorname{vec}^T \left\{ \overset{3}{\otimes} V \right\} \operatorname{vec}\left\{H_6\right\}. \end{aligned}$$

Hence

$$\begin{aligned} I &= \left(\frac{2\pi}{n}\right)^{q/2} |V|^{1/2} \, e^{nH_0} \, E\left[S\right] \\ &= \left(\frac{2\pi}{n}\right)^{q/2} |V|^{1/2} \, e^{nH_0} \left\{ 1 + \frac{a_1}{n} + \frac{a_2}{n^2} + O\left(\frac{1}{n^3}\right) \right\}, \end{aligned}$$

where $a_1 = E\left[C_1\right]$ and $a_2 = E\left[C_2\right]$.

The asymptotic arguments for the multivariate version are identical to those asserted for the univariate version. The terminology introduced for Laplace-related approximations is used here as well. Firstly, the Laplace-I approximation follows from omitting the $O\left(\frac{1}{n}\right)$ terms from $E\left[S\right]$, which could be adequate when n is sufficiently

large. The Laplace-II approximation is derived by relaxing the omissions in $E[S]$ to $O\left(\frac{1}{n^2}\right)$, so

$$I \approx \left(\frac{2\pi}{n}\right)^{q/2} |V|^{1/2}\, e^{nH_0} \left\{1 + \frac{a_1}{n}\right\}.$$

Similarly, the Laplace-III approximation is derived by further relaxing the omissions in $E[S]$ to $O\left(\frac{1}{n^3}\right)$, so

$$I \approx \left(\frac{2\pi}{n}\right)^{q/2} |V|^{1/2}\, e^{nH_0} \left\{1 + \frac{a_1}{n} + \frac{a_2}{n^2}\right\}.$$

Raudenbush et al. (2000) use the Laplace6 approximation instead. By taking expectations after approximating S by $\hat{\hat{\hat{S}}}$, the Laplace-III approximation reduces to Laplace6,

$$\begin{aligned} I &\approx \left(\frac{2\pi}{n}\right)^{q/2} |V|^{1/2}\, e^{nH_0} \\ &\quad \times \left\{1 + \frac{1}{n}\left(\frac{5}{24}\,\mathrm{vec}^T\left\{\overset{3}{\otimes} V\right\}\mathrm{vec}\left\{\overset{2}{\otimes} H_3\right\} + \frac{1}{8}\,\mathrm{vec}^T\left\{\overset{2}{\otimes} V\right\}\mathrm{vec}\{H_4\}\right)\right. \\ &\qquad \left. + \frac{1}{n^2}\left(\frac{1}{48}\,\mathrm{vec}^T\left\{\overset{3}{\otimes} V\right\}\mathrm{vec}\{H_6\}\right)\right\} \\ &= \left(\frac{2\pi}{n}\right)^{q/2} |V|^{1/2}\, e^{nH_0} \left\{1 + \frac{a_1}{n} + \frac{1}{48n^2}\,\mathrm{vec}^T\left\{\overset{3}{\otimes} V\right\}\mathrm{vec}\{H_6\}\right\}. \end{aligned}$$

This expression contains both terms from a_1 and only one of the five terms from a_2. Setting n to one produces the exact form of the expression for the Laplace6 approximation defined in Raudenbush et al. (2000). The next section will investigate whether the inclusion of one or more additional terms improves the accuracy of the approximations.

3.2.2.1 Further extensions

Similar to the univariate version, higher-order approximations to I can be derived for the multivariate version, though the number of terms involved can quickly become prohibitively large in practice. As before, expanding S to $O\left(\frac{1}{n^4}\right)$ yields

$$S = 1 + \frac{C_{1/2}}{n^{1/2}} + \frac{C_1}{n} + \frac{C_{3/2}}{n^{3/2}} + \frac{C_2}{n^2} + \frac{C_{5/2}}{n^{5/2}} + \frac{C_3}{n^3} + \frac{C_{7/2}}{n^{7/2}} + O\left(\frac{1}{n^4}\right),$$

where the yet-to-be-derived coefficients are

$$\begin{aligned} C_3 &= \frac{1}{720}U_3^6 + \frac{1}{24}U_3^4U_4 + \frac{1}{4}U_3^2U_4^2 + \frac{1}{6}U_3^3U_5 + U_3U_4U_5 + \frac{1}{2}U_3^2U_6 + \frac{1}{6}U_4^3 \\ &\quad + U_4U_6 + \frac{1}{2}U_5^2 + U_3U_7 + U_8 \end{aligned}$$

and

$$\begin{aligned} C_{7/2} &= \frac{1}{5040}U_3^7 + \frac{1}{120}U_3^5U_4 + \frac{1}{12}U_3^3U_4^2 + \frac{1}{24}U_3^4U_5 + \frac{1}{2}U_3^2U_4U_5 + \frac{1}{6}U_3U_4^3 + \frac{1}{6}U_3^3U_6 \\ &\quad + U_3U_4U_6 + \frac{1}{2}U_4^2U_5 + \frac{1}{2}U_3U_5^2 + \frac{1}{2}U_3^2U_7 + U_5U_6 + U_4U_7 + U_3U_8 + U_9. \end{aligned}$$

It follows that their expected values are

$$\begin{aligned}
E\left[C_3\right] \;=\; & \frac{85085}{82944}\,\mathrm{vec}^T\left\{\overset{9}{\otimes}\,V\right\}\mathrm{vec}\left\{\overset{6}{\otimes}\,H_3\right\} \\
& +\;\frac{25025}{9216}\,\mathrm{vec}^T\left\{\overset{8}{\otimes}\,V\right\}\mathrm{vec}\left\{\left[\overset{4}{\otimes}\,H_3\right]\otimes H_4\right\} \\
& +\;\mathrm{vec}^T\left\{\overset{7}{\otimes}\,V\right\}\left(\frac{5005}{3072}\,\mathrm{vec}\left\{\left[\overset{2}{\otimes}\,H_3\right]\otimes\left[\overset{2}{\otimes}\,H_4\right]\right\}\right. \\
& \qquad\left. +\;\frac{1001}{1152}\,\mathrm{vec}\left\{\left[\overset{3}{\otimes}\,H_3\right]\otimes H_5\right\}\right) \\
& +\;\mathrm{vec}^T\left\{\overset{6}{\otimes}\,V\right\}\left(\frac{77}{128}\,\mathrm{vec}\left\{H_3\,\otimes\,H_4\,\otimes\,H_5\right\}\right. \\
& \qquad\left. +\;\frac{77}{384}\,\mathrm{vec}\left\{\left[\overset{2}{\otimes}\,H_3\right]\otimes H_6\right\}+\frac{385}{3072}\,\mathrm{vec}\left\{\overset{3}{\otimes}\,H_4\right\}\right) \\
& +\;\mathrm{vec}^T\left\{\overset{5}{\otimes}\,V\right\}\left(\frac{7}{128}\,\mathrm{vec}\left\{H_4\,\otimes\,H_6\right\}+\frac{21}{640}\,\mathrm{vec}\left\{\overset{2}{\otimes}\,H_5\right\}\right. \\
& \qquad\left. +\;\frac{1}{32}\,\mathrm{vec}\left\{H_3\,\otimes\,H_7\right\}\right) \\
& +\;\frac{1}{384}\,\mathrm{vec}^T\left\{\overset{4}{\otimes}\,V\right\}\mathrm{vec}\left\{H_8\right\}
\end{aligned}$$

and $E\left[C_{7/2}\right]=0$. Hence

$$I \;=\; \left(\frac{2\pi}{n}\right)^{q/2}|V|^{1/2}\,e^{nH_0}\left\{1+\frac{a_1}{n}+\frac{a_2}{n^2}+\frac{a_3}{n^3}+O\left(\frac{1}{n^4}\right)\right\},$$

where $a_3=E\left[C_3\right]$. The Laplace-IV approximation is therefore defined as the extension of Laplace's method that is derived by relaxing the omissions in $E\left[S\right]$ to $O\left(\frac{1}{n^4}\right)$, so

$$I \;\approx\; \left(\frac{2\pi}{n}\right)^{q/2}|V|^{1/2}\,e^{nH_0}\left\{1+\frac{a_1}{n}+\frac{a_2}{n^2}+\frac{a_3}{n^3}\right\}.$$

An alternative extension of the Laplace approximation follows the work of Raudenbush et al. (2000), by further expansion of the approximation to $\widehat{S}$. Such refinements would be adequate when $\widehat{S}$ is a reasonably accurate approximation to S, the actual function in the expectation of Eq. 3.2; this occurs when the integrand represents a normal-like function. Now since every odd moment is zero, the expected value of any odd-ordered expansion is equivalent to the expected value of the preceding even-ordered expansion. As a result, the simplest extension from the Laplace6 approximation is derived by expanding $\widehat{S}$ to the eighth-order based on the elements of $\boldsymbol{x}$,

$$\begin{aligned}
\widehat{\widehat{S}} \;=\; & 1+\frac{1}{n^{1/2}}U_3+\frac{1}{n}\left(\frac{1}{2}U_3^2+U_4\right)+\frac{1}{n^{3/2}}\left(U_3U_4+U_5\right) \\
& +\frac{1}{n^2}\left(U_3U_5+\frac{1}{2}U_4^2+U_6\right)+\frac{1}{n^{5/2}}U_7+\frac{1}{n^3}U_8.
\end{aligned}$$

Therefore

$$
\begin{aligned}
I \;\approx\; & \left(\frac{2\pi}{n}\right)^{q/2} |V|^{1/2}\, e^{nH_0} \\
& \times \left\{1 + \frac{1}{n}\left(\frac{5}{24}\,\mathrm{vec}^T\left\{\overset{3}{\otimes} V\right\}\mathrm{vec}\left\{\overset{2}{\otimes} H_3\right\} + \frac{1}{8}\,\mathrm{vec}^T\left\{\overset{2}{\otimes} V\right\}\mathrm{vec}\left\{H_4\right\}\right)\right. \\
& \qquad + \frac{1}{n^2}\left(\mathrm{vec}^T\left\{\overset{4}{\otimes} V\right\}\left[\frac{7}{48}\,\mathrm{vec}\left\{H_3 \otimes H_5\right\} + \frac{35}{384}\,\mathrm{vec}\left\{\overset{2}{\otimes} H_4\right\}\right]\right. \\
& \qquad\qquad \left. + \frac{1}{48}\,\mathrm{vec}^T\left\{\overset{3}{\otimes} V\right\}\mathrm{vec}\left\{H_6\right\}\right) \\
& \qquad \left. + \frac{1}{n^3}\left(\frac{1}{384}\,\mathrm{vec}^T\left\{\overset{4}{\otimes} V\right\}\mathrm{vec}\left\{H_8\right\}\right)\right\} \\
= \; & \left(\frac{2\pi}{n}\right)^{q/2} |V|^{1/2}\, e^{nH_0} \\
& \times \left\{1 + \frac{a_1}{n} + \frac{1}{n^2}\left(\mathrm{vec}^T\left\{\overset{4}{\otimes} V\right\}\left[\frac{7}{48}\,\mathrm{vec}\left\{H_3 \otimes H_5\right\} + \frac{35}{384}\,\mathrm{vec}\left\{\overset{2}{\otimes} H_4\right\}\right]\right.\right. \\
& \qquad\qquad \left. + \frac{1}{48}\,\mathrm{vec}^T\left\{\overset{3}{\otimes} V\right\}\mathrm{vec}\left\{H_6\right\}\right) \\
& \qquad \left. + \frac{1}{n^3}\left(\frac{1}{384}\,\mathrm{vec}^T\left\{\overset{4}{\otimes} V\right\}\mathrm{vec}\left\{H_8\right\}\right)\right\},
\end{aligned}
$$

an expression that will subsequently be referred to as the Laplace8 approximation. This expression contains both terms from a_1, three of the five terms from a_2, and only one of the eleven terms from a_3.

3.2.2.2 Simplification to the univariate version

It is straightforward to obtain the univariate cases from these multivariate results. Of particular note, this includes expressions for the univariate Laplace6 and Laplace8 approximations, which were not discussed in Section 3.1.

3.3 Illustrative examples

Having derived expressions for approximating the values of integrals, the following section illustrates the process involved, giving examples of single and double integrals. Thus far, a positive constant, n, has been included in the exponent of the integrands used in this chapter; for simplicity, n will be set to one from here onwards. In each example presented, particular attention will also be given to the accuracy and speed of the approximations.

3.3.1 Methods of assessing accuracy and speed

The accuracy and speed of the evaluations by Laplace-related approximations will be compared to those by adaptive quadrature (AQ) via R's `integrate` function. Alternative facilities for evaluating both single and multiple integrals in R, such as `adaptIntegrate` function in the `cubature` package (Narasimhan and Johnson, 2017), were found to be much slower for the same level of accuracy. Further details on the implementation of AQ via the `integrate` function, including the accuracy requested (indicated by tolerances), were discussed in Sections 2.2.3.2 and 2.3.2.1 for single and double integrals, respectively. This section uses `integrate` to implement AQ with

default[4] and low tolerances, essentially requesting accuracy to approximately four and twelve decimal places, respectively; AQ techniques which use those tolerances will be denoted by AQ-DT and AQ-LT, respectively.

Estimating the mode of integrands for Laplace-related approximations requires stringent accuracy to ensure the first derivatives differ negligibly from zero. Approximation error may also be amplified if an inaccurate modal estimate is used when evaluating the values for all other functions in the expression. Therefore, an absolute tolerance of 10^{-12} will be stipulated when using the Newton-Raphson method (Burden and Faires, 1993, p. 56) to approximate the mode of every univariate function, and Newton's method for nonlinear systems (Burden and Faires, 1993, p. 555) to approximate the mode of every multivariate function. Selecting such a low (absolute) tolerance not only increases the number of iterations required by the process, but also increases the overall evaluation time. Indeed, a less stringent tolerance may suffice in many applications. However, locating the mode may require relatively little time if (as was the case here) an efficient numerical algorithm is used with reasonable initial values.

Evaluation speeds were assessed with the use of R's `microbenchmark` package (Mersmann, 2018). All computations were performed in R on an iMac 2.8GHz quad-core Intel Core i5 with 8GB of 1867MHz LPDDR3 onboard memory. To minimise the potential impact of external influences on evaluation times, the benchmarking process was run immediately after a hard reboot of the computer and switching off Wi-Fi. No other applications were loaded, and the computer was left, uninterrupted, until well after it had completed the entire process. The average time taken to approximate each integral was considered. Little variability was observed between the vast majority of individual approximation times for evaluating an integral by a particular integration technique. However, the remaining, tiny minority of individual approximation times typically took in the order of hundreds or even many thousands of times longer. This was due to garbage collection (GC), a process whereby memory that is no longer being used by a program is released and made available again when necessary. Importantly, GC is an expected event, so any approximation times affected cannot simply be classified as outliers (Neal, 2014). GC occurs automatically in R, and despite the `gc` function being available for manual GC requests, some memory releases remain stochastic (Wickham, 2015, p. 387). Neal (2014) exemplifies this point in stating that R can automatically adjust the level of memory usage which triggers a GC. The following illustrative examples evaluate each integral a very large number of times (further details are included in each example) to mitigate the risk of disproportionate GC effects on particular integrals or approximation techniques. Between each set of evaluations for a particular integral, the measurement times were written to a file, then the memory associated with those data was released for R to re-use. This was done in order to minimise the risk that memory usage could cause some integrals to require longer evaluation times than others.

3.3.2 Single integrals

Consider the intractable integral

$$I \;=\; \int_{-\infty}^{\infty} f\left(\varepsilon\right)\,\mathrm{d}\varepsilon \;=\; \int_{-\infty}^{\infty} e^{h(\varepsilon)}\,\mathrm{d}\varepsilon,$$

[4]By default, the absolute and relative tolerances are set to `.Machine$double.eps^0.25`, which is 1.2207×10^{-4} on the computer that was used for every application in this thesis.

where

$$h(\varepsilon) \;=\; -\frac{1}{2}\log\left(2\pi\tau^2\right) - \frac{\varepsilon^2}{2\tau^2} - \log\left(1+e^{-\eta-\varepsilon}\right)$$

from Eq. 2.4, having omitted the subscripts i and j from ε_i and η_{ij} for convenience. By inspection, f is a positive, infinitely differentiable function.

For fixed ε and τ^2, the function h, and therefore f, are monotonically increasing. The only component of h that varies (i.e., depends on η) here is

$$-\log\left(1+e^{-\eta-\varepsilon}\right).$$

This component approaches $-\infty$ as $\eta \to -\infty$, thus $h \to -\infty$ and $f \to 0^+$. Conversely, the above component approaches zero as $\eta \to \infty$, so f approaches the PDF of a normal distribution with mean zero and variance τ^2; the Laplace-related approximations approach the analytical solutions to I.

3.3.2.1 Derivation of the approximations for single integrals

Repeated differentiation of h with respect to ε yields

$$h'(\varepsilon) \;=\; -\frac{\varepsilon}{\tau^2} + \left(1+e^{\eta+\varepsilon}\right)^{-1}$$

and

$$h''(\varepsilon) \;=\; -\frac{1}{\tau^2} - \frac{e^{\eta+\varepsilon}}{\left(1+e^{\eta+\varepsilon}\right)^2}.$$

As h'' is always negative, let $\varepsilon = m$ be the maximum of h, which satisfies $h'(m) = 0$. Thus, f is unimodal, maximised at $\varepsilon = m$ as well. Using the notational simplifications $h_k = h^{(k)}(m)$ and $v = (-h_2)^{-1}$, the Laplace-I approximation to I is $\sqrt{2\pi v}\, e^{h_0}$.

Examination of h' reveals that

$$h'(0) \;=\; \left(1+e^{\eta}\right)^{-1} \;>\; 0$$

and

$$h'\left(\tau^2\right) \;=\; -1 + \left(1+e^{\eta+\tau^2}\right)^{-1} \;<\; 0,$$

so the mode, $m \in (0, \tau^2)$. This region must be searched numerically (analytical solutions do not exist) in order to find m such that $h'(m) = 0$, given the values of η and τ^2. The Newton-Raphson method is ideally suited to this task, with the midpoint of $\frac{1}{2}\tau^2$ being selected as a reasonable initial value. By definition, $h'(m) = 0$, so

$$m \;=\; \tau^2\left(1+e^{\eta+m}\right)^{-1},$$

which implies that the mode approaches τ^2 as $\eta \to -\infty$, and zero as $\eta \to \infty$.

Further differentiation of h with respect to ε yields the higher-order derivatives

$$\begin{aligned}
h^{(3)}(\varepsilon) \;&=\; \frac{e^{\eta+\varepsilon}}{\left(1+e^{\eta+\varepsilon}\right)^3}\left(e^{\eta+\varepsilon}-1\right),\\
h^{(4)}(\varepsilon) \;&=\; \frac{e^{\eta+\varepsilon}}{\left(1+e^{\eta+\varepsilon}\right)^4}\left(4e^{\eta+\varepsilon}-e^{2(\eta+\varepsilon)}-1\right),\\
h^{(5)}(\varepsilon) \;&=\; \frac{e^{\eta+\varepsilon}}{\left(1+e^{\eta+\varepsilon}\right)^5}\left(11e^{\eta+\varepsilon}-11e^{2(\eta+\varepsilon)}+e^{3(\eta+\varepsilon)}-1\right),
\end{aligned}$$

$$h^{(6)}(\varepsilon) = \frac{e^{\eta+\varepsilon}}{(1+e^{\eta+\varepsilon})^6}\Big(26e^{\eta+\varepsilon} - 66e^{2(\eta+\varepsilon)} + 26e^{3(\eta+\varepsilon)} - e^{4(\eta+\varepsilon)} - 1\Big),$$

$$h^{(7)}(\varepsilon) = \frac{e^{\eta+\varepsilon}}{(1+e^{\eta+\varepsilon})^7}\Big(57e^{\eta+\varepsilon} - 302e^{2(\eta+\varepsilon)} + 302e^{3(\eta+\varepsilon)} - 57e^{4(\eta+\varepsilon)} + e^{5(\eta+\varepsilon)} - 1\Big),$$

and

$$h^{(8)}(\varepsilon) = \frac{e^{\eta+\varepsilon}}{(1+e^{\eta+\varepsilon})^8}\Big(120e^{\eta+\varepsilon} - 1191e^{2(\eta+\varepsilon)} + 2416e^{3(\eta+\varepsilon)} - 1191e^{4(\eta+\varepsilon)} + 120e^{5(\eta+\varepsilon)} - e^{6(\eta+\varepsilon)} - 1\Big).$$

Additional to the above notational simplifications (using h_k and v), denote

$$a_1 = \frac{5}{24}h_3^2v^3 + \frac{1}{8}h_4v^2,$$

$$a_2 = \frac{385}{1152}h_3^4v^6 + \frac{35}{64}h_3^2h_4v^5 + \frac{7}{48}h_3h_5v^4 + \frac{35}{384}h_4^2v^4 + \frac{1}{48}h_6v^3,$$

and

$$a_3 = \frac{85085}{82944}h_3^6v^9 + \frac{25025}{9216}h_3^4h_4v^8 + \frac{5005}{3072}h_3^2h_4^2v^7 + \frac{1001}{1152}h_3^3h_5v^7 + \frac{77}{128}h_3h_4h_5v^6 + \frac{77}{384}h_3^2h_6v^6 + \frac{385}{3072}h_4^3v^6 + \frac{7}{128}h_4h_6v^5 + \frac{21}{640}h_5^2v^5 + \frac{1}{32}h_3h_7v^5 + \frac{1}{384}h_8v^4,$$

so the Laplace-II, Laplace-III, and Laplace-IV approximations to I are $\sqrt{2\pi v}\, e^{h_0}\,(1+a_1)$, $\sqrt{2\pi v}\, e^{h_0}\,(1+a_1+a_2)$, and $\sqrt{2\pi v}\, e^{h_0}\,(1+a_1+a_2+a_3)$ respectively. Furthermore, the Laplace6 and Laplace8 approximations to I are

$$\sqrt{2\pi v}\, e^{h_0}\left(1 + a_1 + \frac{1}{48}h_6v^3\right)$$

and

$$\sqrt{2\pi v}\, e^{h_0}\left(1 + a_1 + \frac{7}{48}h_3h_5v^4 + \frac{35}{384}h_4^2v^4 + \frac{1}{48}h_6v^3 + \frac{1}{384}h_8v^4\right),$$

respectively.

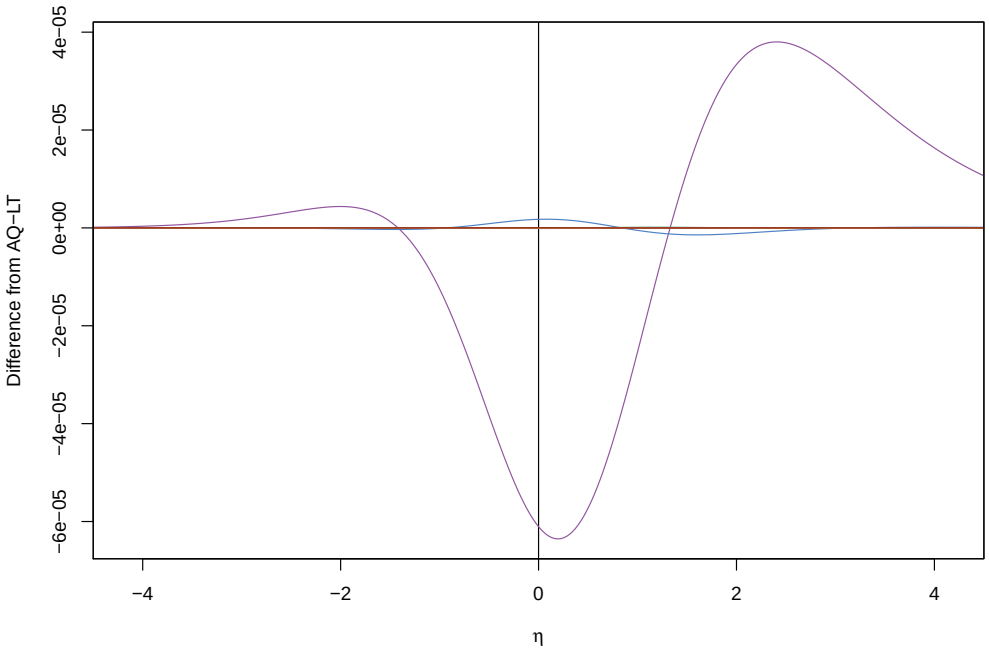
3.3.2.2 Results from the approximations

The R code that was used to evaluate I for specific parameter values is included in Appendix B.1. Following the example in Section 2.2.3.2, τ^2 is set to 0.091809 here. Comparisons between the estimates obtained by AQ-DT and AQ-LT revealed the differences to be remarkably small given the contrasting tolerances used by each AQ technique. In fact, the maximum absolute difference observed between the corresponding AQ estimates was only 10^{-10}. This may be due to how the integrands were mapped from $(-\infty, \infty)$ to $(-1, 1)$ before the integrals were evaluated. Due to such similarities between AQ-DT and AQ-LT here, the accuracy of those AQ-DT estimates will not be discussed further in this example. However, observing those results motivated further investigation into the estimates associated with other values of τ^2. Appendix B.1 contains the R code used for part of this investigation,

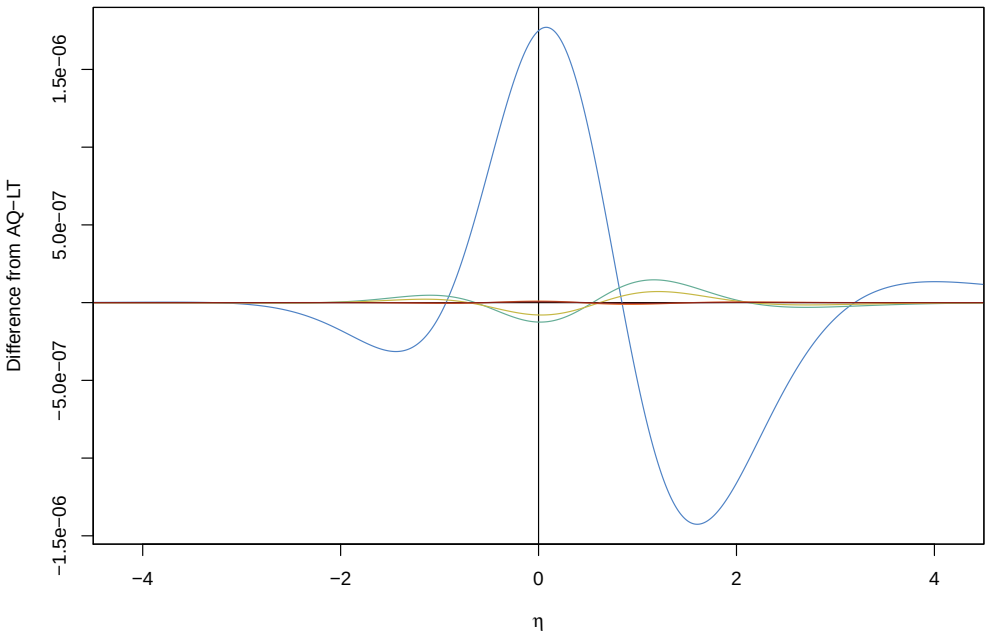
with the resultant plots, shown in Figure B.1, revealing apparent anomalies involving some values of τ^2. A rather large spike, representing the differences between AQ-DT and AQ-LT, appeared in some plots around a different range of values for η. Each spike had a different magnitude, though still remained within the requested tolerance levels. Comparisons between each set of AQ estimates and the corresponding Laplace-related approximations (not included in this thesis) indicated that the spikes were exclusively attributable to the AQ-DT estimates. Essentially, the spikes diminished as the AQ tolerances became increasingly stringent.

Based on a relatively wide range of values for η, Figure 3.1 contains two plots that display the differences between the Laplace-related approximations and AQ-LT estimates. Below both plots is a legend that displays the colour coding for identifying which lines correspond to the Laplace-I approximation and its extensions. The Laplace-I results are included in the Figure 3.1a plot, but excluded from the plot in Figure 3.1b. The latter plot uses a smaller vertical scale, enabling easier inspection of the differences between the extensions to Laplace-I and the AQ-LT estimates. Overall, these plots reveal that the accuracy of Laplace-related approximations improved with each subsequent, higher-order term included. The exclusion of terms from the Laplace-III approximation had a noticeable, albeit fairly slight, detrimental effect on the Laplace6 approximation. Furthermore, the differences between the Laplace-IV and Laplace8 approximations were miniscule here. This example suggests that Laplace-I would be adequate if accuracy to four decimal places was regarded as sufficient because each difference from the corresponding AQ-LT estimate was smaller in magnitude than had actually been requested of AQ-DT in R. Moreover, the same comparator is used for the plots in Figures 2.1b and 3.1, and indicates that the Laplace-I approximation was usually more accurate than the simple approximation technique discussed in Section 2.2.3.1.

For each value of η between ± 4 in 0.1 increments, Figure 3.2 displays the average time from evaluating each of the 81 integrals 250,000 times. Similar to before, the legend displays the colour coding and line types (solid or broken) for identifying the results of each technique. The times are are shown in milliseconds on the base-10 logarithmic scale, which enables the results of each technique to be included on the same plot. For example, the value of -2 on the vertical axis corresponds to $10^{-2} = 0.01$ milliseconds. As expected, AQ-LT was slower than AQ-DT, though given the minimal improvement in accuracy, the latter technique would be preferable here. The presence of plateaus and valleys in the plotted AQ-LT evaluation times may be due to the underlying algorithm used by `integrate` in R. The plots of the times taken for the other techniques reveal them to be fairly constant over the entire range of values considered for η. Closer scrutiny of the overall average times, by technique, revealed that AQ-DT ($0.0286 \approx 10^{-1.54}$ ms) took around four times as long as Laplace-IV ($0.0073 \approx 10^{-2.14}$ ms), and ten times as long as Laplace-I ($0.0027 \approx 10^{-2.57}$ ms). Indeed, it is unsurprising that Laplace-I was fastest, and the evaluation times progressively increased with the inclusion of each subsequent order in the approximation. It is also expected that the evaluation times would increase from Laplace-II to Laplace6 then Laplace-III because the terms contained in each preceding expression is a subset of the terms contained in each subsequent expression. However, the terms contained in Laplace-III and Laplace8 are not subsets of each other, so it is more difficult to predict which technique will usually be quicker. From the present benchmarking procedure, Laplace-III was marginally quicker than Laplace8.



(a) All Laplace-related approximations



(b) Enlargement for extensions to the Laplace approximation

Figure 3.1: Differences from low-tolerance adaptive quadrature of the Laplace-related approximations to the single integrals.

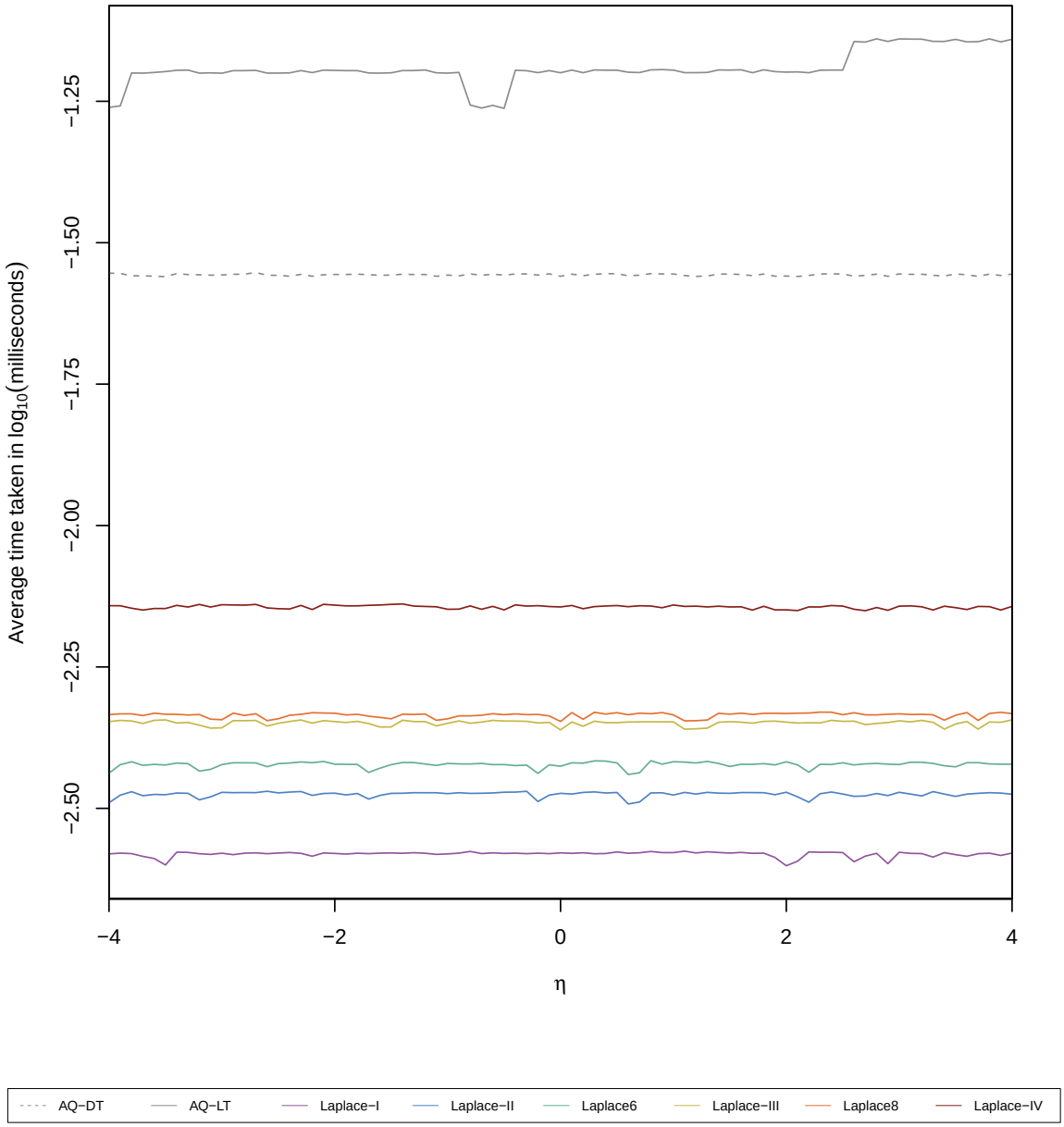


Figure 3.2: The average time taken to approximate the single integrals, in milliseconds on the base-10 logarithmic scale.

3.3.3 Double integrals

Consider the intractable integral

$$I \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\left(\varepsilon_0,\varepsilon_1\right)\,\mathrm{d}\varepsilon_0\,\mathrm{d}\varepsilon_1 \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{h(\varepsilon_0,\varepsilon_1)}\,\mathrm{d}\varepsilon_0\,\mathrm{d}\varepsilon_1,$$

where

$$\begin{aligned} h\left(\varepsilon_0,\varepsilon_1\right) \;=\; &-\log\left(2\pi\sqrt{\tau_0^2\tau_1^2-\tau_{01}^2}\right)-\frac{\varepsilon_0^2\tau_1^2-2\varepsilon_0\varepsilon_1\tau_{01}+\varepsilon_1^2\tau_0^2}{2\left(\tau_0^2\tau_1^2-\tau_{01}^2\right)} \\ &-\log\left(1+e^{-\mu-\alpha_1x_1-\alpha_2x_2-\varepsilon_0-\varepsilon_1x_1}\right) \end{aligned}$$

from Eq. 2.9, having omitted the subscripts i and j from ε_{0i}, ε_{1i}, x_{1ij}, and x_{2ij} for convenience. By inspection, f is a positive, infinitely differentiable function. Based on the Glance et al. (2003) example discussed in Section 2.3.2.1, the parameters are set to $\mu = -8.100$, $\alpha_1 = 1.094$, $\alpha_2 = 0.063$, $\tau_0^2 = 2.412$, $\tau_1^2 = 0.196$, and $\tau_{01} = -0.680$ here. Furthermore, $x_1 = \log\left(x_2+1\right)$, where x_2 represents the Simplified Acute Physiology Score (SAPS) II, an integer-valued predictor of mortality, ranging between 0 and 125 with the median of 30. Hence, the example essentially limits the number of predictor variables to one, the SAPS II score. For fixed ε_0 and ε_1, the function h, and therefore f, are monotonically increasing. The only component in h that varies (i.e., depends on x_1 and x_2 through the SAPS II score) is

$$-\log\left(1+e^{-\mu-\alpha_1x_1-\alpha_2x_2-\varepsilon_0-\varepsilon_1x_1}\right).$$

This component is always negative, with its magnitude reducing as the SAPS II score increases because both α_1 and α_2 are positive. Therefore, as the SAPS II score increases, the function f resembles more closely the PDF of the bivariate normal distribution, with mean vector $\mathbf{0}$ and covariance matrix

$$\Upsilon = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}.$$

3.3.3.1 Derivation of the approximations for double integrals

From here onwards, the simplification

$$\eta \;=\; \mu+\alpha_1x_1+\alpha_2x_2$$

will be used to reduce the algebraic detail involved in subsequent expressions. In doing so,

$$h\left(\varepsilon_0,\varepsilon_1\right) = -\log\left(2\pi\sqrt{\tau_0^2\tau_1^2-\tau_{01}^2}\right)-\frac{\varepsilon_0^2\tau_1^2-2\varepsilon_0\varepsilon_1\tau_{01}+\varepsilon_1^2\tau_0^2}{2\left(\tau_0^2\tau_1^2-\tau_{01}^2\right)}-\log\left(1+e^{-\eta-\varepsilon_0-\varepsilon_1x_1}\right).$$

Now the two-element row vector

$$h^{(1)}(\boldsymbol{\varepsilon}) \;=\; \frac{\partial\left[\mathrm{vec}\left\{h\left(\boldsymbol{\varepsilon}\right)\right\}\right]}{\partial\boldsymbol{\varepsilon}^{\boldsymbol{T}}} \;=\; \frac{\partial\,h(\boldsymbol{\varepsilon})}{\partial\boldsymbol{\varepsilon}^{\boldsymbol{T}}} \;=\; \left[\frac{\partial\,h(\boldsymbol{\varepsilon})}{\partial\varepsilon_0},\;\frac{\partial\,h(\boldsymbol{\varepsilon})}{\partial\varepsilon_1}\right],$$

where

$$\frac{\partial\,h(\boldsymbol{\varepsilon})}{\partial\varepsilon_0} \;=\; -\frac{\varepsilon_0\tau_1^2-\varepsilon_1\tau_{01}}{\tau_0^2\tau_1^2-\tau_{01}^2}+\left(1+e^{\eta+\varepsilon_0+\varepsilon_1x_1}\right)^{-1},$$

and

$$\frac{\partial\, h(\boldsymbol{\varepsilon})}{\partial \varepsilon_1} \;=\; -\frac{\varepsilon_1\tau_0^2 - \varepsilon_0\tau_{01}}{\tau_0^2\tau_1^2 - \tau_{01}^2} + x_1\left(1 + e^{\eta+\varepsilon_0+\varepsilon_1 x_1}\right)^{-1}.$$

Therefore, the 2×2 matrix

$$h^{(2)}(\boldsymbol{\varepsilon}) \;=\; \frac{\partial\left[\mathrm{vec}\left\{h^{(1)}(\boldsymbol{\varepsilon})\right\}\right]}{\partial \boldsymbol{\varepsilon}^{\boldsymbol{T}}} \;=\; \frac{\partial}{\partial \boldsymbol{\varepsilon}^{\boldsymbol{T}}}\begin{bmatrix}\frac{\partial\, h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0}\\ \frac{\partial\, h(\boldsymbol{\varepsilon})}{\partial \varepsilon_1}\end{bmatrix} \;=\; \begin{bmatrix}\frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^2} & \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0 \partial \varepsilon_1}\\ \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0 \partial \varepsilon_1} & \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_1^2}\end{bmatrix},$$

where the second-order partial derivatives of h are

$$\begin{aligned}\frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^2} &= -\frac{\tau_1^2}{\tau_0^2\tau_1^2 - \tau_{01}^2} - \frac{e^{\eta+\varepsilon_0+\varepsilon_1 x_1}}{\left(1 + e^{\eta+\varepsilon_0+\varepsilon_1 x_1}\right)^2},\\ \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0 \partial \varepsilon_1} &= \frac{\tau_{01}}{\tau_0^2\tau_1^2 - \tau_{01}^2} - \frac{x_1\, e^{\eta+\varepsilon_0+\varepsilon_1 x_1}}{\left(1 + e^{\eta+\varepsilon_0+\varepsilon_1 x_1}\right)^2},\end{aligned}$$

and

$$\frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_1^2} \;=\; -\frac{\tau_0^2}{\tau_0^2\tau_1^2 - \tau_{01}^2} - \frac{x_1^2\, e^{\eta+\varepsilon_0+\varepsilon_1 x_1}}{\left(1 + e^{\eta+\varepsilon_0+\varepsilon_1 x_1}\right)^2}.$$

Both $\frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^2}$ and $\frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_1^2}$ are always negative, so let the maximum of h occur at $\boldsymbol{\varepsilon} = \boldsymbol{m} = [m_0, m_1]^T$, where $h^{(1)}(\boldsymbol{m}) = \mathbf{0}^{\boldsymbol{T}}$ and $\left|h^{(2)}(\boldsymbol{m})\right| > 0$. Newton's method for nonlinear systems is ideally suited to the task of locating $\boldsymbol{m}$, with the origin being selected as a reasonable initial estimate.

By denoting $H_k = h^{(k)}(\boldsymbol{m})$ for $k \in \mathbb{Z}_{\geq 0}$, and $V = [-H_2]^{-1}$, the Laplace-I approximation to I is $2\pi|V|^{1/2}e^{H_0}$. This example will now derive the higher-order terms for expanded versions of the Laplace approximation.

The 4×2 matrix

$$h^{(3)}(\boldsymbol{\varepsilon}) \;=\; \frac{\partial\left[\mathrm{vec}\left\{h^{(2)}(\boldsymbol{\varepsilon})\right\}\right]}{\partial \boldsymbol{\varepsilon}^{\boldsymbol{T}}} \;=\; \frac{\partial}{\partial \boldsymbol{\varepsilon}^{\boldsymbol{T}}}\begin{bmatrix}\frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^2}\\ \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0 \partial \varepsilon_1}\\ \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0 \partial \varepsilon_1}\\ \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_1^2}\end{bmatrix} \;=\; \begin{bmatrix}\frac{\partial^3 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^3} & \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^2 \partial \varepsilon_1}\\ \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^2 \partial \varepsilon_1} & \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0 \partial \varepsilon_1^2}\\ \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^2 \partial \varepsilon_1} & \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0 \partial \varepsilon_1^2}\\ \frac{\partial^2 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0 \partial \varepsilon_1^2} & \frac{\partial^3 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_1^3}\end{bmatrix},$$

where the third-order partial derivatives of h are

$$\frac{\partial^3 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^3} \;=\; \frac{e^{\eta+\varepsilon_0+\varepsilon_1 x_1}}{\left(1 + e^{\eta+\varepsilon_0+\varepsilon_1 x_1}\right)^3}\left(e^{\eta+\varepsilon_0+\varepsilon_1 x_1} - 1\right)$$

and

$$\frac{\partial^3 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^{3-k} \partial \varepsilon_1^k} \;=\; x_1^k\, \frac{\partial^3 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^3},$$

for $k = 1$, 2, and 3. Thus, $h^{(3)}(\boldsymbol{\varepsilon}) = G_3\, \frac{\partial^3 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^3}$, where the matrix

$$G_3 \;=\; \begin{bmatrix}1 & x_1\\ x_1 & x_1^2\\ x_1 & x_1^2\\ x_1^2 & x_1^3\end{bmatrix}.$$

Proceeding to the fourth order of differentiation, the 8×2 matrix

$$\begin{aligned} h^{(4)}(\boldsymbol{\varepsilon}) &= \frac{\partial \left[\operatorname{vec}\left\{h^{(3)}(\boldsymbol{\varepsilon})\right\}\right]}{\partial \boldsymbol{\varepsilon}^{\boldsymbol{T}}} \\ &= \operatorname{vec}\{G_3\} \frac{\partial}{\partial \boldsymbol{\varepsilon}^{\boldsymbol{T}}} \left[\frac{\partial^3 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^3}\right] \\ &= \operatorname{vec}\{G_3\} \left[\frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^4}, \ \frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^3 \partial \varepsilon_1}\right]. \end{aligned} \tag{3.5}$$

Since the fourth-order partial derivatives of h are

$$\frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^4} = \frac{e^{\eta+\varepsilon_0+\varepsilon_1 x_1}}{\left(1+e^{\eta+\varepsilon_0+\varepsilon_1 x_1}\right)^4} \left(4e^{\eta+\varepsilon_0+\varepsilon_1 x_1} - e^{2(\eta+\varepsilon_0+\varepsilon_1 x_1)} - 1\right)$$

and

$$\frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^3 \partial \varepsilon_1} = x_1 \frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^4},$$

then

$$\left[\frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^4}, \ \frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^3 \partial \varepsilon_1}\right] = [1, \, x_1] \frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^4}.$$

Therefore, Eq. 3.5 becomes

$$h^{(4)}(\boldsymbol{\varepsilon}) = \begin{bmatrix} 1 \\ x_1 \\ x_1 \\ x_1^2 \\ x_1 \\ x_1^2 \\ x_1^2 \\ x_1^3 \end{bmatrix} [1, \, x_1] \frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^4} = \begin{bmatrix} 1 & x_1 \\ x_1 & x_1^2 \\ x_1 & x_1^3 \\ x_1^2 & x_1^3 \\ x_1 & x_1^2 \\ x_1^2 & x_1^3 \\ x_1^2 & x_1^3 \\ x_1^3 & x_1^4 \end{bmatrix} \frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^4} = G_4 \frac{\partial^4 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^4}.$$

It follows from the preceding derivations, for $k \in \mathbb{Z}_{\geq 3}$, that the sequence

$$h^{(k)}(\boldsymbol{\varepsilon}) = G_k \frac{\partial^k h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^k},$$

where

$$G_k = \begin{cases} \operatorname{vec}\{G_{k-1}\} \, [1, \, x_1] & \text{if } k \geq 4 \\ \begin{bmatrix} 1 & x_1 & x_1 & x_1^2 \\ x_1 & x_1^2 & x_1^2 & x_1^3 \end{bmatrix}^T & \text{if } k = 3 \end{cases}$$

is a $2^{k-1} \times 2$ matrix. Further details regarding the G_k matrices (i.e., $k \geq 5$) have been omitted from this thesis because the number of rows doubles with each subsequent k, thus growing very quickly. Expressions for the $h^{(k)}(\boldsymbol{\varepsilon})$, where $k = 5$, 6, 7, and 8, otherwise follow immediately after obtaining the partial derivatives of h with respect to ε_0,

$$\frac{\partial^5 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^5} = \frac{e^{\eta+\varepsilon_0+\varepsilon_1 x_1}}{\left(1+e^{\eta+\varepsilon_0+\varepsilon_1 x_1}\right)^5} \left(11e^{\eta+\varepsilon_0+\varepsilon_1 x_1} - 11e^{2(\eta+\varepsilon_0+\varepsilon_1 x_1)} + e^{3(\eta+\varepsilon_0+\varepsilon_1 x_1)} - 1\right),$$

$$\begin{aligned}
\frac{\partial^6 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^6} &= \frac{e^{\eta+\varepsilon_0+\varepsilon_1 x_1}}{(1+e^{\eta+\varepsilon_0+\varepsilon_1 x_1})^6}\Big(26e^{\eta+\varepsilon_0+\varepsilon_1 x_1} - 66e^{2(\eta+\varepsilon_0+\varepsilon_1 x_1)} \\
&\qquad + 26e^{3(\eta+\varepsilon_0+\varepsilon_1 x_1)} - e^{4(\eta+\varepsilon_0+\varepsilon_1 x_1)} - 1\Big), \\
\frac{\partial^7 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^7} &= \frac{e^{\eta+\varepsilon_0+\varepsilon_1 x_1}}{(1+e^{\eta+\varepsilon_0+\varepsilon_1 x_1})^7}\Big(57e^{\eta+\varepsilon_0+\varepsilon_1 x_1} - 302e^{2(\eta+\varepsilon_0+\varepsilon_1 x_1)} + 302e^{3(\eta+\varepsilon_0+\varepsilon_1 x_1)} \\
&\qquad - 57e^{4(\eta+\varepsilon_0+\varepsilon_1 x_1)} + e^{5(\eta+\varepsilon_0+\varepsilon_1 x_1)} - 1\Big),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^8 h(\boldsymbol{\varepsilon})}{\partial \varepsilon_0^8} &= \frac{e^{\eta+\varepsilon_0+\varepsilon_1 x_1}}{(1+e^{\eta+\varepsilon_0+\varepsilon_1 x_1})^8}\Big(120e^{\eta+\varepsilon_0+\varepsilon_1 x_1} - 1191e^{2(\eta+\varepsilon_0+\varepsilon_1 x_1)} \\
&\qquad + 2416e^{3(\eta+\varepsilon_0+\varepsilon_1 x_1)} - 1191e^{4(\eta+\varepsilon_0+\varepsilon_1 x_1)} \\
&\qquad + 120e^{5(\eta+\varepsilon_0+\varepsilon_1 x_1)} - e^{6(\eta+\varepsilon_0+\varepsilon_1 x_1)} - 1\Big).
\end{aligned}$$

Additional to the above notational simplifications (using H_k and V), denote

$$\begin{aligned}
a_1 &= \frac{5}{24}\operatorname{vec}^T\left\{\overset{3}{\otimes} V\right\}\operatorname{vec}\left\{\overset{2}{\otimes} H_3\right\} + \frac{1}{8}\operatorname{vec}^T\left\{\overset{2}{\otimes} V\right\}\operatorname{vec}\{H_4\}, \\
a_2 &= \frac{385}{1152}\operatorname{vec}^T\left\{\overset{6}{\otimes} V\right\}\operatorname{vec}\left\{\overset{4}{\otimes} H_3\right\} \\
&\quad + \frac{35}{64}\operatorname{vec}^T\left\{\overset{5}{\otimes} V\right\}\operatorname{vec}\left\{\left[\overset{2}{\otimes} H_3\right]\otimes H_4\right\} \\
&\quad + \operatorname{vec}^T\left\{\overset{4}{\otimes} V\right\}\left(\frac{7}{48}\operatorname{vec}\{H_3\otimes H_5\} + \frac{35}{384}\operatorname{vec}\left\{\overset{2}{\otimes} H_4\right\}\right) \\
&\quad + \frac{1}{48}\operatorname{vec}^T\left\{\overset{3}{\otimes} V\right\}\operatorname{vec}\{H_6\},
\end{aligned}$$

and

$$\begin{aligned}
a_3 &= \frac{85085}{82944}\operatorname{vec}^T\left\{\overset{9}{\otimes} V\right\}\operatorname{vec}\left\{\overset{6}{\otimes} H_3\right\} \\
&\quad + \frac{25025}{9216}\operatorname{vec}^T\left\{\overset{8}{\otimes} V\right\}\operatorname{vec}\left\{\left[\overset{4}{\otimes} H_3\right]\otimes H_4\right\} \\
&\quad + \operatorname{vec}^T\left\{\overset{7}{\otimes} V\right\}\left(\frac{5005}{3072}\operatorname{vec}\left\{\left[\overset{2}{\otimes} H_3\right]\otimes\left[\overset{2}{\otimes} H_4\right]\right\}\right. \\
&\qquad\qquad \left. + \frac{1001}{1152}\operatorname{vec}\left\{\left[\overset{3}{\otimes} H_3\right]\otimes H_5\right\}\right) \\
&\quad + \operatorname{vec}^T\left\{\overset{6}{\otimes} V\right\}\left(\frac{77}{128}\operatorname{vec}\{H_3\otimes H_4\otimes H_5\}\right. \\
&\qquad\qquad \left. + \frac{77}{384}\operatorname{vec}\left\{\left[\overset{2}{\otimes} H_3\right]\otimes H_6\right\} + \frac{385}{3072}\operatorname{vec}\left\{\overset{3}{\otimes} H_4\right\}\right) \\
&\quad + \operatorname{vec}^T\left\{\overset{5}{\otimes} V\right\}\left(\frac{7}{128}\operatorname{vec}\{H_4\otimes H_6\} + \frac{21}{640}\operatorname{vec}\left\{\overset{2}{\otimes} H_5\right\}\right. \\
&\qquad\qquad \left. + \frac{1}{32}\operatorname{vec}\{H_3\otimes H_7\}\right) \\
&\quad + \frac{1}{384}\operatorname{vec}^T\left\{\overset{4}{\otimes} V\right\}\operatorname{vec}\{H_8\},
\end{aligned}$$

so the Laplace-II, Laplace-III, and Laplace-IV approximations to I are $2\pi|V|^{1/2}e^{H_0}(1+a_1)$, $2\pi|V|^{1/2}e^{H_0}(1+a_1+a_2)$, and $2\pi|V|^{1/2}e^{H_0}(1+a_1+a_2+a_3)$, respectively. Furthermore, the Laplace6 and Laplace8 approximations to I are

$$2\pi|V|^{1/2}e^{H_0}\left(1+a_1+\frac{1}{48}\,\mathrm{vec}^T\left\{\overset{3}{\otimes}V\right\}\mathrm{vec}\,\{H_6\}\right)$$

and

$$2\pi|V|^{1/2}e^{H_0}\left(1+a_1+\mathrm{vec}^T\left\{\overset{4}{\otimes}V\right\}\left[\frac{7}{48}\,\mathrm{vec}\,\{H_3\otimes H_5\}+\frac{35}{384}\,\mathrm{vec}\left\{\overset{2}{\otimes}H_4\right\}\right]+\frac{1}{48}\,\mathrm{vec}^T\left\{\overset{3}{\otimes}V\right\}\mathrm{vec}\,\{H_6\}+\frac{1}{384}\,\mathrm{vec}^T\left\{\overset{4}{\otimes}V\right\}\mathrm{vec}\,\{H_8\}\right),$$

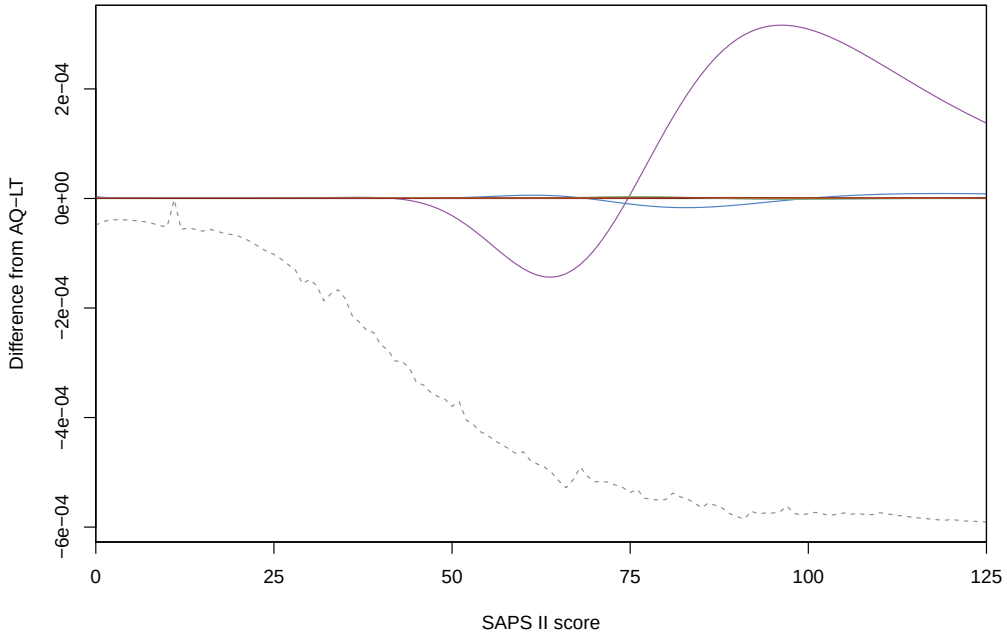
respectively.

3.3.3.2 Results from the approximations

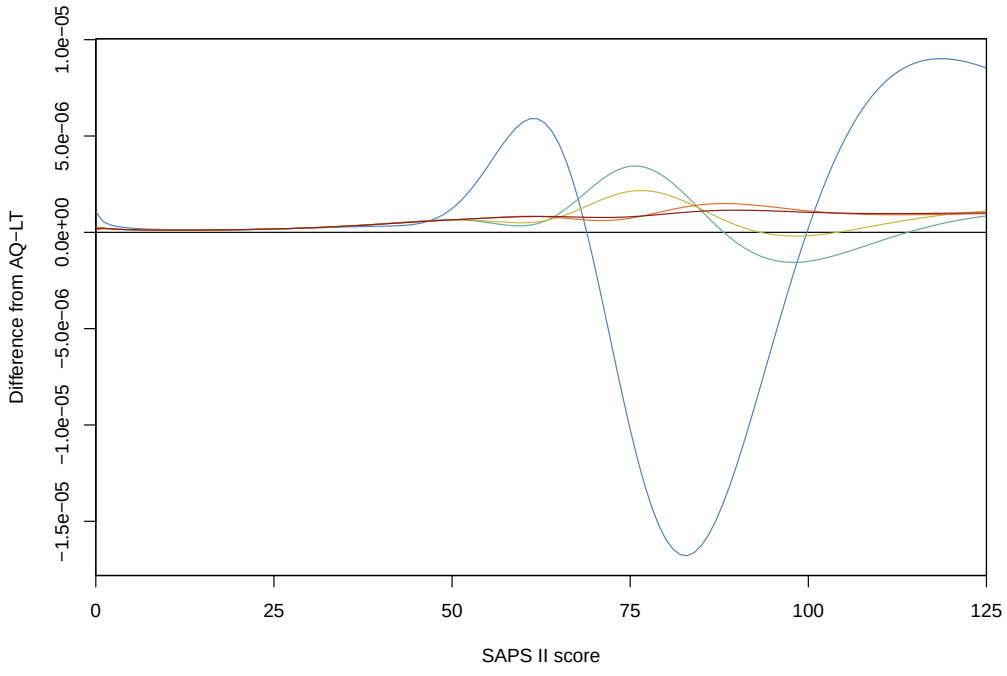
The R code that was used to evaluate I for specific parameter values is included in Appendix B.2. Figure 3.3 contains two plots that display the differences from the AQ-LT estimates of both AQ-DT and Laplace-related approximations. It would seem reasonable to expect that AQ-LT estimates are extremely close to the actual values due to the extreme accuracy requested by this technique. The legend included below both plots displays the colour coding and line types for identifying the results associated with each approximation. For all possible SAPS II scores in Glance et al. (2003), the AQ-DT and Laplace-I results are included in the Figure 3.3a plot, but excluded from the plot in Figure 3.3b. The latter plot uses a smaller vertical scale, enabling easier inspection of the differences between the extensions to Laplace-I and the AQ-LT estimates.

The estimates by AQ-DT were consistently lower than AQ-LT. In most instances, the differences increased along with the SAPS II scores, up to a maximum of 0.0006. This example demonstrates that AQ-DT was accurate to around four decimal places, thus closely matching the accuracy actually requested of R (by default), and contrasting with the much greater accuracy achieved in the previous example (Section 3.3.2.2). One striking feature of the AQ-DT line was the sharp spike at the score of 11, where the difference between the AQ-DT and AQ-LT estimates was minimised. Closer scrutiny revealed that the difference was 2×10^{-6}, which may be attributable to the algorithm used by the `integrate` function in R. An investigation into the differences between the AQ estimates, similar to Section 3.3.2.2, was conducted here. Intermediate (non-integer) values for the SAPS II scores were considered, along with different values for the elements of matrix Υ (i.e., the values for τ_0^2, τ_1^2, and τ_{01}). Appendix B.2 contains the R code used for part of this investigation, with the resultant plots shown in Figure B.2. These plots reveal apparent anomalies involving regions, corresponding to SAPS II scores, where the differences between AQ-DT and AQ-LT seemed remarkably large or small. Nevertheless, the magnitude of each difference remained consistently within the requested tolerance levels.

Noting that the median SAPS II score was 30, the estimates by all Laplace-related approximations were extremely accurate across the majority of patients' scores, with very slight positive differences from the corresponding AQ-LT estimates until the scores reached at least 41. The range of differences between the Laplace-I and AQ-LT estimates was $(-0.0001, 0.0003)$, though the most extreme differences only occurred at much higher SAPS II scores. Comparisons between Figures 2.2b and 3.3a reveals



(a) AQ-DT and all Laplace-related approximations



(b) Enlargement for extensions to the Laplace approximation

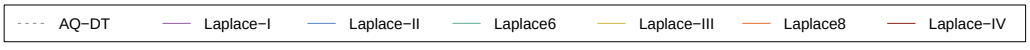


Figure 3.3: Differences from low-tolerance adaptive quadrature of the other approximations to the double integrals.

that the Laplace-I approximation was generally much more accurate than the simple approximation technique used by Glance et al. (2003). Returning to Figure 3.3b, this plot reveals that, overall, the accuracy of extensions to Laplace-I improved with each subsequent, higher-order term included. The lines associated with each subsequent extension to the Laplace approximation flattened out towards zero for all differences from AQ-LT; however, the Laplace8 and Laplace-IV estimates remained consistently, albeit very slightly, above the AQ-LT estimates. From these results, the use of Laplace-related approximations may raise concerns that mortality probabilities could always (or at least almost always) be overestimated, despite each discrepancy seeming rather miniscule. Potential objections that may arise from using these techniques would therefore include how the consistent (or near-consistent) overestimation of each patient's mortality risks may give the impression that each provider performed slightly better than was actually the case.

For every SAPS II score between 0 and 125, Figure 3.4 displays the average time for evaluating each of the 126 double integrals 5,000 times. This example ran far fewer repetitions (only 2%) than the preceding, single integral example because each double integral took considerably longer to evaluate; in fact, the entire process required six days to complete. As before, the legend displays the colour coding and line types for identifying the results of each technique, and the times are are shown in milliseconds on the base-10 logarithmic scale. Excluding AQ-DT, the plots of the times taken for each technique were fairly constant across over the entire range of SAPS II scores. The AQ-DT times fluctuated dramatically, though increased overall, until reaching SAPS II scores in the forties (i.e., considerably above the median), then flattening out through the remaining scores. A sharp spike for AQ-DT at the score of 11 is present in this plot too, which indicates that these evaluations were much quicker and more accurate. In this example, AQ-DT was quicker than Laplace-IV for the majority of patients' scores, though much less accurate than even Laplace-I. Another particularly striking difference from the single integral example is that, here, Laplace8 was quicker than Laplace-III. This may be attributable to the additional computational expense required for matrix operations in R. Closer scrutiny of the overall average times for the fastest and slowest techniques revealed that AQ-LT ($566.5 \approx 10^{2.75}$ ms) took around 22 times as long as Laplace-IV ($25.36 \approx 10^{1.40}$ ms), and 17,593 times as long as Laplace-I ($0.0322 \approx 10^{-1.49}$ ms).

3.4 Future work

Subsequent work in this thesis (in Chapter 4) will use the Laplace approximation to predict mortality probabilities based on previously fitted models. The most complicated usage is for predicting probabilities by integrating over posterior distributions. Each probability is calculated using two applications of the Laplace-IV approximation. One approximation is for integrating the probability function over the kernel of the posterior distribution, an unrecognisable distribution from its PDF. A second approximation is for evaluating the normalising constant of the posterior distribution's PDF to scale the first approximation.

The Laplace approximation is also a common option available in statistical applications for fitting generalised linear mixed models. Raudenbush et al. (2000) implemented the Laplace6 extension to fit rather small models to moderate-sized datasets. The results obtained by Laplace6 were relatively fast computationally and accurate when compared to the alternative approaches (AQ, etc.) that they consid-

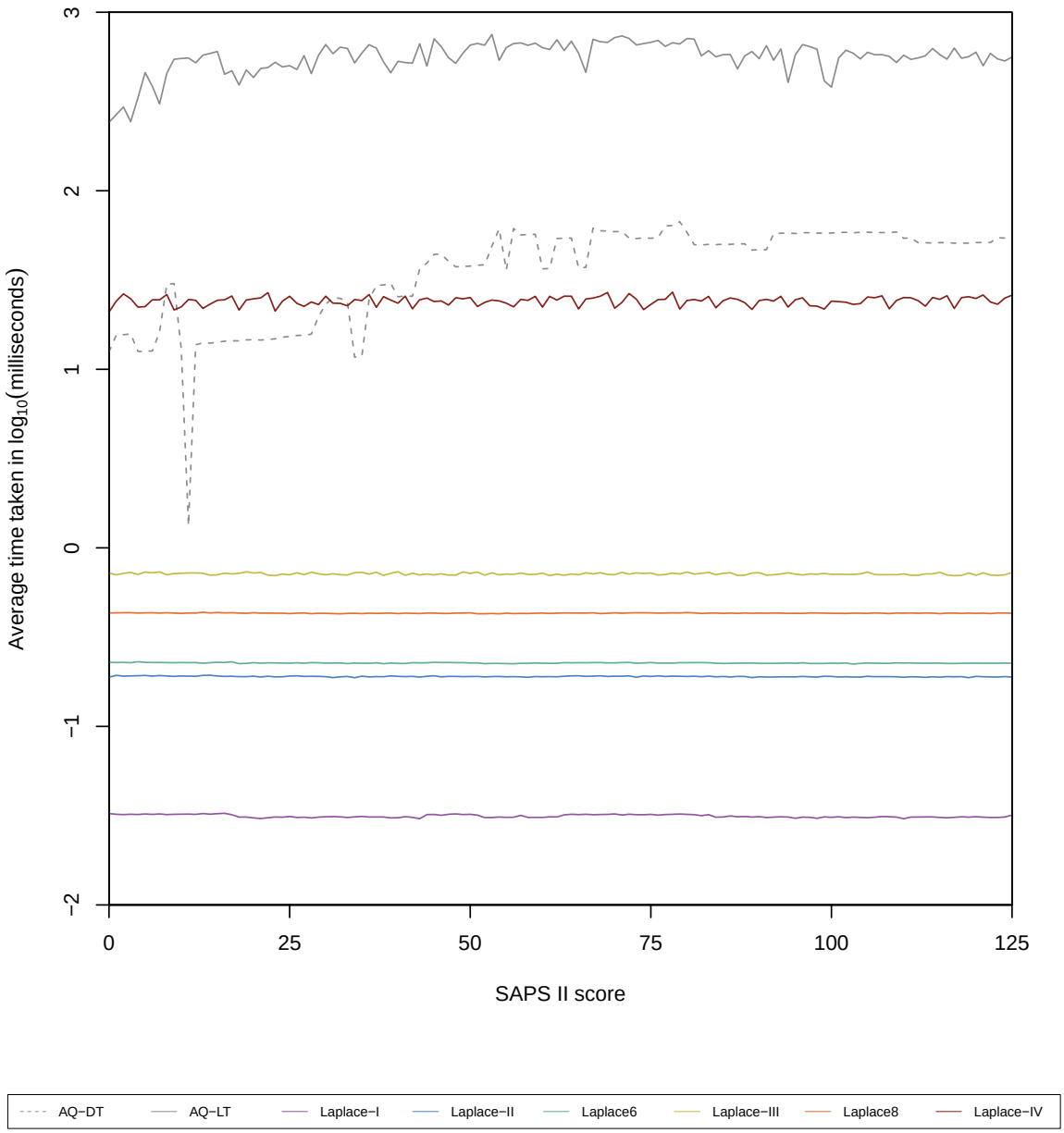


Figure 3.4: The average time taken to approximate the double integrals, in milliseconds on the base-10 logarithmic scale.

ered. Known software implementations of extensions to the Laplace approximation remain limited to Laplace6; see, for example, HLM 8 (Raudenbush et al., 2019). Following the work of Raudenbush et al. (2000), future model-fitting applications could incorporate other extensions to the Laplace approximation that were derived in this chapter. Through further investigation, guidelines may be constructed for recommending particular extensions based on the competing criteria of (approximate) accuracy and speed. Where applications require greater accuracy than the Laplace-related approximations could achieve, these techniques would still be useful in the model-fitting process. The estimates could serve as reasonably accurate, initial values to substantially reduce the overall time needed to fit models with greater accuracy.

Chapter 4

Application to the ANZICS APD

Chapter 2 described various statistics that may be used when comparing provider performance. The standardised mortality ratio (SMR) statistics are of particular interest here, primarily due to their prevalent usage in the critical care context. This chapter will examine the results obtained from applying them to critical care data from Australia and New Zealand (NZ). Due to the computational intensity involved in this application, the Laplace-related approximations developed in Chapter 3 will be implemented here.

This application uses a subset of the critical care data from the Adult Patient Database (APD), a clinical registry that is administered by the Centre for Outcome and Resource Evaluation (CORE) of the Australian and New Zealand Intensive Care Society (ANZICS). The APD was constructed to benchmark and audit adult intensive care unit (ICU) performance, with the aim of promoting improvements in patient outcomes (ANZICS CORE, 2019). For convenience, adult ICUs will hereinafter be referred to as ICUs. Since its inception in 1992, the APD has collected de-identified data on patients' admissions, submitted voluntarily by Australian and NZ ICUs (Stow et al., 2006). In fact, despite it being voluntary in principle, most ICUs currently submit data. ANZICS CORE have assisted this process by providing ICUs with data collection software and support to encourage them to contribute to the registry (ANZICS CORE, 2017). The APD is updated quarterly and retrospectively, and now contains details on over two million patient admissions, making it one of the largest databases of its kind in the world (ANZICS, 2019).

The APD contains numerous fields for the storage of extensive administrative and clinical data pertaining to each patient's ICU admission. ICUs that submit accurate and thorough data to the registry greatly assist ANZICS CORE with one of their objectives: to inform all healthcare stakeholders, including clinicians, providers, and funders (ANZICS CORE, 2019). Matters to consider include the demographics of patients, their illness severities, and treatments required. All potential contributions to the APD undergo automatic data checking, with the ICU notified of necessary validations or corrections before their submission will be accepted; however, the ICU's submission is automatically rejected if over twenty percent of their reported admissions are missing essential entries for SMR calculations (ANZICS, 2019). Problems with data quality in accepted contributions may not be investigated until the affected ICUs have been identified as potential under-performers (McClean et al., 2017). It would seem reasonable to expect occasional missing or inaccurate values in such a large database, particularly since many ICUs lack a dedicated data collector (Stow et al., 2006). Still, missing data can be especially problematic for the contributing ICU. Missing (or otherwise invalid) values for some clinical variables, such as those

requiring pathology results, are automatically replaced by normal values for healthy subjects (ANZICS CORE, 2018). It follows that some patients will be classified as (much) healthier than they actually were, and hence be assigned unrealistically low mortality probabilities. As a result, the ICUs that treated those patients will be more likely to have overestimated SMRs.

Attempts to quantify mortality risks for ICU patients have lead to the development of various prognostic models, which enable most patients' risks to be represented by illness severity scores or mortality probabilities. Section 2.3.2.1 briefly described the second version of one such prognostic system, the Simplified Acute Physiology Score (SAPS) II (Le Gall et al., 1993). Another commonly used system is the Acute Physiology and Chronic Health Evaluation (APACHE), the first of its kind to be developed (Glance et al., 2003), in 1981 (Knaus et al., 1981), and regularly refined since then (Knaus et al., 1985, 1991; Zimmerman et al., 2006). This chapter uses the most recent APACHE scores contained in the APD: from the tenth calibration of APACHE III (ANZICS CORE, 2018), the third version of APACHE (Knaus et al., 1991).[1] For each patient, their APACHE III score was set to the highest value (corresponding to the highest predicted mortality risk) calculated from a wide variety of administrative and (mainly) clinical details pertaining to the first 24 hours in the ICU. Knaus et al. (1991) constructed the age categories, whereby patients in older age categories had progressively more points added to their APACHE III scores calculated. Becker et al. in Kramer and Zimmerman (2008) noted that one of the (many) introductions with APACHE III was the use of separate models for coronary artery bypass graft (CABG) and non-CABG patients. Based on the APD, Paul et al. (2012) demonstrated that APACHE III maintained good discrimination between hospital outcomes from 2000 to 2009, despite deteriorating calibration between observed and predicted outcomes for individual patients.

4.1 Data selection

Chapter 1 detailed the matters relating to ethics and permission which needed to be addressed before any data could be obtained and used for this application. The dataset used here is a subset of the ANZICS APD that John Moran obtained access to in January 2018. He undertook preliminary work on the dataset before supplying me with a partially-cleaned copy in February 2018, where

- all re-admissions were removed, so only the first admission to an ICU could be considered for each patient, and it would be reasonable to assume that the patients are independent;
- all patients were at least 16 years old, and every patient's record contained valid entries for sex and ICU admission source; and
- 31 consolidated diagnostic categories had been constructed, which he recommended using rather than simpler alternatives, such as those in Kasza et al. (2013) or Moran and Solomon (2012).

All relevant variables will be discussed in greater detail in Section 4.2. Furthermore, Section 4.5 will consider limitations associated with the exclusion of patients due to missingness for particular variables and ICUs not reporting.

[1] The latest version is APACHE IV (Zimmerman et al., 2006), which cannot be used because the APD did not collect the data for all of the necessary variables (Paul et al., 2012).

This chapter analyses data pertaining to patients who were admitted to an Australian or NZ ICU between 2011 and 2015 inclusive. In February 2018, these were the most recent five years of the available data which extended back to the year 2000. A variety of issues were considered when selecting this study period. Firstly, the use of a shorter study period keeps the application topical and relevant. Essentially, it will be assumed that ICU effects, as indicated by hospital types, were constant over the entire study period. Such an assumption is likely to be unrealistic in longer study periods. While this assumption could also be unrealistic for the five years covered in the present application, at least one event (death) needed to be observed for each variable contained in the model. Particular attention was required for diagnostic categories, even in the full dataset, because some categories contained very few cases, deaths, or both. A model fitted to diagnostic-category data would automatically exclude all records pertaining to mortality-free categories, and hence not estimate the corresponding regression coefficients. Such a problem would be more likely with the planned bootstrap procedure (detailed in Section 2.2.7) due to the distinct possibility of no resampled deaths from a diagnostic category with very few deaths. Minimal, further consolidation of exceptionally rare diagnostic categories with the fewest observed deaths was approved by John Moran in June 2019. Each diagnostic category then contained at least 12 deaths, and it seemed highly unlikely that any model-fitting problems would occur in this application. The most frequently used guideline in the statistical literature for a minimum number of events per variable (EPV) is ten (van Smeden et al., 2016).

The resulting dataset contained 420,769 patient admissions at 170 ICUs from 2011 to 2015. The exclusion criteria which had yet to be applied to these data follow the criteria applied by the Solomon et al. (2014) study of the APD. Thus, the additional, excluded patients were those with: unknown hospital outcome (1,368); ICU length-of-stay (LOS) unknown (108) or at most four hours (4,240); CABG (42,332); or missing APACHE III score (1,918), diagnostic category (1,332), or ventilation status (16). Solomon et al. (2014) noted that an original APACHE III requirement (Knaus et al., 1991) excluded patients whose LOS $\leq$ 4 hours. Also noted earlier in this chapter, APACHE III scores of CABG and non-CABG patients were calculated using separate models; by considering non-CABG patients only, the study is restricted to the much larger, more general subset of patients whose scores were calculated using the same model.[2] However, the present application imposes no constraint on ICU sample sizes, because a specific interest of this application lies in the shrinkage effect on the SMRs due to smoothing observed outcomes. Such smoothing was neither used in Solomon et al. (2014), nor their preceding work in Kasza et al. (2013), and potentially problematic discrepancies were avoided by excluding patients if, during their year of admission, the ICU contained fewer than 150 complete patient records. The final dataset considered for this application contained 370,554 patient admissions at 170 ICUs.

4.2 Data description

This section will describe the data selected in terms of the variables subsequently involved in the modelling. Each of the variables can be classified generally as either a response (the outcome of interest), another patient-level, or an ICU-level variable.

[2]CABG patients also had much lower mortality rates (1.1%) than non-CABG patients (10.6%).

The decision to use these particular variables largely follows previous investigations of the APD by Kasza et al. (2013), Solomon et al. (2014), and Moran and Solomon (2014).

4.2.1 Response variable

The response variable used in this application is in-hospital mortality. This binary variable is coded as one for patients who died in hospital, and zero for patients who survived to hospital discharge. The use of shorter-term mortality measures such as in-ICU mortality would raise concerns of potential gaming strategies (Bottle and Aylin, 2016, p. 65). For example, ICUs would appear better than was actually the case if they transferred patients to hospital wards when survival seemed extremely unlikely. Although the use of longer-term mortality measures such as 30 days after hospital discharge may be preferable, Kasza et al. (2013) noted that this is not currently implementable with the APD because it lacks the necessary linkage with governmental death registers.

Of the 370,554 patients in the selected dataset, 39,107 patients (10.6%) died in hospital, while the remaining 331,447 patients (89.4%) survived to hospital discharge. For convenience, in-hospital mortality will hereinafter be referred to as mortality, with the two mortality outcomes as died and survived.

4.2.2 Other patient-level variables

This application uses seven other patient-level variables: age, APACHE III score, sex, ventilation status, ICU source, diagnostic category, and ICU admission year. Using the order in this list, each variable will be described below. The first two variables in this list (age and APACHE III score) will be treated as continuous in the modelling; all other variables will be treated as categorical.

4.2.2.1 Age

The APD Data Dictionary (ANZICS CORE, 2018) indicated that the variable age contains a patient's age at the time of admission to ICU[3], rounded down to one decimal place. The Data Dictionary advised ICUs to estimate patients' ages if they were unknown. Patients were aged from 16.0 to 110.0 years old, with the mean age of 61.0 and the standard deviation (SD) of 18.4. The median age was 64.4, so the distribution of ages was negatively skewed. Figure 4.1 displays a histogram of patient admissions by age. The heights of the green bars indicate the total number of patients in a five-year age bin; the heights of the red bars indicate the corresponding number of mortalities observed. It was noted that the heights of the red bars generally increased relative to the corresponding green bars, indicative (not surprisingly) of raw mortality increasing with age.

4.2.2.2 APACHE III score

The APACHE III scoring system was briefly introduced earlier in this chapter when discussing various currently-available systems which attempt to quantify mortality

[3]Datasets containing admissions from 2016 onwards used the time of admission to hospital rather than ICU (ANZICS CORE, 2018). Almost all age differences would likely be minimal because over 95% of patients in the current dataset were admitted to ICU within nine days of hospital admission.

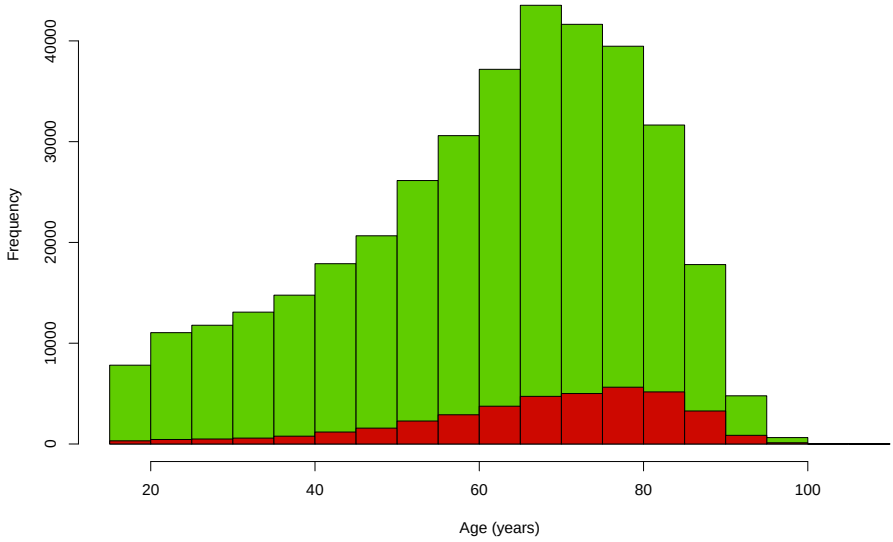


Figure 4.1: Histogram of patient admissions by age (years). The heights of the green bars indicate the total patients per five-year histogram bin; the heights of the red bars indicate the corresponding mortalities.

risks for ICU patients. Quantitatively, APACHE III scores are integer-valued, from 0 to 299, with higher values corresponding to higher predicted mortality. Knaus et al. (1991) stated that APACHE III scores take into account various patient characteristics: age contributes up to 24 units to the scores, chronic health conditions contribute up to 23 units to the scores, and 17 physiological measurements contribute up to 252 units to the scores. However, similar to other commonly used systems for quantifying illness severities (e.g. SAPS II and APACHE IV), the calculation of an APACHE III score does not take into account the patient's sex (this variable is described in Section 4.2.2.3).

For this dataset, the patients' APACHE III scores ranged between 0 and 222 inclusive, with the mean score of 56.6 and the SD of 27.8. The median score was 51, so the distribution of scores was positively skewed. Figure 4.2 displays a histogram of patient admissions by APACHE III scores. The heights of the green bars indicate the total number of patients in each score bin (with a width of ten); the heights of the red bars indicate the corresponding number of mortalities observed. Not surprisingly, the heights of the red bars increased relative to the corresponding green bars. As Solomon et al. (2014) noted, with reference to their preceding work, the APACHE III score was the most important predictor of mortality for the APD. The present dataset contained a weak, positive correlation of 0.3078 between ages and APACHE III scores. Indeed, some positive correlation is expected due to age being one of the variables involved in APACHE III calculations.

4.2.2.3 Sex

The variable sex is used to distinguish biologically between male and female patients (ANZICS CORE, 2018). The dataset contained admissions records of substantially more males than females: 209,324 males (56.5%) and 161,230 females (43.5%). Furthermore, the raw mortality rate for males, 11.1%, was slightly higher than for females, 9.8%. A summary of the patient admissions and mortalities by sex is displayed in Table 4.1.

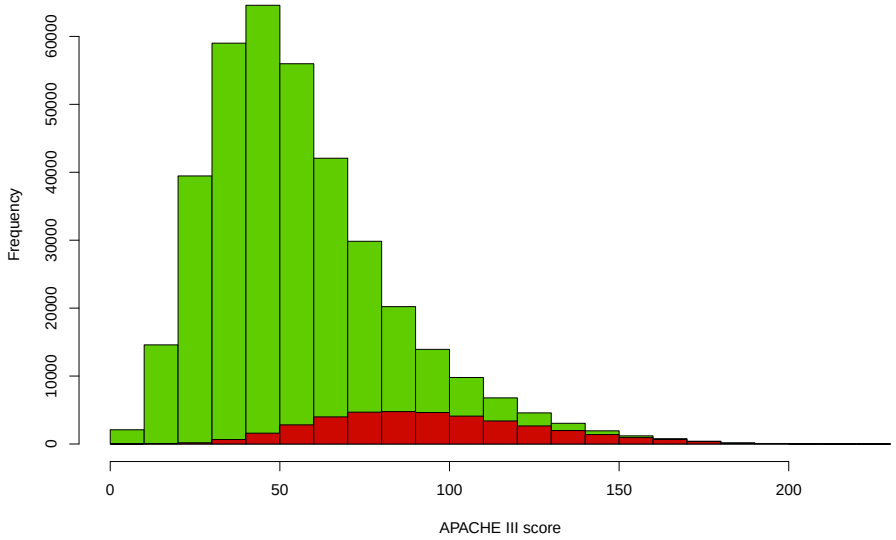


Figure 4.2: Histogram of patient admissions by APACHE III score. The heights of the green bars indicate the total patients per histogram bin with a width of ten; the heights of the red bars indicate the corresponding mortalities.

4.2.2.4 Ventilation status

Ventilation status is a binary variable indicating whether patients were on a mechanical ventilator within their first 24 hours in ICU (Moran and Solomon, 2014). Mechanical ventilation is a form of respiratory support which sustains the lives of patients who could not otherwise breathe adequately (Tobin and Manthous, 2017). For convenience, patients who received mechanical ventilation within their first 24 hours in ICU will hereinafter be referred to as having been ventilated. Nearly half of the patients, 163,479 (44.1%), were ventilated; the remaining 207,075 (55.9%) patients were non-ventilated. Not surprisingly, the raw mortality rate for ventilated patients, 16.2%, was much higher than for non-ventilated patients, 6.1%. A summary of the patient admissions and mortalities by ventilation status is displayed in Table 4.1.

4.2.2.5 ICU source

The variable ICU source was constructed consistently with previous investigations by Solomon et al. (2014), Kasza et al. (2013), and Moran and Solomon (2012). ICU source is coded as a binary variable, assigned the value one for patients who had been transferred from another hospital, and assigned the value zero otherwise. There were relatively few hospital transfers, 30,946 (8.4%), in comparison with non-transfers, 339,608 (91.6%). However, the raw mortality rate for hospital transfers, 15.0%, was considerably higher than for non-transfers, 10.2%. A summary of the patient admissions and mortalities by ICU source is displayed in Table 4.1.

It is important to note, however, that a similar limitation highlighted by Kasza et al. (2013) affects the present study. There were 30,946 hospital transfers, with 4,138 of them being directly between ICUs. The APD provides no linkage for patient records between hospitals, so there is the potential for up to 1.1% of the total patients being evaluated twice. Nevertheless, Kasza et al. (2013) also noted that there will be little impact on the results for particular ICUs when the number of patients is small, as is the case here.

Table 4.1: Admissions and mortalities (with percentages of admissions) by the patient-level, binary variables: sex, ventilation status, and ICU source. Overall, there were 370,554 admissions, with 39,107 (10.6%) mortalities.

	Admissions (%)	**Mortalities (%)**
Sex		
Male	209,324 (56.5)	23,271 (11.1)
Female	161,230 (43.5)	15,836 (9.8)
Ventilation status		
Not ventilated	207,075 (55.9)	12,576 (6.1)
Ventilated	163,479 (44.1)	26,531 (16.2)
ICU source		
No transfer	339,608 (91.6)	34,476 (10.2)
Hospital transfer	30,946 (8.4)	4,631 (15.0)

4.2.2.6 Diagnostic category

Section 4.1 briefly discussed the diagnostic categories that will be used in this application. This dataset contained 31 such categories, constructed by John Moran, to consolidate the diagnostic categories of the APACHE III algorithm (Moran and Solomon, 2014). These categories were separated by patient surgical statuses: whether patients had been admitted to ICU following surgery, either elective or emergency, or for non-surgical treatment. Table 4.2 sets out the patient admissions and mortalities by the variable diagnostic category.[4]

Initial examination of Table 4.2 will be limited to the overall, surgical statuses. Almost half of the admissions were non-surgical, 175,959 (47.5%), with the next-most admissions being from elective surgery, 135,727 (36.6%), then emergency surgery, 58,868 (15.9%). The raw mortality rate was highest for non-surgical patients, at 17.5%, with the next-highest mortality rate being for emergency-surgical patients, at 10.1%, followed by elective-surgical patients, at 1.8%.

Table 4.2 reveals that, within surgical statuses, there were very few patients admitted from the haematologic emergency (70) or elective (79) surgery categories. Furthermore, there were extremely few mortalities in the metabolic emergency-surgical category (five) and the haematologic elective-surgical category (one). Section 4.1 discussed the need for minimal, additional consolidation due to potential difficulties arising from using a bootstrap procedure in the presence of rare cases, deaths, or both. To address this problem, the haematologic and metabolic categories were consolidated within a particular surgical type for each of the emergency and elective groups. The further-consolidated categories became haematologic/metabolic emergency surgical and haematologic/metabolic elective surgical. Thus, the number of diagnostic categories used in the application was reduced to 29, with each category having at least 518 admissions (haematologic/metabolic emergency) and 12 deaths (haematologic/metabolic elective). No difficulties were encountered during the subsequent implementation of the bootstrap.

[4]In this chapter, percentages are rounded to one decimal place. As a result, some percentages do not add up to the exact values expected.

Table 4.2: Admissions and mortalities (with percentages of admissions) by diagnostic category. The haematologic and metabolic categories were consolidated within a particular surgical type for each of the emergency and elective groups. Thus, the further-consolidated categories became haematologic/metabolic emergency surgery and haematologic/metabolic elective surgery.

Diagnostic category	**Admissions (%)**		**Mortalities (%)**	
Non-surgical	**175,959**	**(47.5)**	**30,720**	**(17.5)**
Cardiovascular	31,460	(8.5)	8,956	(28.5)
Respiratory	39,584	(10.7)	6,399	(16.2)
Liver/Gastrointestinal	11,292	(3.0)	2,199	(19.5)
Central Nervous System	21,511	(5.8)	4,720	(21.9)
Sepsis	24,624	(6.6)	5,060	(20.5)
Trauma	15,135	(4.1)	1,464	(9.7)
Metabolic Hormonal	22,226	(6.0)	650	(2.9)
Haematologic	1,558	(0.4)	417	(26.8)
Renal/Genitourinary	5,869	(1.6)	669	(11.4)
Other medical disorders	1,819	(0.5)	107	(5.9)
Musculoskeletal/Skin	881	(0.2)	79	(9.0)
Emergency surgical	**58,868**	**(15.9)**	**5,946**	**(10.1)**
Cardiovascular	7,297	(2.0)	837	(11.5)
Thoracic	4,124	(1.1)	208	(5.0)
Gastrointestinal	22,534	(6.1)	2,567	(11.4)
Central Nervous System	6,803	(1.8)	1,140	(16.8)
Traumatic/Orthopaedic	6,945	(1.9)	713	(10.3)
Renal/Genitourinary	2,317	(0.6)	81	(3.5)
Gynaecological	2,787	(0.8)	17	(0.6)
Musculoskeletal/Skin	5,543	(1.5)	368	(6.6)
Haematologic	70	(0.0)	10	(14.3)
Metabolic	448	(0.1)	5	(1.1)
Elective surgical	**135,727**	**(36.6)**	**2,441**	**(1.8)**
Cardiovascular	44,663	(12.1)	943	(2.1)
Thoracic	12,810	(3.5)	210	(1.6)
Gastrointestinal	27,285	(7.4)	689	(2.5)
Central Nervous System	25,052	(6.8)	272	(1.1)
Traumatic/Orthopaedic	874	(0.2)	43	(4.9)
Renal/Genitourinary	5,870	(1.6)	58	(1.0)
Gynaecological	3,212	(0.9)	17	(0.5)
Musculoskeletal/Skin	13,588	(3.7)	197	(1.4)
Haematologic	79	(0.0)	1	(1.3)
Metabolic	2,294	(0.6)	11	(0.5)

4.2.2.7 ICU admission year

The ICU admission year variable indicates the calendar year, between 2011 and 2015 inclusive, in which a given patient was admitted to an ICU. Table 4.3 sets out the number of patients admitted per year, along with the corresponding number of mortalities observed. The total number of admissions generally increased over time (excluding 2013), while the raw mortality consistently decreased. Possible factors explaining the increase in admissions over time include population growth during this period (ABS, 2019) and the patients in this dataset representing an increasing number of ICUs reporting data to ANZICS. The decrease in mortality over time may be attributed to improvements in the quality of care (as discussed in the Chapter 2 preamble), though would also be dependent on patients' actual mortality risks.

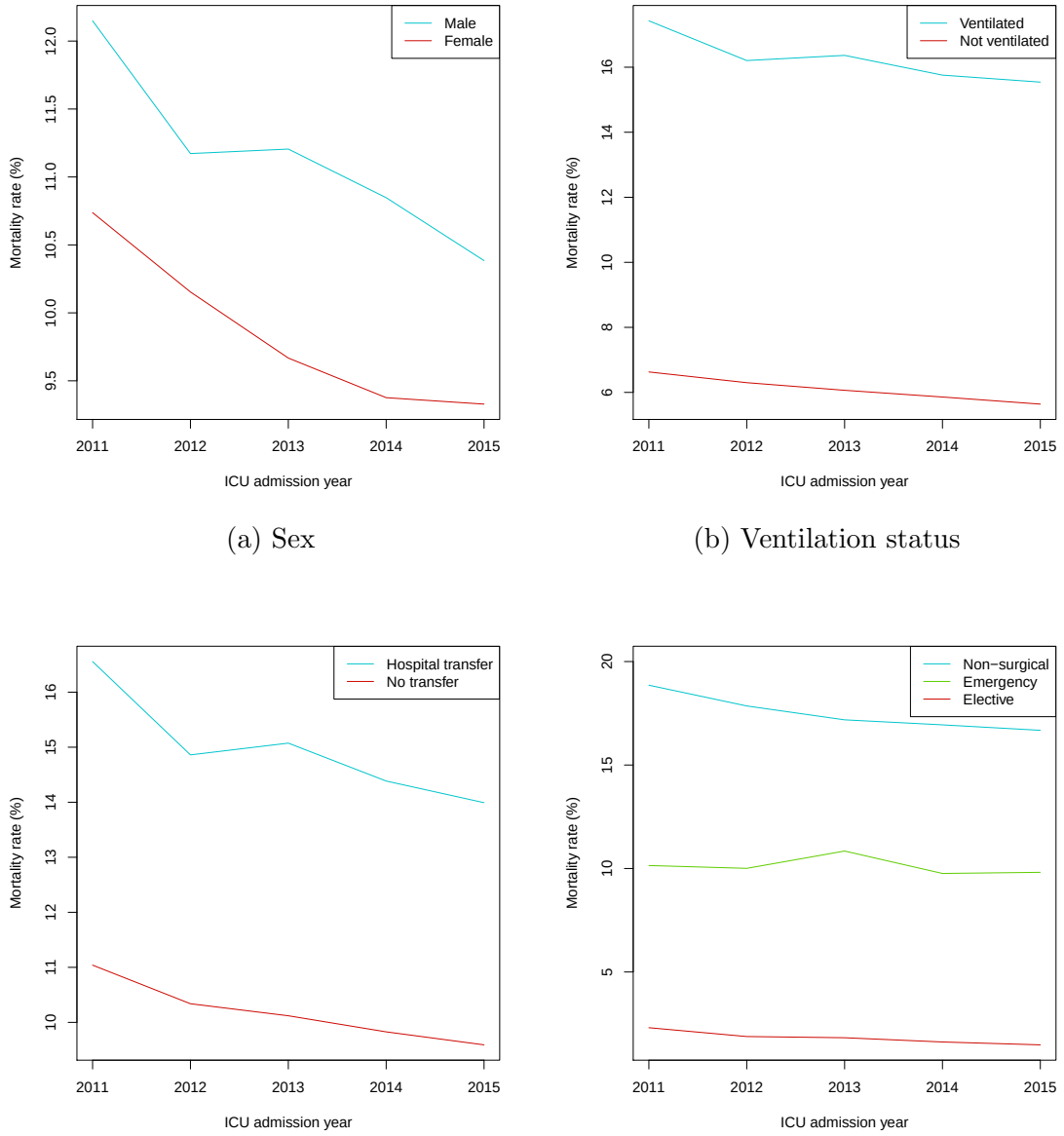
A summary of the annual admissions by patient-level, continuous and binary variables is displayed in Table 4.4. This table indicates that the mean and SD of age remained fairly constant over time. However, the mean and SD of APACHE III score consistently decreased, albeit slightly, in each subsequent year. Table 4.4 also indicates that the percentages of males and females admitted to ICU were roughly constant over time. Meanwhile, the percentage of ventilated patients usually decreased, and the percentage of hospital transfers consistently decreased. As shown in Figure 4.3, it nevertheless appears that the annual, raw mortality rates generally decreased over time, roughly uniformly between the sexes, ICU sources, and ventilation statuses; similar trends seemed apparent between non-surgical and elective-surgical statuses, but not emergency-surgical, which fluctuated comparatively.

Table 4.3: Admissions and mortalities (with percentages of admissions) by ICU admission year.

Year	**Admissions (%)**	**Mortalities (%)**
2011	67,139 (18.1)	7,745 (11.5)
2012	73,091 (19.7)	7,842 (10.7)
2013	72,536 (19.6)	7,646 (10.5)
2014	77,244 (20.8)	7,879 (10.2)
2015	80,544 (21.7)	7,995 (9.9)

Table 4.4: Annual patient admissions by patient-level, continuous and binary variables. The abbreviations used in the table header are APIII for APACHE III score and MV for mechanical ventilation (i.e. ventilated).

Year	**Age Mean (SD)**	**APIII Mean (SD)**	**Male %**	**MV %**	**Hospital transfer %**
2011	61.1 (18.5)	58.0 (28.3)	56.6	45.4	9.0
2012	61.1 (18.5)	57.2 (28.1)	56.5	44.7	8.6
2013	61.0 (18.5)	56.4 (27.8)	56.8	43.5	8.5
2014	60.9 (18.3)	56.0 (27.5)	56.0	43.9	8.2
2015	61.0 (18.3)	55.7 (27.3)	56.6	43.3	7.6



(a) Sex

(b) Ventilation status

(c) ICU source

(d) Surgical status

Figure 4.3: Annual mortality rates by sex, ventilation status, ICU source, and surgical status. Note that the y-axes differ between plots.

4.2.3 ICU-level variables

The dataset included a numerical (dummy ID) variable which demarks different ICUs, but does not actually identify them. Neither individual patients nor their treating ICUs are identifiable from the APD. Complete patient admissions records (sample sizes) in the 170 ICUs ranged from 3 to 12,226, with the mean and median of 2,179.7 and 1,492, respectively. Subject to the exclusion criteria of Section 4.1, this application retains all ICUs, even those with the smallest sample sizes. The decision to retain these ICUs for subsequent modelling is consistent with the approach recommended by Ash et al. (2012), as discussed in Section 2.1.3.

The other ICU-level variable that will be used in this application is hospital type. This is a categorical variable which describes the type of hospital in which the ICU is located: tertiary/teaching, metropolitan, rural/regional, and private. The last type on that list, private, refers to privately-funded hospitals; the other three types refer to publicly-funded hospitals. ICUs assign themselves to one of these categories (ANZICS CORE, 2018). Hereinafter, for convenience, tertiary/teaching will be referred to as tertiary, and rural/regional will be referred to as rural. Of the 170 ICUs in the dataset, there were 38 tertiary, 35 metropolitan, 41 rural, and 56 private.

Table 4.5 sets out the patient admissions and mortalities by the hospital type. Almost half of the admissions were in tertiary ICUs, 166,416 (44.9%), with the next-most admissions being to private ICUs, 104,486 (28.2%), followed by metropolitan, 62,953 (17.0%), then rural, 36,699 (9.9)%. Raw mortality rates between the public hospital categories (i.e., tertiary, metropolitan, and rural) were similar, from 13.1% to 13.4%, but much lower in private hospitals, at 3.6%. A summary of the admissions to each hospital type by patient-level, continuous and binary variables is displayed in Table 4.6. On average, patients in private ICUs were slightly older than patients in public ICUs, but had much lower mortality risks based on APACHE III scores; the distributions of ages and APACHE III scores were also narrower (smaller corresponding SDs) for private ICUs. The ICUs within each hospital type treated more male than female patients, with the percentage of males being somewhat higher in tertiary than non-tertiary ICUs. The percentage of ventilated patients was also much higher in tertiary ICUs, followed by quite similar percentages in the other public (i.e. metropolitan and rural) ICU categories, then a much lower percentage in private ICUs. Meanwhile, tertiary and metropolitan ICUs had the highest, and almost equal, percentages of hospital-transferred patients; private ICUs had the lowest percentage of hospital transfers. Furthermore, examination of admissions and mortalities within each hospital type revealed that the annual number of ICU admissions generally increased (Figure 4.4a), while the corresponding (raw) mortality rates generally decreased (Figure 4.4b).

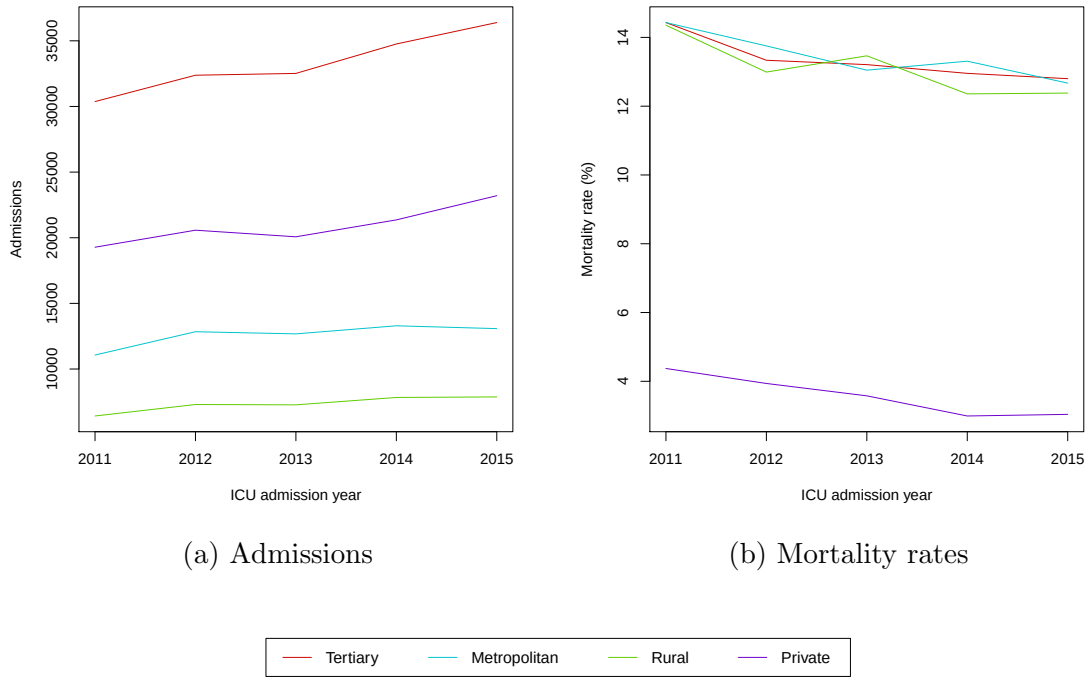
This application limited the number of ICU-level variables to the one that was directly supplied by the APD. It was noted that numerous earlier versions of the APD also contained the geographic locality variable, which indicated each ICU's location by the state or territory in Australia or the whole of NZ. Indeed, patient mortality can be significantly affected by geographic location due to climatic effects, as was detected by Solomon et al. (2014). Nevertheless, the inclusion of locality in the APD can be problematic due to the potential risk of inadvertently identifying individual ICUs. Another ICU-level variable that can still be constructed from the APD is patient volume. Within each ICU, annual volumes of patients were computed directly from the data that John Moran provided, but before I applied any exclusion

Table 4.5: Admissions and mortalities (with percentages of admissions) by hospital type.

Hospital type	**Admissions (%)**	**Mortalities (%)**
Tertiary	166,416 (44.9)	22,157 (13.3)
Metropolitan	62,953 (17.0)	8,443 (13.4)
Rural	36,699 (9.9)	4,792 (13.1)
Private	104,486 (28.2)	3,715 (3.6)

Table 4.6: Admissions to each hospital type by patient-level, continuous and binary variables. The abbreviations used in the table header are APIII for APACHE III score and MV for mechanical ventilation (i.e. ventilated).

Hospital type	**Age Mean (SD)**	**APIII Mean (SD)**	**Male %**	**MV %**	**Hospital transfer %**
Tertiary	58.2 (18.7)	60.6 (28.7)	59.8	60.0	11.1
Metropolitan	60.6 (19.3)	62.7 (30.0)	53.7	41.7	11.4
Rural	60.8 (18.8)	60.1 (30.2)	55.8	39.3	6.5
Private	65.8 (16.1)	45.4 (19.6)	53.1	21.9	2.8



(a) Admissions

(b) Mortality rates

Figure 4.4: Admissions and mortality rates by ICU admission year and hospital type.

criteria in this application. However, this thesis will give little further attention to annual ICU volume (treated as a continuous variable) because its inclusion in the modelling yielded no significant improvement. Although an inverse volume-outcome relationship was detected, it lacked significance ($P > 0.2$) and had negligible magnitude.

4.3 Model selection

This section will explain the process that was followed in selecting a model for comparing ICUs by their SMRs in the current application.

4.3.1 Methods

The aim of the methods described below was to construct an empirical model that is fit for the purpose of this application. Each candidate model was fitted to every observation in the final dataset (i.e., no separation into training and test sets), and had the same general structure as the augmented model in Section 2.2.2. Briefly, these were logistic models that contain: two levels, representing patients within ICUs; variables for patient and ICU characteristics; and a random intercept for ICUs. All models were fitted in Stata using the `melogit` command and (the default) seven-point Gaussian quadrature (StataCorp, 2015). Stata subsequently provides facilities for calculating each predicted mortality probability by plugging the mean (the default) or mode of the posterior distribution of random effects into the probability expression. However, it is preferable to calculate each probability by integrating the expression over the posterior distribution (Rabe-Hesketh and Skrondal, 2012, p. 551), essentially yielding smoothed outcomes as discussed in Chapter 2. Each probability could be readily obtained via Stata in this manner if the model had been fitted using the alternative, user-written `gllamm` command (Rabe-Hesketh and Skrondal, 2012, p. 527). Nevertheless, the present task of fitting rather large random-effects models to huge datasets proved prohibitively slow with `gllamm`, while `melogit` remained feasible.

After fitting each model using `melogit`, the probabilities were evaluated using the Laplace-IV approximation, an integration technique developed in Chapter 3.[5] Two implementations of the Laplace-IV approximation were required for evaluating each mortality probability. One Laplace-IV approximation corresponds to each ICU, that is, to evaluate the normalising constant of the posterior distribution for each ICU's random effect. The other Laplace-IV approximation corresponds to each patient, that is, to integrate their mortality probability over the posterior distribution of their ICU's random effect. Self-written functions were implemented in R for this purpose. In the present application, it was noted that the predicted probabilities evaluated by Laplace-IV generally differed minimally from the corresponding probabilities obtained directly from Stata (after using `melogit`). Well over 95% of the corresponding estimates of the probabilities differed by less than 0.0005. This would not be surprising because, as mentioned in Chapter 3, posterior distributions are generally asymptotically normal.

[5]The expression for each Laplace-related approximation uses the mode of the integrand. As was done in Chapter 3, the application in the current chapter selected the particularly stringent, absolute tolerance of 10^{-12} for approximating every mode.

In the model-development stage of this application, it seemed appropriate, for comparability, to use the same model for calculating SMRs by each method. This imposes an additional restriction on candidate models. As previously discussed in Section 2.3.1, SMR methods that involve average comparators cannot be used with any model which contains cross-level interactions (i.e., between patient- and ICU-level variables). Thus, no cross-level interactions were considered here, despite some of these effects being highly significant in the data (e.g. between ICU source and hospital type). It is noted that interactions between variables were also limited to the patient level in Moran and Solomon (2014). For the present application, the inclusion of patient-level interactions was initially guided by the corresponding interactions detailed in Kasza et al. (2013) and Moran and Solomon (2014). Each study limited modelling to the same two levels as in this application, but used earlier versions of the APD.

Model selection was also guided by discrimination, calibration, statistical significance, and information criteria. As Nattino et al. (2016) noted, the first two criteria, discrimination and calibration, are considered fundamental for logistic models to satisfy. The discrimination of a model indicates its ability to separate patients by their outcomes (Bottle and Aylin, 2016, p. 115). Discrimination was measured using the area under the receiver operating characteristic (ROC) curve with R's `pROC` package (Robin et al., 2011). Another criteria for assessing model adequacy is calibration, which indicates a model's ability to predict mortality risks across the entire distribution of risk (Bottle and Aylin, 2016, p. 116). Paul et al. (2013) demonstrated that the commonly used, Hosmer-Lemeshow test for calibration is not recommended when datasets of more than 25,000 records are involved due to the test's over-sensitivity. A viable alternative when larger datasets (as well as smaller ones) are involved is to use calibration belts (Nattino et al., 2016). These also provide graphical assistance for identifying regions of possible miscalibration, that is, where mortality was over- or under-predicted. Calibration belts were constructed using R's `givitiR` package (Nattino et al., 2017) to assess model calibration.[6]

Finally, all conclusions regarding statistical significance were determined at the 5% level ($P \leq 0.05$). Wald tests were used for deciding whether individual terms would be retained in the model. Where multiple terms were involved, the Akaike Information Criterion (AIC) was used for nested models, and the Bayesian Information Criterion (BIC) was used for non-nested models.

4.3.2 Risk adjustment

The final risk-adjustment model included each predictor variable discussed in Sections 4.2.2 and 4.2.3. Further details are provided later in this section on how each variable was represented (as numerical or categorical) in the model and which higher-order and interaction terms were included. Table C.1 displays the parameter estimates of the fitted model that was selected for the ANZICS application. The selected model only contains significant terms. There was significant variation in mortality across ICU sites; the estimated variance was 0.0678, which corresponds to an estimated standard deviation of 0.2604. Section 4.3.3 will assess whether the modelling assumption of normality seemed reasonable.

[6]These calibration belts were developed by the GiViTI (Gruppo Italiano per la Valutazione degli interventi in Terapia Intensiva), the Italian group for the evaluation of interventions in ICUs (Nattino et al., 2017).

Both age and APACHE III score were treated as numeric (continuous) variables: the values of each variable were centred (subtracting the variable's mean), then scaled by dividing the resultant values by ten.[7] Attempts to categorise the age variable lead to poorer-fitting models; examples of such categorisations included those in the original APACHE III scoring system and in Solomon et al. (2014). All other predictor variables were categorical. These variables, with their reference levels in parentheses, were sex (male), ventilation status (not ventilated), ICU source (no transfer), diagnostic category (cardiovascular non-surgical), ICU admission year (2015), and hospital type (tertiary). Representing year as a categorical variable imposed no assumption on annual trends, and improved the prediction for each of the five years in the dataset.

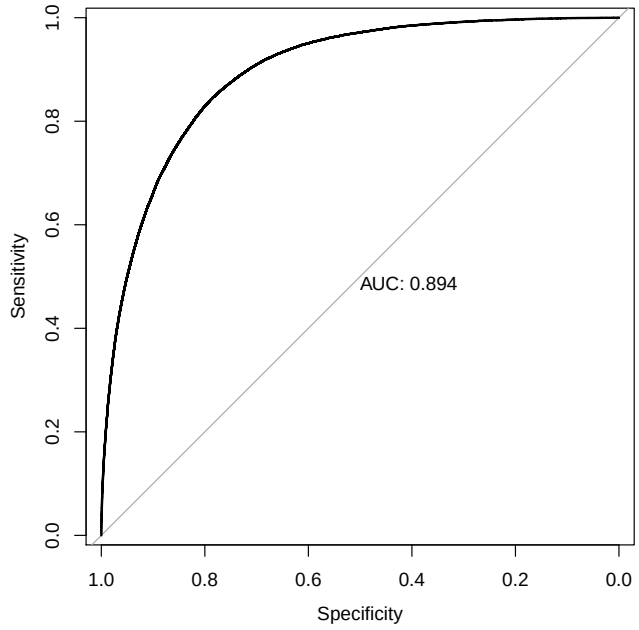
Some curvature in the effects of age and APACHE III score was accommodated in the model through the inclusion of squared terms for each variable. An interaction term between age and APACHE III score was included too. The remaining (two-way) interactions included in the model were: APACHE III score with each of ventilation status, ICU source, and diagnostic category; and ventilation status with each of age and sex.

4.3.3 Model adequacy

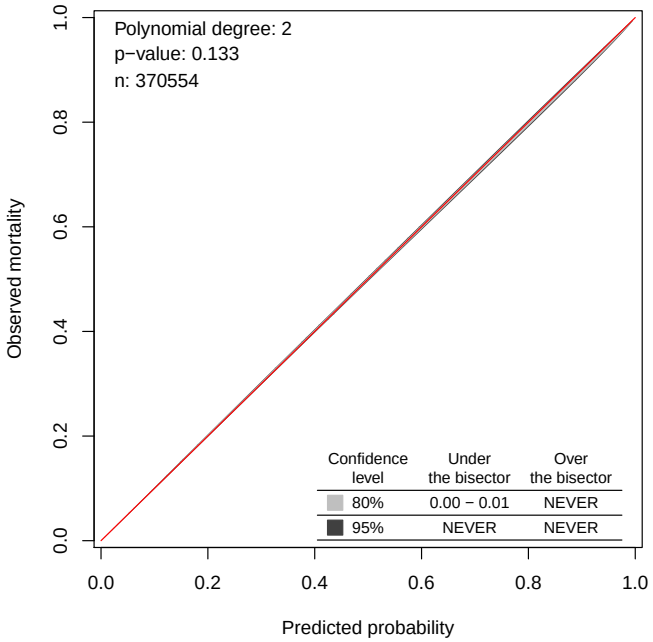
Diagnostic tests were performed to assess the adequacy of the final model selected. The ROC curve (Figure 4.5a) displays the model's sensitivity and specificity at each cut-off point for the mortality probability. In the current context, sensitivity measured the proportion of patients to be correctly classified as having died in hospital, while specificity measured the proportion of patients to be correctly classified as having survived to hospital discharge. The area under the curve (AUC) was quite high, at 0.894; the selected model had excellent discrimination according to the guidelines of Hosmer et al. (2013, p. 177).

The calibration belt plot (Figure 4.5b) displays the relationship between the model's mortality probabilities and the observed proportion of mortalities. The polynomial degree, which has the value two, refers to the degree of the polynomial in the calibration curve, where the degree was selected by maximum likelihood (Nattino et al., 2017). With the P-value of 0.133, the null hypothesis of good calibration was retained (Nattino et al., 2016). Examination of the confidence levels for the calibration belt yields the same conclusion. The bisector of perfect calibration was encompassed by the 95% calibration belt across the entire range of the model's predictions. This was also the case for the 90% calibration belt; however, that was not included on the displayed plot, because the 90% calibration belt appeared almost imperceptibly narrower than the 95% calibration belt. Incidentally, the 80% calibration belt (a default inclusion in `givitiR`) fell under the bisector for predicted probabilities below 0.01; it would follow from applying such a weak (80%) confidence level that mortality rates were significantly overestimated at very low probabilities. Of further note, the calibration belts (by confidence levels) in Figure 4.5b were very narrow, primarily due to the large sample size (n = 370,554) used and internal validation (same sample used for fitting and predicting).

[7] For this application, scaling values to reduce their magnitude resulted in `melogit` achieving dramatically faster (almost trebled) model-fitting speeds when higher-order (polynomial) terms were included. The increased speed corresponded with fewer iterations required by the numerical algorithm for fitting the models. Thus, the use of scaling may be remarkably beneficial in circumstances where it appears to be taking an inordinate amount of time to fit a model.



(a) Receiver operating characteristic (ROC) curve where the cut-off point being varied is the predicted mortality probability



(b) Calibration belt plot displaying the relationship between the predicted mortality probability and the observed proportion of mortalities

Figure 4.5: Diagnostic plots for assessing the adequacy of the selected model by its discrimination and calibration.

Both the discrimination and calibration of the selected model were deemed to be adequate for this application. Through further use of an ROC curve and a calibration belt, Appendix C.2 demonstrates how the base model, which limits the predictors to (linear) APACHE III score, achieved nearly as good discrimination (the AUC was only 0.021 lower), but unacceptably poor calibration.

Model adequacy was further assessed by its normality assumption. The posterior means of the random effects (predicted random intercepts of the ICUs) were obtained directly from Stata using the same quadrature as before. Although this was done for convenience, those predicted values could have been calculated by adapting the self-written R code (described in Section 4.3.1) because one of the functions constructed the necessary posterior distributions. Figure 4.6a displays the normal quantile-quantile (Q-Q) plot of the standardised posterior means of the random effects. The values were standardised by subtracting their approximately-zero mean of -1.9401×10^{-8} then dividing by their standard error (SE) of 0.2279. On inspection of the normal Q-Q plot, the points followed a roughly straight line, so the normality assumption seemed reasonable for this model.

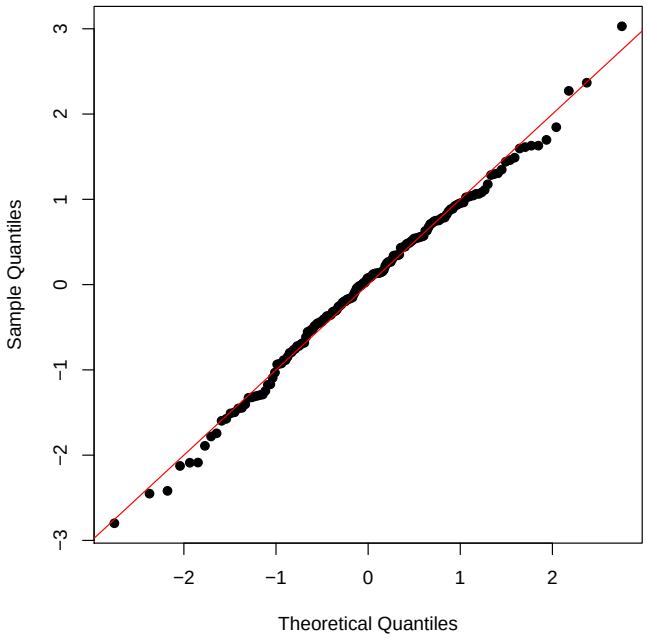
A final check of the model's adequacy involved an examination of the estimated quadratic surface that quantified the effect of age and APACHE III score towards the log odds of mortality. Figure 4.6b displays a heat map of the estimated log odds of mortality for patients across the entire ranges of age and APACHE III score contained in the dataset, conditional on them having reference-level attributes for all categorical variables. Thus, the estimates shown in the plot are restricted to non-ventilated, non-transferred, cardiovascular non-surgical, male patients who were admitted to a tertiary ICU during 2015. Since mortality generally increases with age and APACHE III score, the heat map provided plausible estimates. This conclusion appropriately avoids the fairly small areas of extrapolation beyond the data. Otherwise, the slight curvature at the bottom right of the heat map (youngest patients and highest APACHE III scores in the dataset) may be of some concern.

4.4 Comparisons of ICU performance

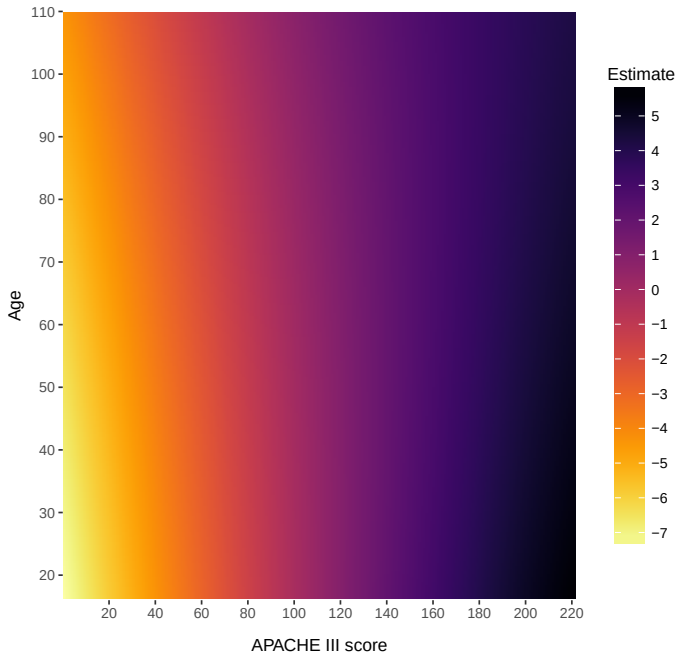
This application predominantly focuses on the performance of ICUs after adjusting for the patient casemix. Using the terminology introduced in Section 2.2.2, the ICUs were randomly numbered $i = 1, \ldots, I$, where $I = 170$. The only ICU-level variable included in the final risk-adjustment model was hospital type, and tertiary (the largest category) was set as the reference level. ICU i's characteristics were denoted by the vector $\boldsymbol{z_i} = [z_{1i}, z_{2i}, z_{3i}]^T$, with indicator variables z_1, z_2, and z_3 representing whether ICU i was metropolitan, rural, and private, respectively. Corresponding to the $\boldsymbol{z_i}$ was the vector of risk parameters $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \gamma_3]^T$. From the fitted model, $\boldsymbol{\gamma}$ was estimated by $\hat{\boldsymbol{\gamma}} = [-0.2240, -0.2639, -0.3155]^T$, and there was overwhelming evidence that each parameter in $\boldsymbol{\gamma}$ differed significantly from zero. Given a patient's risk characteristics, it was concluded that their odds of mortality would be higher if they had been treated in a tertiary ICU rather than a non-tertiary one.

4.4.1 Estimating the SMRs

With the use of the risk-adjustment model, the SMR for each ICU can be estimated by the four methods that were detailed in Section 2.2. For convenience, these SMR



(a) Normal quantile-quantile (Q-Q) plot of the standardised posterior means of the random effects (standardised by subtracting their approximately-zero mean then dividing by their standard error)



(b) Heat map displaying the estimated log odds of mortality for patients by age and APACHE III score, conditional on them having reference-level attributes for all categorical variables (i.e., non-ventilated, non-transferred, cardiovascular non-surgical, male patients who were admitted to a tertiary ICU during 2015)

Figure 4.6: Diagnostic plots for further assessing the appropriateness of the selected model by its normality assumption and plausibility.

methods will be numbered in the following multi-hierarchical order[8].

1. Standard approach, actual comparator (from Section 2.2.3)
2. Smoothed approach, actual comparator (from Section 2.2.4)
3. Standard approach, average comparator (from Section 2.2.6)
4. Smoothed approach, average comparator (from Section 2.2.5)

Trivial calculations from previous work were all that remained in evaluating the numerator of each SMR. For the standard approaches, Methods 1 and 3, the numerator is the number of (ICU-admitted) patients who died in hospital. For the smoothed approaches, Methods 2 and 4, the numerator is the sum of the smoothed outcomes, which were already evaluated in Section 4.3.1.

The denominator of each SMR is the total expected mortalities. Obtaining these values required further non-trivial calculations from previous work. The mortality probability expression for each patient was integrated over the prior distribution (normal distribution with the mean of zero and estimated variance of $\hat{\tau}^2 = 0.0678$) to obtain their expected probability of death if they were treated at an actual or average comparator ICU. Similar to before, each expected probability was evaluated using the Laplace-IV approximation. Average ICUs were those where the sum of the expected mortalities for all 370,554 patients equalled the total of 39,107 observed deaths. The necessary calculations for identifying average ICUs were detailed in Section 2.2.5 and solved by an iterative process[9]. An average ICU was estimated to be any ICU whose attributes, identified by $\boldsymbol{z^*} = [z_1^*, z_2^*, z_3^*]^T$, satisfied the constraint $\hat{\boldsymbol{\gamma}}^{\boldsymbol{T}}\boldsymbol{z^*} = -0.1285$, subject to the conditions $z_1^* + z_2^* + z_3^* \leq 1$ and $z_1^*, z_2^*, z_3^* \geq 0$. The proportions of a metropolitan, rural, and private ICU comprising a particular, average ICU were represented by z_1^*, z_2^*, and z_3^*, respectively; the corresponding proportion of a tertiary ICU was $1 - z_1^* - z_2^* - z_3^*$. An example of an average ICU was estimated as 50.6% tertiary, 19.1% metropolitan, 19.1% rural, and 11.2% private.

Using the above calculations, a point estimate of each SMR was obtained for every ICU. The uncertainty associated with each estimate was addressed by implementing the bootstrap procedure detailed in Section 2.2.7 with 1000 bootstrap samples. No difficulties were encountered with model fitting and convergence, though the entire process took just over 18 days to complete on the same computer that had previously been used for the benchmarking in Chapter 3. Nevertheless, all ICUs that observed fewer than ten mortalities during the study period were not considered in the subsequent analysis. This decision followed from the previously discussed assumption (in Section 2.2.7) that the SMRs on the logarithmic scale (LSMRs) are approximately normally distributed. For ICUs that observed few or no mortalities, many or all of the bootstrap SMRs which were calculated by a standard approach (Method 1 or 3) would be zero. This becomes particularly problematic when attempting to approximate the SEs of the LSMRs. Effectively 14 of the 170 ICUs (8.2%) were not considered further. This comprised three tertiary, one rural, and ten private ICUs, and corresponded to 1,384 of the 370,554 patients (0.4%) and 37 of the 39,107 deaths (0.1%).

[8]The methods are ordered by actual then average comparators, within which, the methods are ordered by standard then smoothed approaches.

[9]This application implemented the Method of False Position algorithm (Burden and Faires, 1993, pp.64-65) and used the (possibly excessively stringent) absolute tolerance of 10^{-12} for the difference between each total.

4.4.2 Examining the SMRs

Figure 4.7 sets out the boxplots of the estimated SMRs, by hospital type, for each method used here. Excluding outliers, the SMRs within each boxplot were distributed fairly symmetrically. The boxplots for Methods 1 and 2 (Figures 4.7a and 4.7b, respectively) were centred close to the SMR of one. Meanwhile, the boxplots for Methods 3 and 4 (Figures 4.7c and 4.7d, respectively) were centred above one for tertiary ICUs and below one for the other ICU categories. Tertiary was the reference category for the hospital-type variable, and the estimated coefficients for the non-tertiary categories were all negative. The estimated hospital-type coefficients in the fitted model affected the direction and magnitude that the boxplots shifted along the SMR-axis. Thus, the boxplots were shifted downwards most for private ICUs, followed by rural, then metropolitan ICUs; the boxplots were shifted upwards for tertiary ICUs.

Further examination of the boxplots in Figure 4.7 revealed substantial spread in the SMRs for each method. As a result, few outliers were detected in these boxplots, and all corresponded to SMRs exceeding one. The spread of the SMRs was much wider for private ICUs than each public-ICU category. However, the shrinkage effect from smoothing observed outcomes was demonstrated throughout by markedly lower variability in the SMRs. As a result, there were fewer apparent outliers for Methods 2 and 4 than Methods 1 and 3, respectively.

As previously discussed in Section 2.2, funnel plots are often used for comparing the performance of healthcare providers. In the present context, a funnel plot can be used for displaying the estimated SMR of each ICU, while also incorporating an estimate of the uncertainty associated with that SMR. Figures 4.8 and 4.9 display the funnel plots of the SMRs for the actual- and average-comparator methods, respectively. The horizontal axis of each funnel plot is the effective sample size: a measure of the variability of the estimated LSMR for an ICU relative to the total variability of the estimated LSMRs for all ICUs (Kasza et al., 2013). Effective sample sizes give smooth limits, which is why they have been used here. Calculation of the effective sample sizes follows the details contained in Kasza et al. (2013). Suppose, for convenience, that the 156 ICUs being examined were numbered $i = 1, \ldots, 156$, with the other 14 ICUs being numbered $i = 157, \ldots, 170$. The effective sample size of ICU i is

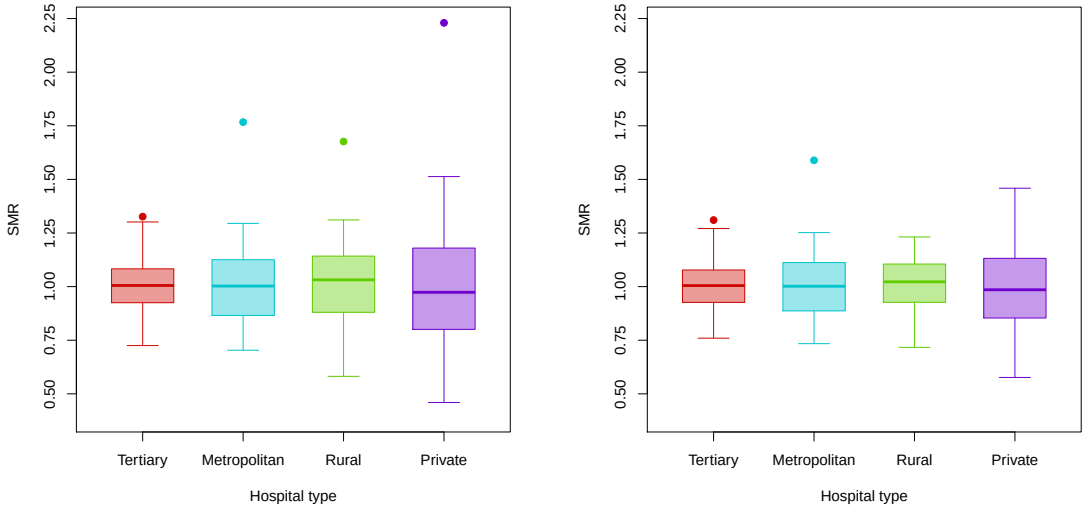
$$\widetilde{n}_i \;=\; \frac{1}{156\,\widehat{\sigma}_i^2} \sum_{l=1}^{156} n_l\,\widehat{\sigma}_l^2,$$

where

$$\widehat{\sigma}_i^2 \;=\; \left(\widehat{SE}_{boot}\left[\widehat{LSMR}_i\right]\right)^2$$

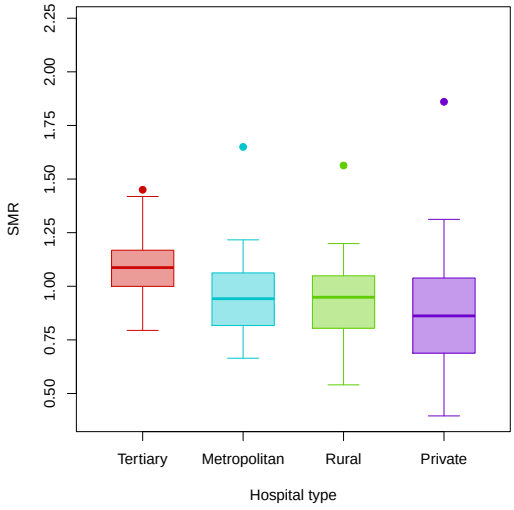
is the bootstrap estimated variance of the estimated LSMR of ICU i, as calculated in Section 2.2.7. Now the way in which effective sample sizes are defined allows for their values to be greater than or less than the corresponding actual sample sizes. The definition is such that, on average, there is no difference between weighting the variability of the estimated LSMRs by actual or effective sample sizes.

The vertical axis of each funnel plot is the un-transformed SMR, even though LSMR uncertainty was estimated. Calculating the 95% and Bonferroni limits involved back-transforming the corresponding endpoints on the logarithmic scale. The 95% limits (solid lines) control the individual error rate at 5% for each ICU, while the Bonferroni limits (broken lines) adjust for multiple comparisons by controlling the family-wise error rate at 5% for all 156 ICUs. Between these two extreme sets

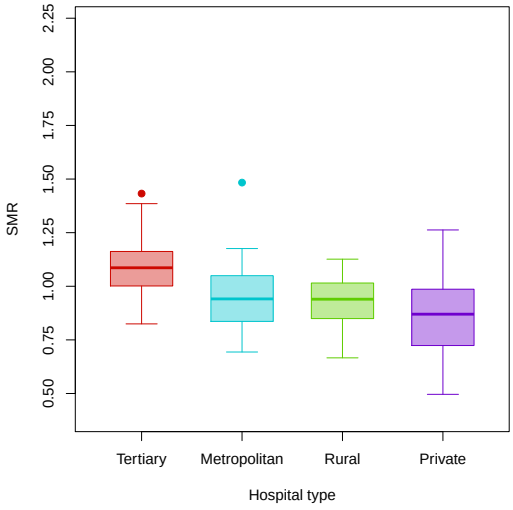


(a) Method 1

(b) Method 2

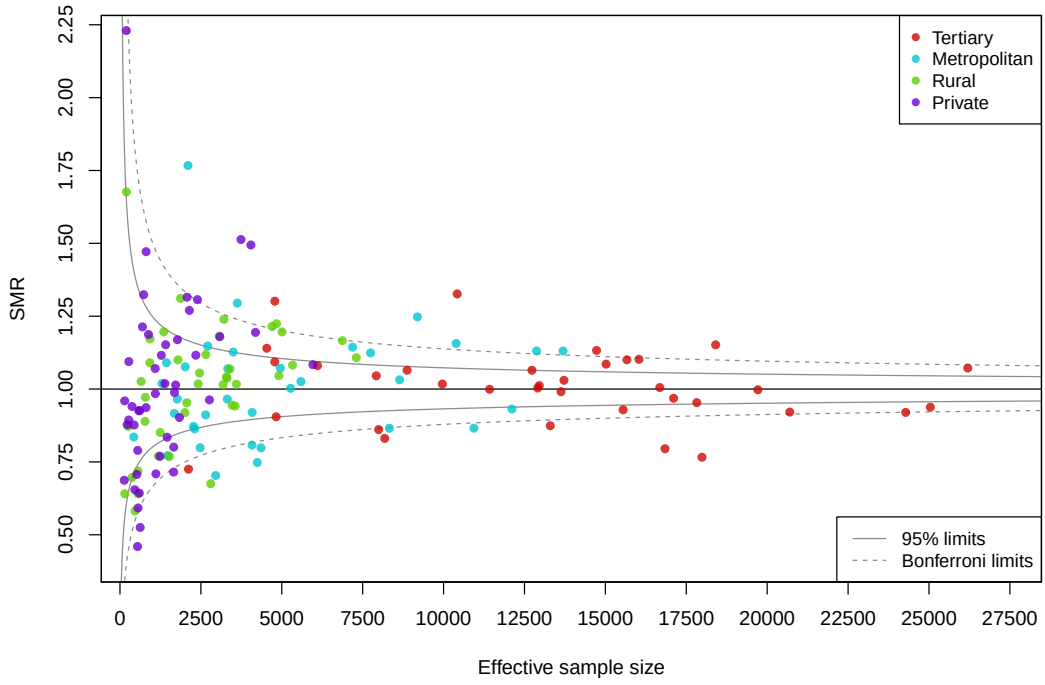


(c) Method 3

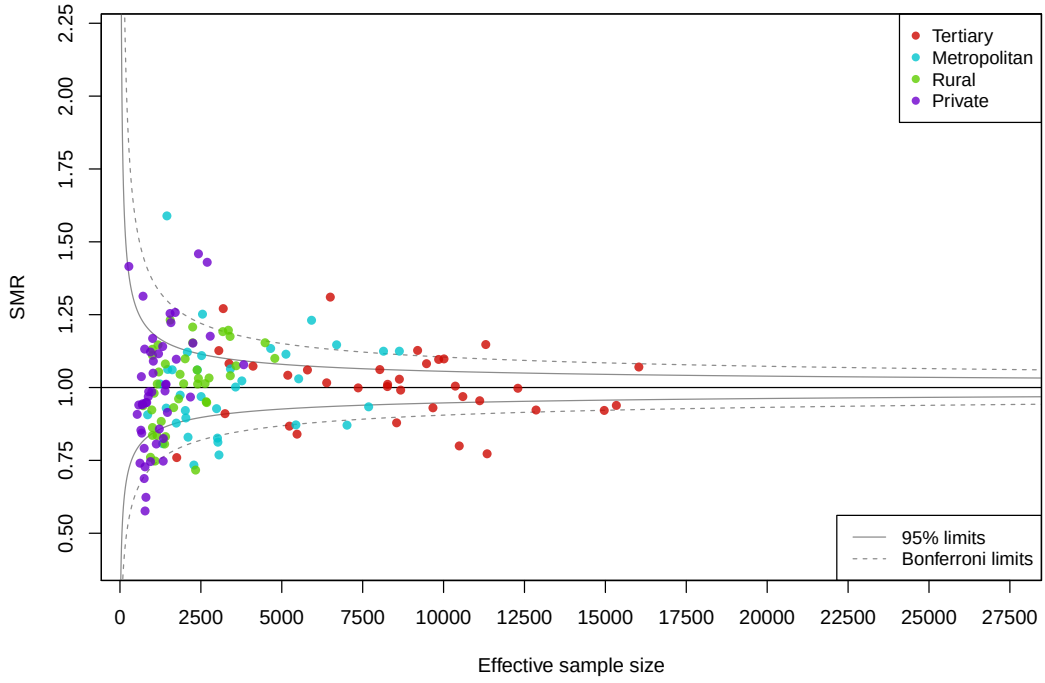


(d) Method 4

Figure 4.7: Boxplots of the estimated SMRs, by hospital type, for each method: standard approach, actual comparator (Method 1); smoothed approach, actual comparator (Method 2); standard approach, average comparator (Method 3); and smoothed approach, average comparator (Method 4).

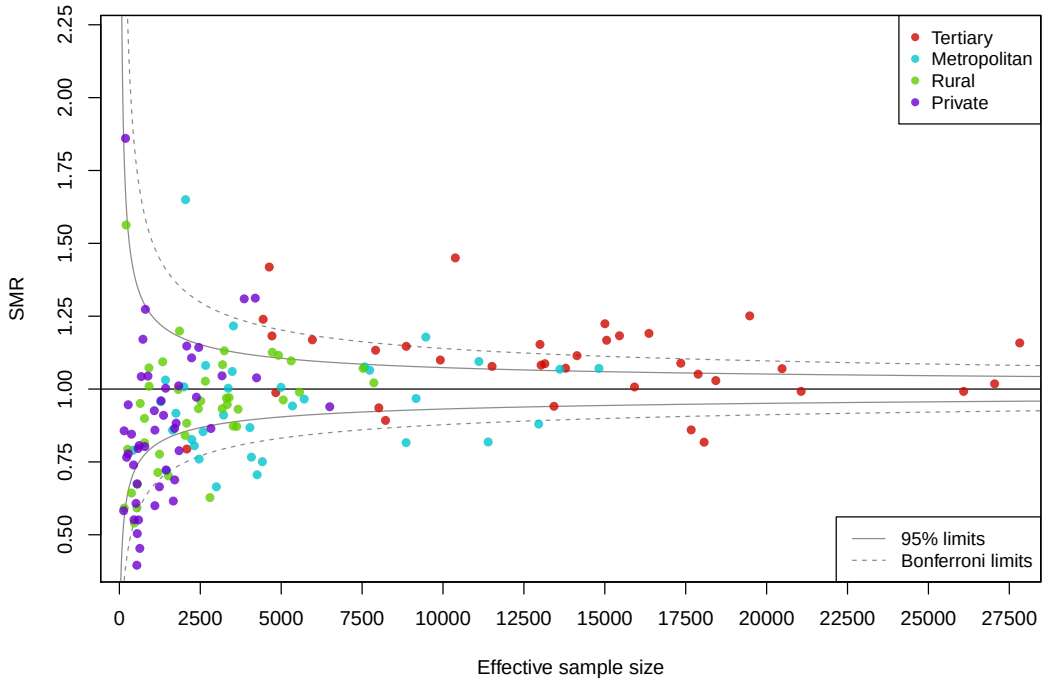


(a) Standard approach (Method 1)

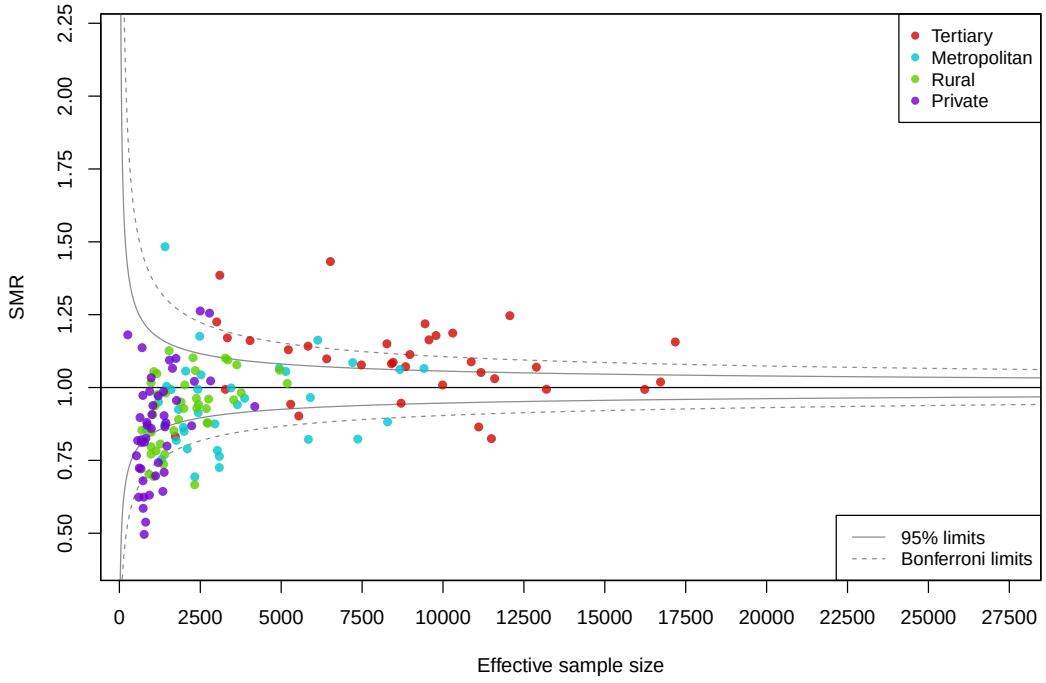


(b) Smoothed approach (Method 2)

Figure 4.8: Funnel plots of the estimated SMRs by actual comparators.



(a) Standard approach (Method 3)



(b) Smoothed approach (Method 4)

Figure 4.9: Funnel plots of the estimated SMRs by average comparators.

of limits, one could look at the false discovery rate (Benjamini and Hochberg, 1995), as was done in Kasza et al. (2013) and Solomon et al. (2014). An alternative graphical presentation to funnel plots is shown in Appendix C.1 (Figures C.1 and C.2), using the actual rather than effective sample sizes, and displaying the estimated SMRs and corresponding 95% confidence intervals (CIs) of each ICU. The use of actual sample sizes removes any intrinsic linkage between SMR variability and the horizontal axis.

Scrutinising the funnel plots (Figures 4.8 and 4.9) for each method revealed that just over half of the SMRs were within the 95% limits, and between 73.1% and 80.1% of the SMRs were within the Bonferroni limits. Table 4.7 summarises, by hospital type, the ICUs with SMRs outside of the 95% and Bonferroni limits for each SMR method. The funnels in Figures 4.8a and 4.8b were marginally narrower than the corresponding funnels in Figures 4.9a and 4.9b, respectively; however, this was not readily detectable from the plots due to the proximities of the coordinates in respective funnels. Indeed, it would seem reasonable that the variability of the SMRs was greater (albeit only narrowly here) when comparisons were to an overall average comparator rather than hospital-specific averages. As would also be expected, smoothing the observed outcomes resulted in the SMRs being shifted towards one, with the largest influences generally on the values for ICUs of smaller (actual and effective) sample sizes. Smoothing also reduced the variability of the results. This was demonstrated by the shrunken funnels of Figures 4.8b and 4.9b in comparison with the corresponding funnels of Figures 4.8a and 4.9a, respectively. A similar pattern appeared in the 95% CI widths of Figures C.1 and C.2.

The combined effect of shrinking the SMR point estimates and their variability had a relatively small but important impact on outlier detection in this application. This was confirmed by comparing the results in Tables 4.7a and 4.7c with those in Tables 4.7b and 4.7d, respectively. The number of ICUs whose SMRs fell outside of the 95% or Bonferroni limits, in total and by hospital type, sometimes differed according to whether or not smoothing was used.

Unsurprisingly, the use of actual or average comparators also influenced outlier detection. With the use of actual comparators, the number of outlying ICUs based on (individual) 95% limits was fairly uniform across all hospital types. The effects on the SMRs, when shifting from actual to average comparators, corresponded with the estimated coefficients for each hospital type in the fitted model. Thus, the SMRs of tertiary ICUs increased, with many such SMRs moved from inside the funnels to above their upper limits, and some other SMRs were moved into the funnels from below their lower limits. The opposite effect applied to non-tertiary ICUs: their SMRs decreased, with many such SMRs moved from inside the funnels to below their lower limits, and many other SMRs were moved into the funnels from above their upper limits.

4.5 Discussion

For each SMR method, there was a considerable proportion of ICUs whose SMRs differed significantly from one. There are many reasons why it would be expected for ICUs to differ in their SMRs. Benchmarking was always to an average performance, differing in the SMR methods by whether the benchmark accounted for each individual ICU's actual characteristics, in this case, hospital type for the actual comparator, or average ICU characteristics for the average comparator. As discussed in Section

Table 4.7: Summary by hospital type of the ICUs with SMRs outside of the 95% and Bonferroni limits for each SMR method. The 95% limits control the individual error rate at 5% for each ICU, while the Bonferroni limits adjust for multiple comparisons by controlling the family-wise error rate at 5% for all 156 ICUs.

(a) Method 1: standard approach, actual comparator

		95% limits		Bonferroni limits	
		Below	Above	Below	Above
Hospital type	Tertiary	10	10	7	4
	Metropolitan	10	10	6	6
	Rural	6	9	1	2
	Private	10	10	3	3
	Total	36	39	17	15

(b) Method 2: smoothed approach, actual comparator

		95% limits		Bonferroni limits	
		Below	Above	Below	Above
Hospital type	Tertiary	10	9	7	4
	Metropolitan	10	8	6	6
	Rural	7	8	1	1
	Private	10	9	4	2
	Total	37	34	18	13

(c) Method 3: standard approach, average comparator

		95% limits		Bonferroni limits	
		Below	Above	Below	Above
Hospital type	Tertiary	4	22	2	10
	Metropolitan	13	6	8	2
	Rural	11	3	3	0
	Private	14	3	8	2
	Total	42	34	21	14

(d) Method 4: smoothed approach, average comparator

		95% limits		Bonferroni limits	
		Below	Above	Below	Above
Hospital type	Tertiary	4	22	2	12
	Metropolitan	13	6	9	2
	Rural	12	0	4	0
	Private	17	2	11	2
	Total	46	30	26	16

2.2.1, the choice of comparator, actual or average, is dependent on the purpose of the investigation. In other words, why the ICUs are to be compared determines how they ought to be compared. The performance of all ICUs in the dataset reflects on the benchmark of average performance. For example, if all ICUs provided high-quality care, then the benchmark would be especially high too.

This chapter presents the first known application of SMRs calculated by a smoothed approach, where the fitted model contained provider-level attributes, despite Ash et al. (2012) suggesting its use with average comparators. On balance, smoothing, to allow for sampling variability, is recommended because the bias introduced through shrinkage is well compensated for by the reduced variability. This was most readily verifiable in the 95% CI plots (Figures C.1 and C.2) for the ICUs with relatively small sample sizes. For this application, the smoothing segment of the code was by far the most time-consuming to complete due to the computational expense involved, notwithstanding the considerable time saved by implementing a Laplace-related approximation technique. This application implemented the Laplace-IV approximation, the highest-order extension to the Laplace approximation that was articulated in Chapter 3. Of course, the application would have been completed more quickly if a lower-order extension (or indeed no extension) to the Laplace-related approximation had been used.[10] If time was exceptionally limited, one could approximate the smoothed outcomes even quicker, but less accurately, by plugging the mean or mode of the posterior distribution of the random effects into each mortality probability expression. This technique is available in Stata and was briefly discussed in Section 4.3.1. It is noted that Kasza et al. (2013) and Solomon et al. (2014) used the modal values in the above manner to estimate the expected outcomes for the denominators of their SMRs.[11]

The boxplots and funnel plots of Section 4.4.2 indicate that there was substantial variability in the SMRs within hospital types. Moreover, the SMR variability was remarkably higher for private ICUs than for each public subgroup of ICUs. Now based on the funnel plots, exactly 5% of the points would be expected to lie outside of the 95% limits simply by chance alone, that is, if the true mean SMR of each ICU equalled one. Insufficient risk adjustment would partially explain the over-dispersion, whereby far more ICUs lay outside of the funnel limits (Spiegelhalter, 2005). Additional measures for managing over-dispersion may entail expanding the funnel limits by reducing the influence of outliers, then calculating an additive or multiplicative factor to apply to each estimated variance (Spiegelhalter, 2005). It is noted that funnel limits can be expanded by adjusting for over-dispersion and/or multiple comparisons because each type of adjustment applies to different components of the funnel calculations. Examples of applications that adjusted for over-dispersion include those by two United Kingdom (UK) based entities, the National Health Service (NHS) and the Intensive Care National Audit and Research Centre (ICNARC), on their respective websites (NHS Digital, 2019; ICNARC, 2019). Both the NHS and ICNARC used multiplicative factors; those applied by the NHS were generally much larger than those applied by the ICNARC. Returning to the current application in this chapter, adjustments for over-dispersion were not applied, while adjustments for multiple comparisons were illustrated by the Bonferroni limits in Figures 4.8 and 4.9.

[10] For the illustrative example of Section 3.3.2, the choice of Laplace-related approximation had little effect on the accuracy achieved, though noticeably affected the evaluation speed.

[11] At the second stage of the investigations by Kasza et al. (2013) and Solomon et al. (2014).

Table 4.8: Approximate 95% funnel limits at various sample sizes for a parametric (Poisson) method, Method 1, and Method 3. For simpler comparisons between methods, equality was assumed between the actual and effective sample sizes.

		95% lower limit			95% upper limit		
		Poisson method	Method 1	Method 3	Poisson method	Method 1	Method 3
Sample size	200	0.6527	0.6080	0.6051	1.5321	1.6449	1.6526
	400	0.7396	0.7034	0.7010	1.3521	1.4218	1.4265
	800	0.8079	0.7797	0.7779	1.2378	1.2825	1.2855
	1,600	0.8600	0.8387	0.8373	1.1628	1.1924	1.1944
	3,200	0.8988	0.8830	0.8820	1.1125	1.1325	1.1338
	6,400	0.9274	0.9158	0.9150	1.0783	1.0920	1.0929
	12,800	0.9481	0.9397	0.9391	1.0548	1.0642	1.0648
	25,600	0.9630	0.9570	0.9566	1.0384	1.0450	1.0454

The funnel limits in Figures 4.8a and 4.9a for Methods 1 and 3, respectively, were constructed with the use of a bootstrap procedure but involved no smoothing of observed deaths. These funnel limits will now be compared with those obtained by a simpler parametric technique. Suppose that the observed deaths, denoted O, were approximated by a Poisson distribution with the mean (and variance) of λn, where $\lambda = \frac{39107}{370554}$ was the overall study-wide mortality rate, and n was the number of patients treated. Approximating the SE of the LSMR by $1/\sqrt{O}$, as used in Breslow and Day (1987, p. 67), yields approximate $(1-\alpha)\times 100\%$ funnel limits of $\exp\left\{\pm\frac{z_{\alpha/2}}{\sqrt{\lambda n}}\right\}$, where the upper $100\times\alpha/2$ percentile of the standard normal distribution is $z_{\alpha/2}$.[12] Table 4.8 sets out the approximate 95% funnel limits at various sample sizes for this parametric (Poisson) method; the corresponding limits for Methods 1 and 3 are also displayed. For simpler comparability, it was also assumed, somewhat roughly, that the actual and effective sample sizes were equal. Table 4.8 demonstrates that the Poisson-method limits were narrower than those for Method 1, which in turn were marginally narrower than those for Method 3. Furthermore, the distances between the limits decrease, both absolutely and relatively, as the sample sizes increase.

All results of this application should be viewed cautiously due to a variety of inherent limitations and potential improvements that may follow from future work. Firstly, the model contained no cross-level interactions, a necessary restriction for the SMRs calculated using average comparators. Thus, the modelling did not accommodate for cross-level interactions that were found to be statistically significant after fitting additional models, so the use of average comparators with the selected dataset appears to be problematic. Nevertheless, the ICUs of tertiary hospitals had generally higher risk-adjusted mortality based on each of the fitted models. Take the case of an expanded model that resulted from adding the interaction between ICU source and hospital type to the selected model. All differences between the corresponding parameter estimates for the selected model and the expanded model were slight. This would be expected due to the relatively small proportion of admissions

[12]This follows from the approximate $(1-\alpha)\times 100\%$ CI for the SMR of $\exp\left\{LSMR\pm\frac{z_{\alpha/2}}{\sqrt{O}}\right\}$ because the funnel limits are calculated around the SMR of one (LSMR of zero), and therefore O is set to λn if n patients were treated.

records which involved hospital transfers. The additional interaction was statistically significant for metropolitan and private hospitals, but not for rural hospitals. Furthermore, the cross-level interaction term had just over half of the magnitude of the main effect for metropolitan hospitals and the same sign (negative), and almost the same magnitude as the main effect for private hospitals but the opposite sign (positive).

The selected model limited the longitudinal effects being adjusted for to (categorical) ICU admission years, with 2015 set as the reference year. Risk-adjusted mortality was lowest in 2015 based on the positive estimates for each coefficient in the fitted model. Furthermore, the estimated coefficients decreased in magnitude for each subsequent year, except for 2012, so risk-adjusted mortality generally decreased over time. However, mortality was only (strongly) significantly higher for 2011 than the other years. Three years appeared to be the longest study period in the medical and statistical literature where any provider's performance was assessed without reference to longitudinal patterns; see, for example, the work by the Centers for Medicare and Medicaid Services (CMS) in Normand et al. (2016). Future work could involve the inclusion of more complicated longitudinal effects, using the ICU admission dates in the APD, which could improve the model in the present application. To illustrate, previous studies of the APD by Solomon et al. (2014) and Moran and Solomon (2014) detected significant linear trends and seasonal cycles associated with the ICU mortality of respiratory-disease patients. Moreover, these effects varied across geographic locations (Solomon et al., 2014), which were details not contained in the present dataset, as has been discussed in Section 4.2.3. Indeed, a limitation of the present study is the potential impact on the results by a lurking variable for ICU locality.

The APD contains other variables which have not been included in the modelling, despite their potentially significant effects on patient mortality. Two examples of these variables relate to lifestyle (smoking) and background (race). Smoking is detrimental to health and directly attributable for many ICU admissions (McGain et al., 2018). Although there were two smoking-related variables in the dataset, smoking status and smoking intensity, data contributors were (and are) not required to report them (ANZICS CORE, 2018, Table 3). The present dataset was missing smoking-related information on 74.4% of patients. Perhaps the additional effects of smoking on mortality were adequately captured by the clinical variables used. To illustrate, APACHE III scores account for physiological variables including those pertaining to the functioning of the heart and lungs; these organs are known to be adversely affected by smoking.

Indigenous status was another variable contained in the APD that seemed to have considerable reporting issues associated with it. This was a binary variable describing whether patients identified as indigenous. As a bi-national database, the coding of indigenous was specific to the country in which the treating ICU is located, indicating patients who identified as Aboriginal or Torres Strait Islander in Australian ICUs, and Māori in NZ ICUs (ANZICS CORE, 2018). However, the present dataset indicated that 6.2% of patients identified as indigenous, 72.4% as non-indigenous, and 21.4% had unknown status. Numerous studies, including Phillips et al. (2017), have quantified the substantially lower life expectancy of indigenous than non-indigenous peoples in both Australia and NZ. The socio-economic status (SES) of indigenous people in both countries was generally lower than the SES of non-indigenous people (AIHW, 2019; NZ Ministry of Health, 2018).

Stringhini et al. (2017) conducted a meta-analysis of studies from Australia, Europe, and the USA, which concluded that low SES was associated with lower life expectancy. Risk adjustment by race or SES is nevertheless controversial. Without making such adjustments, some providers may argue that they are being unfairly penalised because some of their patients had the racial or SES characteristics associated with poorer outcomes. However, by making such adjustments, there is an implicit "soft bigotry of lower expectations" for disadvantaged patients, on the basis of race or SES, and potential to excuse any providers who gave poorer-quality care to such patients (Jha and Zaslavsky, 2014). The generally higher mortality risks for patients of disadvantaged groups would be partially accounted for through slightly higher APACHE III scores based on clinical observations. Similar to the selection of comparator, the risks to be adjusted for ought to be dictated by the purpose of the investigation as well. Adjusting for controversial risk factors may seem reasonable if doing so was solely for assessing the performance of providers that treated disadvantaged patients. Nevertheless, the exclusion of such risk factors would assist investigators to identify providers requiring additional resources and funding. On balance, it seemed appropriate to exclude race and SES from the risk-adjustment models used in assessing provider performance.

To some extent, private health insurance is a proxy for SES and influences where patients receive treatment. The considerable cost of private health insurance generally precludes lower SES patients from being treated at private hospitals. Without reference to SES, Section 4.2.3 illustrated the considerably different patient casemixes for public and private ICUs. One major benefit of being privately insured is the reduced waiting times for elective surgery. It was therefore unsurprising that the proportion of elective-surgical patients was much higher for private hospitals (75.7%) than public hospitals (27.8% tertiary, 10.8% metropolitan, 9.4% rural). Within diagnostic categories, there were also enormous differences between the distributions of patients treated in each hospital type. For example, of the 22,954 trauma/orthopaedic ICU admissions across surgical and non-surgical statuses, there were 18,229 (79.4%) tertiary, 2,433 (10.6%) rural, 1,813 (7.9%) metropolitan, and 479 (2.1%) private. Section 4.2.3 discussed the variability of patient casemixes between hospital types. The private ICUs had markedly lower raw mortality rates than those of the public ICUs; however, this would be expected due to the private-ICU patients generally having much lower illness severities as well. It also seems likely that resources would usually be less constrained for private than public hospitals. Perhaps the propensity to admit less severely ill patients to an ICU was considerably greater at private hospitals?

It is noted that a small proportion of the Australian and New Zealand ICUs do not contribute data, while some have missing periods in their records. Such ICUs were usually rural or private. This may be explained by the rural ICUs and the smaller private ICUs being less likely to contribute because of constrained data collection resources (ANZICS CORE, 2017). A small proportion of all records that are contributed to the APD can be problematic (as discussed in this chapter's preamble). In the partially-cleaned subset of the APD that John Moran supplied me with, 1.0% of the patient admissions records from 2011 to 2015 were missing a value for at least one of the relevant variables.[13] Those records were omitted, and the subsequent data

[13]This was the case irrespective of whether the subset had included patients with CABG, LOS $\leq$ 4 hours, or both. Such patients were excluded from the study for the reasons explained in Section 4.1.

analysis essentially treated all missingness as completely at random. Future work to address the missing data issue could involve the use of multiple imputation, a technique that would appear to require considerable computational time, particularly due to the large database involved (Goldstein, 2010, p. 312).

Finally, this application used both Stata and R because each has its relative advantages and disadvantages. The multilevel logistic models were able to be fitted more quickly and easily using Stata. However, the coding was usually found to be much less cumbersome in R, and this chapter's application required a considerable amount of code to be written. Future work could aim to improve the efficiency of this application by automating the entire process.

Chapter 5

Conclusion

The focus of this thesis is the performance of critical care providers in hospital intensive care units (ICUs). Given the high mortality risks of patients in critical care, mortality-related outcomes are typically used in assessing their healthcare providers.

Chapter 2 contained a review of how healthcare provider performance can be assessed using standardised mortality statistics based on general risk-adjustment models. Most of Chapter 2 was devoted to four inter-related methods of calculating standardised mortality ratios (SMRs) using a model which adjusted for the patient casemix and provider attributes. The SMR methods were obtained from the four possible combinations of whether to smooth the observed mortalities to allow for sampling variation, and whether the benchmark of expected performance should account for the actual or average provider attributes. The SMR terminology that was used in this thesis labelled SMR

- approaches as smoothed or standard, depending, respectively, on whether or not smoothing was used; and
- comparators as actual or average, depending, respectively, on whether the benchmark accounted for actual or average provider attributes.

The choice of comparator, actual or average, is dependent on the purpose of the investigation. Nevertheless, the decision to use average comparators restricts the set of candidate models by excluding those that accommodate cross-level interactions between patient and provider effects. Thus, the ability to appropriately model the data before calculating SMRs is more likely to be problematic if average rather than actual comparators must be used.

Evaluating the SMRs is complicated by the intractable integrals that they contain, which represent the mortality probabilities of the patients. A non-parametric bootstrap procedure was described in Section 2.2.7 for addressing the challenge of estimating SMR uncertainty. The computational intensity of bootstrapping, in conjunction with the need to evaluate intractable integrals, emphasises the importance of speed and accuracy when assessing provider performance. Simple (and quick) approximations may be sufficiently accurate in some cases. However, more complicated (and slower) evaluations become increasingly important when the estimated between-provider variability in patient mortality is greater. Details were provided in Chapter 3 on how the intractable integrals can be evaluated by the Laplace approximation, and how this approximation technique can be extended beyond that covered in the literature, to any desired extent. The examples in Section 3.3 demonstrated the potentially favourable speed and accuracy of Laplace-related approximations in

comparison with adaptive quadrature, a commonly available option for performing numerical integration. Further work remains to be done towards developing guidelines on the use of particular approximations according to speed and accuracy criteria. In association with such guidelines, the incorporation of multiple Laplace-related approximations may enhance the model-fitting procedures in statistical software.

Chapter 4 consisted of an application that used much of the material which was covered in the earlier chapters. This application assessed the performance of Australian and New Zealand adult ICUs using a five-year subset of the Australian and New Zealand Intensive Care Society (ANZICS) Adult Patient Database (APD) from 2011 to 2015. As the investigation was limited to adult ICUs, for convenience they are simply referred to as ICUs. Patient- and ICU-level information were used to adjust for the in-hospital mortality of ICU patients. The risk-adjustment model contained patient-level variables, for many administrative and clinical attributes, and some corresponding higher-order and interaction terms. This model also contained one (statistically significant) ICU-level variable, hospital type, a categorical variable with the possible values of tertiary, metropolitan, rural, and private. Details pertaining to another ICU-level variable, annual ICU volume, were directly inferred from the dataset; however, annual ICU volume was excluded from the modelling due to a lack of evidence that its effect on mortality was statistically significant. For comparability between SMR methods, the selected model contained no cross-level interactions, and the adequacy of the model was supported by the diagnostic tests which were conducted.

Raw mortality consistently decreased over time, while risk-adjusted mortality decreased during most of the study period. Furthermore, risk-adjusted mortality was significantly higher for the ICUs of tertiary hospitals. This was reflected in the estimated SMRs when average comparators were used. Of course, the choice of comparator (actual or average) influences the notion of how ICU performance should be assessed and thus what constitutes an outlier. Irrespective of the comparator used, there was substantial variability in the estimated SMRs. This was predominantly due to the variability within each hospital type, and was largest in the private hospital category. However, the variability of the estimated SMRs, overall and within hospital types, was markedly lower when the smoothed approach was used. For each SMR method, just over half of the ICUs were within the 95% limits and around three-quarters of the ICUs were within the Bonferroni limits. Smoothing had a twofold effect on the estimated SMRs, shifting them towards one (indicative of expected performance) and reducing their variability. This had a relatively small but important impact on detecting ICU outliers in the APD. The largest impact was generally on the ICUs of smaller sample sizes. On balance, smoothing is recommended for the process of identifying unusually performing ICUs because the bias introduced through shrinkage is well compensated for by the reduced variability. As noted in Section 4.5, this thesis presents the first known application of SMRs calculated by a smoothed approach, where the fitted model contained provider-level attributes, despite Ash et al. (2012) suggesting its use with average comparators.

The results of the application in Chapter 4 should be viewed cautiously due to a variety of inherent limitations and potential improvements that may follow from future work. By enabling comparisons between the SMR methods, significant cross-level interactions were excluded from the modelling, so the use of average comparators with the APD appears to be problematic. Nevertheless, the ICUs of tertiary hospitals also had generally higher risk-adjusted mortality based on each of the fit-

ted models that was considered, whether or not the cross-level terms were included. It would be of interest to what extent this could be due to underfunding, understaffing, or the proportion of inexperienced staff in tertiary hospitals, rather than those hospitals tending to receive the more complicated cases.

Future work could involve the search for an improved risk-adjustment model. Such a model may incorporate more detailed longitudinal information to adjust for particular trends (linear or non-linear) and cycles (e.g. annual, monthly, and weekly). However, to do this effectively would require additional information than was available in the dataset. In particular, each hospital's geographic location or region is relevant because climatic conditions may have some effect on patient mortality.

Adequate risk adjustment is vital for ensuring ICU performance is accurate and fair. The procedures that have been explored in this thesis may assist with identifying which ICUs (or healthcare providers more generally) performed significantly better or worse than expected. Where significantly unusual performance is identified, further investigations may be appropriate. Perhaps the reasons for poorer performance can be remedied and the reasons for better performance can be learned from?

Appendix A

Derivations and proofs

A variety of results that were referred to in Chapters 2 and 3 are derived or proved in this appendix.

A.1 Central moments of normal distributions

This section derives a formula for the central moments of normal distributions. Such a formula, see Forbes et al. (2011) for example, is often not given in textbooks.

The probability density function (PDF) of the normally distributed random variable ε, with mean zero and variance τ^2, is

$$f(\varepsilon) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{\varepsilon^2}{2\tau^2}\right\}.$$

With odd k, $E\left[\varepsilon^k\right] = 0$ by symmetry. Now considering even k,

$$\begin{aligned} E\left[\varepsilon^k\right] &= \int_{-\infty}^{\infty} \varepsilon^k f(\varepsilon)\, \mathrm{d}\varepsilon \\ &= \int_{-\infty}^{\infty} \frac{\varepsilon^k}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{\varepsilon^2}{2\tau^2}\right\} \mathrm{d}\varepsilon \\ &= \sqrt{\frac{\tau^2}{2\pi}} \int_{-\infty}^{\infty} \varepsilon^{k-1} \left(\frac{\varepsilon}{\tau^2} \exp\left\{-\frac{\varepsilon^2}{2\tau^2}\right\}\right) \mathrm{d}\varepsilon. \end{aligned}$$

Using integration by parts,

$$\begin{aligned} E\left[\varepsilon^k\right] &= \sqrt{\frac{\tau^2}{2\pi}} \left\{ \left[-\varepsilon^{k-1} \exp\left\{-\frac{\varepsilon^2}{2\tau^2}\right\}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} (k-1)\,\varepsilon^{k-2} \exp\left\{-\frac{\varepsilon^2}{2\tau^2}\right\} \mathrm{d}\varepsilon \right\} \\ &= \sqrt{\frac{\tau^2}{2\pi}} \left\{ 0 + (k-1) \int_{-\infty}^{\infty} \varepsilon^{k-2} \exp\left\{-\frac{\varepsilon^2}{2\tau^2}\right\} \mathrm{d}\varepsilon \right\} \\ &= (k-1)\,\tau^2 \int_{-\infty}^{\infty} \frac{\varepsilon^{k-2}}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{\varepsilon^2}{2\tau^2}\right\} \mathrm{d}\varepsilon \\ &= (k-1)\,\tau^2\, E\left[\varepsilon^{k-2}\right]. \end{aligned}$$

By recursion to $k = 0$ and noting that $E\left[\varepsilon^0\right] = 1$,

$$E\left[\varepsilon^k\right] = (k-1)\,\tau^2\,(k-3)\,\tau^2 \ldots 3\,\tau^2 . 1\,\tau^2\, E\left[\varepsilon^0\right]$$

$$
\begin{aligned}
&= (k-1)(k-3)\,\ldots\,3\,\tau^k \\
&= \frac{k!\,\tau^k}{k(k-2)(k-4)\,\ldots\,4\,.\,2} \\
&= \frac{k!\,\tau^k}{\left[2\left(\frac{k}{2}\right)\right]\left[2\left(\frac{k}{2}-1\right)\right]\left[2\left(\frac{k}{2}-2\right)\right]\,\ldots\,[2(2)]\,[2(1)]} \\
&= \frac{k!\,\tau^k}{2^{k/2}\,\left(\frac{k}{2}\right)!}.
\end{aligned}
$$

Therefore, the k-th centred moment of the normal distribution with variance τ^2 is

$$
E\left[\varepsilon^k\right] \;=\; \mu_{(k)} \;=\; \begin{cases} 0 & \text{if } k \text{ is odd} \\ \frac{k!\,\tau^k}{2^{k/2}\left(\frac{k}{2}\right)!} & \text{if } k \text{ is even} \end{cases}. \tag{A.1}
$$

A.2 Simple approximation to expected outcome has odd-functional bias

Eq. 2.5 stated the bias from approximating the expected outcome for patient j treated by provider i,

$$
E_{ij} \;=\; \sum_{k=0}^{\infty} \frac{\tau^{2k}}{2^k\,k!}\, g^{(2k)}(\eta_{ij})
$$

by

$$
g\left(\eta_{ij}\right) \;=\; \left(1+e^{-\eta_{ij}}\right)^{-1}
$$

is

$$
h\left(\eta_{ij}\right) \;=\; -\sum_{k=1}^{\infty} \frac{\tau^{2k}}{2^k k!}\, g^{(2k)}(\eta_{ij})\,.
$$

The following properties of the derivatives of symmetric (odd and even) functions will be used to prove that h is an odd function. Letting a be an odd function whose first derivative exists, then by the chain rule,

$$
\begin{aligned}
& a\left(-x\right) \;=\; -a\left(x\right) \\
\Rightarrow \quad & a'(-x)\,.\,(-1) \;=\; -a'(x) \\
\Rightarrow \quad & a'(-x) \;=\; a'(x)\,,
\end{aligned}
$$

so the first derivative of an odd function is even. Now supposing that the second derivative of a exists as well, then by the chain rule,

$$
\begin{aligned}
& a''(-x)\,.\,(-1) \;=\; a''(x) \\
\Rightarrow \quad & a''(-x) \;=\; -a''(x)\,,
\end{aligned}
$$

so the second derivative of an odd function is also odd. Thus, the derivative of an odd function is even and the derivative of an even function is odd, provided that those derivatives exist. It follows that if a is an infinitely-differentiable odd function, then every odd-numbered derivative is an even function, and every even-numbered derivative is an odd function.

Inspection of function h reveals that it contains all even-ordered derivatives of the infinitely-differentiable function g. The first two derivatives of g with respect to η_{ij} are

$$g'(\eta_{ij}) \;=\; \frac{e^{\eta_{ij}}}{(1+e^{\eta_{ij}})^2}$$

and

$$g''(\eta_{ij}) \;=\; -\frac{e^{\eta_{ij}}}{(1+e^{\eta_{ij}})^3}\left(e^{\eta_{ij}}-1\right).$$

Now the lowest even-ordered derivative is an odd function because

$$\begin{aligned} g''(-\eta_{ij}) \;&=\; -\frac{e^{-\eta_{ij}}}{(1+e^{-\eta_{ij}})^3}\left(e^{-\eta_{ij}}-1\right) \\ &=\; \frac{e^{-\eta_{ij}}}{(e^{-\eta_{ij}}+1)^3}\left(1-e^{-\eta_{ij}}\right) \\ &=\; \frac{e^{2\eta_{ij}}e^{-\eta_{ij}}}{\left[e^{\eta_{ij}}\left(e^{-\eta_{ij}}+1\right)\right]^3}e^{\eta_{ij}}\left(1-e^{-\eta_{ij}}\right) \\ &=\; \frac{e^{\eta_{ij}}}{(1+e^{\eta_{ij}})^3}\left(e^{\eta_{ij}}-1\right) \\ &=\; -g''(\eta_{ij}), \end{aligned}$$

so all even-ordered derivatives are odd functions as well. Hence, the bias, h, is an odd function because it is a linear combination of odd functions.

A.3 Proof of Eq. 3.4

This section proves Eq. 3.4,

$$E\left[\prod_{i=1}^{l} U_{k_i}\right] \;=\; \begin{cases} 0 & \text{if } K \text{ is odd} \\ \frac{K!}{\left(\prod_{i=1}^{l} k_i!\right) 2^{K/2}\left(\frac{K}{2}\right)!} & \\ \times \operatorname{vec}^T\left\{\overset{K/2}{\underset{}{\otimes}} V\right\} \operatorname{vec}\left\{H_{k_1} \otimes \ldots \otimes H_{k_l}\right\} & \text{if } K \text{ is even} \end{cases},$$

using Eq. 3.3,

$$U_{k_i} \;=\; \frac{1}{k_i!}\left[\overset{k_i-1}{\underset{}{\otimes}} \boldsymbol{W}^T\right] H_{k_i}\boldsymbol{W},$$

where $k_i \in \mathbb{Z}_{\geq 3}$, and $\boldsymbol{W} = (W_1, \ldots, W_q)^T$ has the q-dimensional multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix V. The above expressions use the "vec" operator and Kronecker product, and these can be defined as follows. The "vec" operator concatenates the columns of a matrix to form a single column (Searle, 1982, p. 332). Suppose that A is a $p \times q$ matrix with elements a_{ij} ($i = 1, \ldots, p$; $j = 1, \ldots, q$), and B is any $m \times n$ matrix. Then the Kronecker product of A and B, denoted $A \otimes B$, is

$$\begin{bmatrix} a_{11}B & \cdots & a_{1q}B \\ \vdots & \ddots & \vdots \\ a_{p1}B & \cdots & a_{pq}B \end{bmatrix},$$

a $pm \times qn$ matrix (Searle, 1982, p. 265).

Firstly, an equivalent expression to the product inside the expection of Eq. 3.4 will be derived. Denoting

$$k_{\{l\}} = \sum_{i=1}^{l} k_i,$$

it will be proven inductively that the proposition

$$\mathrm{P}(l): \qquad \prod_{i=1}^{l} U_{k_i} = \left[\prod_{i=1}^{l} \frac{1}{k_i!}\right] \left[\overset{k_{\{l\}}}{\otimes} \boldsymbol{W}\right]^T \mathrm{vec}\left\{H_{k_1} \otimes \ldots \otimes H_{k_l}\right\} \tag{A.2}$$

is true for all $l \in \mathbb{N}$. This proof will use the following properties. Let A, B, C, and D be matrices; $\boldsymbol{a}$ and $\boldsymbol{b}$ be vectors; and c be a scalar. All of the following properties hold, provided that the relevant matrix multiplications exist in properties P2, P3, P4, and P8, and the matrices are square in property P6.

$$\begin{aligned}
c &= \mathrm{tr}\,(c) && \text{(P1)}\\
(AB)^T &= B^T A^T && \text{(P2)}\\
\mathrm{tr}\,(AB) &= \mathrm{tr}\,(BA) && \text{(P3)}\\
(A \otimes B)(C \otimes D) &= AC \otimes BD && \text{(P4)}\\
(A \otimes B)^T &= A^T \otimes B^T && \text{(P5)}\\
\mathrm{tr}\,(A \otimes B) &= \mathrm{tr}\,(A)\,\mathrm{tr}\,(B) && \text{(P6)}\\
\mathrm{vec}\left(\boldsymbol{a}\boldsymbol{b}^{\boldsymbol{T}}\right) &= \boldsymbol{b} \otimes \boldsymbol{a} && \text{(P7)}\\
\left(\mathrm{vec}^T A\right)(\mathrm{vec}\,B) &= \mathrm{tr}\left(A^T B\right) && \text{(P8)}
\end{aligned}$$

Property P1 is trivial, while Magnus and Neudecker (2007, pp. 6, 11, 32, 34) state properties P2, P3, P4–6, and P7–8, respectively.

Now the simplest case of the proposition in Eq. A.2, P(1), is true because

$$\begin{aligned}
U_{k_1} &= \frac{1}{k_1!}\left[\overset{k_1-1}{\otimes} \boldsymbol{W}^T\right] H_{k_1}\boldsymbol{W}\\
&= \frac{1}{k_1!}\,\mathrm{tr}\left\{\left[\overset{k_1-1}{\otimes} \boldsymbol{W}^T\right] H_{k_1}\boldsymbol{W}\right\} && \text{(from P1)}\\
&= \frac{1}{k_1!}\,\mathrm{tr}\left\{\boldsymbol{W}\left[\overset{k_1-1}{\otimes} \boldsymbol{W}^T\right] H_{k_1}\right\} && \text{(from P3)}\\
&= \frac{1}{k_1!}\,\mathrm{tr}\left\{\boldsymbol{W}\left[\overset{k_1-1}{\otimes} \boldsymbol{W}\right]^T H_{k_1}\right\} && \text{(from P5)}\\
&= \frac{1}{k_1!}\,\mathrm{tr}\left\{\left(\left[\overset{k_1-1}{\otimes} \boldsymbol{W}\right]\boldsymbol{W}^{\boldsymbol{T}}\right)^T H_{k_1}\right\} && \text{(from P2)}\\
&= \frac{1}{k_1!}\,\mathrm{vec}^T\left\{\left[\overset{k_1-1}{\otimes} \boldsymbol{W}\right]\boldsymbol{W}^{\boldsymbol{T}}\right\}\mathrm{vec}\left\{H_{k_1}\right\} && \text{(from P8)}\\
&= \frac{1}{k_1!}\left(\boldsymbol{W} \otimes \left[\overset{k_1-1}{\otimes} \boldsymbol{W}\right]\right)^T \mathrm{vec}\left\{H_{k_1}\right\} && \text{(from P7)}\\
&= \frac{1}{k_1!}\left[\overset{k_1}{\otimes} \boldsymbol{W}\right]^T \mathrm{vec}\left\{H_{k_1}\right\}.
\end{aligned} \tag{A.3}$$

It is assumed that P(r) is true, so

$$\begin{aligned}
\prod_{i=1}^{r} U_{k_i} &= \left[\prod_{i=1}^{r} \frac{1}{k_i!}\right] \left[\overset{k_{\{r\}}}{\otimes} \boldsymbol{W}\right]^T \operatorname{vec}\{H_{k_1} \otimes \ldots \otimes H_{k_r}\} \\
&= \left[\prod_{i=1}^{r} \frac{1}{k_i!}\right] \left(\left[\overset{r}{\otimes} \boldsymbol{W}\right] \otimes \left[\overset{k_{\{r\}}-r}{\otimes} \boldsymbol{W}\right]\right)^T \operatorname{vec}\{H_{k_1} \otimes \ldots \otimes H_{k_r}\} \\
&= \left[\prod_{i=1}^{r} \frac{1}{k_i!}\right] \operatorname{vec}^T\left\{\left[\overset{k_{\{r\}}-r}{\otimes} \boldsymbol{W}\right] \left[\overset{r}{\otimes} \boldsymbol{W}\right]^T\right\} \operatorname{vec}\{H_{k_1} \otimes \ldots \otimes H_{k_r}\} \\
&\qquad \text{(from P7)} \\
&= \left[\prod_{i=1}^{r} \frac{1}{k_i!}\right] \operatorname{tr}\left\{\left(\left[\overset{k_{\{r\}}-r}{\otimes} \boldsymbol{W}\right] \left[\overset{r}{\otimes} \boldsymbol{W}\right]^T\right)^T (H_{k_1} \otimes \ldots \otimes H_{k_r})\right\} \\
&\qquad \text{(from P8)} \\
&= \left[\prod_{i=1}^{r} \frac{1}{k_i!}\right] \operatorname{tr}\left\{\left(\left[\overset{r}{\otimes} \boldsymbol{W}\right] \left[\overset{k_{\{r\}}-r}{\otimes} \boldsymbol{W}\right]^T\right) (H_{k_1} \otimes \ldots \otimes H_{k_r})\right\}. \\
&\qquad \text{(from P2)}
\end{aligned}$$

For P($r+1$),

$$\begin{aligned}
\prod_{i=1}^{r+1} U_{k_i} &= \left[\prod_{i=1}^{r} U_{k_i}\right] U_{k_{r+1}} \\
&= \left[\prod_{i=1}^{r} \frac{1}{k_i!}\right] \operatorname{tr}\left\{\left(\left[\overset{r}{\otimes} \boldsymbol{W}\right] \left[\overset{k_{\{r\}}-r}{\otimes} \boldsymbol{W}\right]^T\right) (H_{k_1} \otimes \ldots \otimes H_{k_r})\right\} U_{k_{r+1}},
\end{aligned}$$

by the assumption. Furthermore,

$$U_{k_{r+1}} = \frac{1}{k_{r+1}!} \operatorname{tr}\left\{\boldsymbol{W} \left[\overset{k_{r+1}-1}{\otimes} \boldsymbol{W}\right]^T H_{k_{r+1}}\right\}$$

from Eq. A.3. It follows that P($r+1$) is true because

$$\begin{aligned}
\prod_{i=1}^{r+1} U_{k_i} &= \left[\prod_{i=1}^{r+1} \frac{1}{k_i!}\right] \operatorname{tr}\left\{\left[\overset{r}{\otimes} \boldsymbol{W}\right] \left[\overset{k_{\{r\}}-r}{\otimes} \boldsymbol{W}\right]^T (H_{k_1} \otimes \ldots \otimes H_{k_r})\right\} \\
&\qquad \times \operatorname{tr}\left\{\boldsymbol{W} \left[\overset{k_{r+1}-1}{\otimes} \boldsymbol{W}\right]^T H_{k_{r+1}}\right\} \\
&= \left[\prod_{i=1}^{r+1} \frac{1}{k_i!}\right] \operatorname{tr}\left\{\left(\left[\overset{r}{\otimes} \boldsymbol{W}\right] \left[\overset{k_{\{r\}}-r}{\otimes} \boldsymbol{W}\right]^T (H_{k_1} \otimes \ldots \otimes H_{k_r})\right)\right. \\
&\qquad \left. \otimes \left(\boldsymbol{W} \left[\overset{k_{r+1}-1}{\otimes} \boldsymbol{W}\right]^T H_{k_{r+1}}\right)\right\} \qquad \text{(from P6)} \\
&= \left[\prod_{i=1}^{r+1} \frac{1}{k_i!}\right] \operatorname{tr}\left\{\left[\left(\left[\overset{r}{\otimes} \boldsymbol{W}\right] \left[\overset{k_{\{r\}}-r}{\otimes} \boldsymbol{W}\right]^T\right) \otimes \left(\boldsymbol{W} \left[\overset{k_{r+1}-1}{\otimes} \boldsymbol{W}\right]^T\right)\right]\right. \\
&\qquad \left. \times \left[(H_{k_1} \otimes \ldots \otimes H_{k_r}) \otimes H_{k_{r+1}}\right]\right\} \qquad \text{(from P4)}
\end{aligned}$$

$$
\begin{aligned}
&= \left[\prod_{i=1}^{r+1} \frac{1}{k_i!}\right] \operatorname{tr}\left\{\left(\left[\overset{r}{\underset{\otimes}{}} \boldsymbol{W}\right] \otimes \boldsymbol{W}\right)\left(\left[\overset{k_{\{r\}}-r}{\underset{\otimes}{}} \boldsymbol{W}\right]^T \otimes \left[\overset{k_{r+1}-1}{\underset{\otimes}{}} \boldsymbol{W}\right]^T\right)\right. \\
&\qquad \left. \times \left(H_{k_1} \otimes \ldots \otimes H_{k_r} \otimes H_{k_{r+1}}\right)\right\} \qquad \text{(from P4)} \\
&= \left[\prod_{i=1}^{r+1} \frac{1}{k_i!}\right] \operatorname{tr}\left\{\left[\overset{r+1}{\underset{\otimes}{}} \boldsymbol{W}\right]\left[\overset{k_{\{r+1\}}-(r+1)}{\underset{\otimes}{}} \boldsymbol{W}\right]^T \left(H_{k_1} \otimes \ldots \otimes H_{k_{r+1}}\right)\right\} \\
&= \left[\prod_{i=1}^{r+1} \frac{1}{k_i!}\right] \operatorname{tr}\left\{\left(\left[\overset{k_{\{r+1\}}-(r+1)}{\underset{\otimes}{}} \boldsymbol{W}\right]\left[\overset{r+1}{\underset{\otimes}{}} \boldsymbol{W}\right]^T\right)^T \left(H_{k_1} \otimes \ldots \otimes H_{k_{r+1}}\right)\right\} \\
&\qquad \text{(from P2)} \\
&= \left[\prod_{i=1}^{r+1} \frac{1}{k_i!}\right] \operatorname{vec}^T\left\{\left[\overset{k_{\{r+1\}}-(r+1)}{\underset{\otimes}{}} \boldsymbol{W}\right]\left[\overset{r+1}{\underset{\otimes}{}} \boldsymbol{W}\right]^T\right\} \operatorname{vec}\left\{H_{k_1} \otimes \ldots \otimes H_{k_{r+1}}\right\} \\
&\qquad \text{(from P8)} \\
&= \left[\prod_{i=1}^{r+1} \frac{1}{k_i!}\right] \left(\left[\overset{r+1}{\underset{\otimes}{}} \boldsymbol{W}\right] \otimes \left[\overset{k_{\{r+1\}}-(r+1)}{\underset{\otimes}{}} \boldsymbol{W}\right]\right)^T \operatorname{vec}\left\{H_{k_1} \otimes \ldots \otimes H_{k_{r+1}}\right\} \\
&\qquad \text{(from P7)} \\
&= \left[\prod_{i=1}^{r+1} \frac{1}{k_i!}\right] \left[\overset{k_{\{r+1\}}}{\underset{\otimes}{}} \boldsymbol{W}\right]^T \operatorname{vec}\left\{H_{k_1} \otimes \ldots \otimes H_{k_{r+1}}\right\}.
\end{aligned}
$$

Since P(1) is true, and P(r) implies that P($r+1$) is true, then by the principle of mathematical induction, P(l) is true for all $l \in \mathbb{N}$.

By using the above result, and denoting

$$K \;=\; k_{\{l\}} \;=\; \sum_{i=1}^{l} k_i,$$

the left-hand side of Eq. 3.4 becomes

$$
\begin{aligned}
E\left[\prod_{i=1}^{l} U_{k_i}\right] &= E\left[\left(\prod_{i=1}^{l} \frac{1}{k_i!}\right)\left(\overset{K}{\underset{\otimes}{}} \boldsymbol{W}\right)^T \operatorname{vec}\left\{H_{k_1} \otimes \ldots \otimes H_{k_l}\right\}\right] \\
&= \left(\prod_{i=1}^{l} \frac{1}{k_i!}\right)\left(E\left[\overset{K}{\underset{\otimes}{}} \boldsymbol{W}\right]\right)^T \operatorname{vec}\left\{H_{k_1} \otimes \ldots \otimes H_{k_l}\right\}. \qquad \text{(A.4)}
\end{aligned}
$$

The expectation in the right-hand side of Eq. A.4 can be expressed alternatively as

$$
\begin{aligned}
E\left[\overset{K}{\underset{\otimes}{}} \boldsymbol{W}\right] &= E\left[\boldsymbol{W} \otimes \left(\overset{K-1}{\underset{\otimes}{}} \boldsymbol{W}\right)\right] \\
&= E\left[\operatorname{vec}\left\{\left(\overset{K-1}{\underset{\otimes}{}} \boldsymbol{W}\right)\boldsymbol{W}^{\boldsymbol{T}}\right\}\right] \qquad \text{(from P7)} \\
&= \operatorname{vec}\left\{E\left[\left(\overset{K-1}{\underset{\otimes}{}} \boldsymbol{W}\right)\boldsymbol{W}^{\boldsymbol{T}}\right]\right\}.
\end{aligned}
$$

Now Raudenbush et al. (2000, Appendix) stated that

$$E\left[\left(\overset{K-1}{\underset{\otimes}{}} \boldsymbol{W}\right)\boldsymbol{W}^{\boldsymbol{T}}\right] = \mu_{(K)},$$

the K-th centred moment of the q-dimensional multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix V, so

$$E\left[\overset{K}{\underset{\otimes}{}}\boldsymbol{W}\right] \;=\; \operatorname{vec}\left\{\mu_{(K)}\right\}.$$

Furthermore, Yang in Raudenbush et al. (2000, Appendix) showed that

$$\operatorname{vec}\left\{\mu_{(K)}\right\} \;=\; \begin{cases} \mathbf{0} & \text{if } K \text{ is odd} \\ (K-1)(K-3)\,\ldots\,3\,\operatorname{vec}\left\{\overset{K/2}{\underset{\otimes}{}}V\right\} & \text{if } K \text{ is even} \end{cases},$$

where $K \in \mathbb{Z}_{\geq 3}$. Since

$$\begin{aligned} (K-1)(K-3)\,\ldots\,3 \;&=\; \frac{K!}{K(K-2)(K-4)\,\ldots\,4\,.\,2} \\ &=\; \frac{K!}{\left[2\left(\frac{K}{2}\right)\right]\left[2\left(\frac{K}{2}-1\right)\right]\left[2\left(\frac{K}{2}-2\right)\right]\,\ldots\,[2(2)]\,[2(1)]} \\ &=\; \frac{K!}{2^{K/2}\,\left(\frac{K}{2}\right)!}, \end{aligned}$$

then

$$\operatorname{vec}\left\{\mu_{(K)}\right\} \;=\; \begin{cases} \mathbf{0} & \text{if } K \text{ is odd} \\ \frac{K!}{2^{K/2}\left(\frac{K}{2}\right)!}\operatorname{vec}\left\{\overset{K/2}{\underset{\otimes}{}}V\right\} & \text{if } K \text{ is even} \end{cases}.$$

Thus, Eq. A.4 becomes

$$E\left[\prod_{i=1}^{l} U_{k_i}\right] \;=\; \begin{cases} 0 & \text{if } K \text{ is odd} \\ \frac{K!}{\left(\prod_{i=1}^{l} k_i!\right) 2^{K/2}\left(\frac{K}{2}\right)!} & \\ \times \operatorname{vec}^T\left\{\overset{K/2}{\underset{\otimes}{}}V\right\}\operatorname{vec}\left\{H_{k_1}\otimes\,\ldots\,\otimes H_{k_l}\right\} & \text{if } K \text{ is even} \end{cases},$$

which is Eq. 3.4.

Appendix B

Evaluating the integrals in R

This appendix presents the R code that was used to evaluate the integrals in Section 3.3 when comparing the results obtained by adaptive quadrature (AQ) and Laplace-related approximations. Examples are also provided to illustrate the apparent anomalies which can arise from using R's `integrate` function.

B.1 Single integrals

As referred to in Section 3.3.2.2, the following function was used to obtain AQ and Laplace-related approximations of

$$I \;=\; \int_{-\infty}^{\infty} f\left(\varepsilon\right) \mathrm{d}\varepsilon \;=\; \int_{-\infty}^{\infty} e^{h(\varepsilon)} \,\mathrm{d}\varepsilon,$$

where

$$h\left(\varepsilon\right) \;=\; -\frac{1}{2}\log\left(2\pi\tau^2\right) - \frac{\varepsilon^2}{2\tau^2} - \log\left(1+e^{-\eta-\varepsilon}\right),$$

given particular values of η (`eta`) and τ^2 (`taus`).

```
single_integral = function(eta, taus)
{
  TOL    = 1e-12  # absolute and relative tolerances for AQ-LT
  MAXSUB = 1000   # maximum number of subintervals for AQ-LT

  h = function(eps) - (1/2) * log(2 * pi * taus) -
                      (eps ^ 2) / (2 * taus) - log(1 + exp(-eta - eps))
  f = function(eps) exp(h(eps))

  # Implement the Newton-Raphson method to locate the mode of the integrand
  # using the tolerance of TOL.

  m0 = taus / 2  # initial estimate of the mode
  ITER = 1e9     # maximum iterations allowed (never actually reached)
  repeat
  {
    EXP = exp(eta + m0)
    INCEXP = 1 + EXP
    m = m0 - (- m0 / taus + 1 / INCEXP) / (- 1 / taus - EXP / (INCEXP ^ 2))
    if (abs(m - m0) < TOL) break()
    ITER = ITER - 1
    if (!ITER)
    {
```

```
    cat("Unable to find the mode - maximum iterations exceeded\n")
    return(Inf)
  }
  m0 = m
}

EXP = exp(eta + m)
EXP2 = EXP * EXP
EXP3 = EXP2 * EXP
EXP4 = EXP3 * EXP
EXP5 = EXP4 * EXP
EXP6 = EXP5 * EXP

INCEXP = 1 + EXP

v = - 1 / (- 1 / taus - EXP / (INCEXP ^ 2))
Laplace_I = sqrt(2 * pi * v) * exp(h(m))

H3 = (EXP - 1) * EXP / (INCEXP ^ 3)
H4 = (4 * EXP - EXP2 - 1) * EXP / (INCEXP ^ 4)
H5 = (11 * EXP - 11 * EXP2 + EXP3 - 1) * EXP / (INCEXP ^ 5)
H6 = (26 * EXP - 66 * EXP2 + 26 * EXP3 - EXP4 - 1) * EXP / (INCEXP ^ 6)
H7 = (57 * EXP - 302 * EXP2 + 302 * EXP3 - 57 * EXP4 + EXP5 - 1) *
     EXP / (INCEXP ^ 7)
H8 = (120 * EXP - 1191 * EXP2 + 2416 * EXP3 - 1191 * EXP4 +
      120 * EXP5 - EXP6 - 1) * EXP / (INCEXP ^ 8)

v2 = v  * v
v3 = v2 * v
v4 = v3 * v
v5 = v4 * v
v6 = v5 * v
v7 = v6 * v
v8 = v7 * v
v9 = v8 * v

a1 = (5/24) * (H3 ^ 2) * v3 + (1/8) * H4 * v2

a2 = (385/1152) * (H3 ^ 4) * v6 + (35/64) * (H3 ^ 2) * H4 * v5 +
     (7/48) * H3 * H5 * v4 + (35/384) * (H4 ^ 2) * v4 + (1/48) * H6 * v3

a3 = (85085/82944) * (H3 ^ 6) * v9 + (25025/9216) * (H3 ^ 4) * H4 * v8 +
     (5005/3072) * (H3 ^ 2) * (H4 ^ 2) * v7 + (1001/1152) * (H3 ^ 3) * H5 * v7 +
     (77/128) * H3 * H4 * H5 * v6 + (77/384) * (H3 ^ 2) * H6 * v6 +
     (385/3072) * (H4 ^ 3) * v6 + (7/128) * H4 * H6 * v5 +
     (21/640) * (H5 ^ 2) * v5 + (1/32) * H3 * H7 * v5 + (1/384) * H8 * v4

Laplace_II  = Laplace_I * (1 + a1)
Laplace_III = Laplace_I * (1 + a1 + a2)
Laplace_IV  = Laplace_I * (1 + a1 + a2 + a3)
Laplace6    = Laplace_I * (1 + a1 + (1/48) * H6 * v3)
Laplace8    = Laplace_I * (1 + a1 +
                           (7/48) * H3 * H5 * v4 + (35/384) * (H4 ^ 2) * v4 +
                           (1/48) * H6 * v3 + (1/384) * H8 * v4)

AQ_DT = integrate(f, -Inf, Inf)$value
AQ_LT = integrate(f, -Inf, Inf, subdivisions = MAXSUB,
                  rel.tol = TOL, abs.tol = TOL)$value
```

```
  c(AQ_DT, AQ_LT,
    Laplace_I, Laplace_II, Laplace_III, Laplace_IV, Laplace6, Laplace8)
}
```

The following R code was used for part of an investigation into the differences between the AQ-DT and AQ-LT estimates of I when $\tau^2 = 0.1$, 0.2, 0.5, and 1. Examination of the plots produced (Figure B.1) indicates that the differences between corresponding AQ estimates were notably poorer for a small range of values for η, and that range differed dramatically with each value of τ^2 considered. However, each difference still remained well within the requested, default (absolute and relative) tolerances of AQ-DT, 1.2207×10^{-4}.

```
PlotDifferences = function(taus)
{
  TOL    = 1e-12            # absolute and relative tolerances for AQ-LT
  MAXSUB = 1000             # maximum number of subintervals for AQ-LT
  etas = seq(-3, 0, 0.001)  # set of values to be used for eta
  differences = NULL        # stores differences between AQ-DT and AQ-LT

  for (eta in etas)
  {
    f = function(eps) exp(- (1/2) * log(2 * pi * taus)
                          - (eps ^ 2) / (2 * taus) - log(1 + exp(-eta - eps)))
    AQ_DT = integrate(f, -Inf, Inf)$value
    AQ_LT = integrate(f, -Inf, Inf, subdivisions = MAXSUB,
                      rel.tol = TOL, abs.tol = TOL)$value
    differences = c(differences, AQ_DT - AQ_LT)
  }

  plot(NA, xlab = expression(eta), ylab = "Difference",
           xlim = range(etas), ylim = range(differences),
           xaxs = "i", xaxt = "n", type = "b")
  axis(1, -3:0)
  lines(etas, differences)
}

PlotDifferences(0.1)
PlotDifferences(0.2)
PlotDifferences(0.5)
PlotDifferences(1)
```

B.2 Double integrals

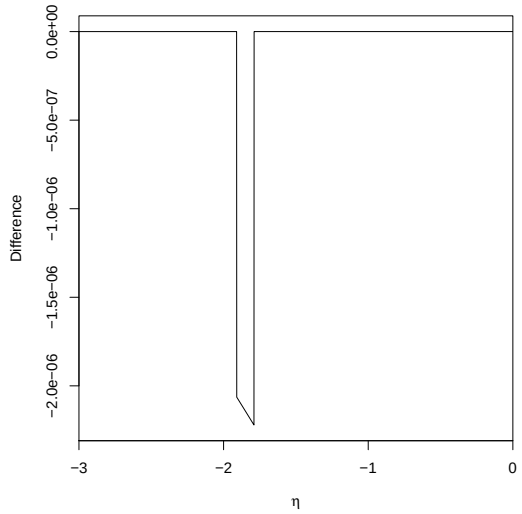
As referred to in Section 3.3.3.2, the following function was used to obtain AQ and Laplace-related approximations of

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\varepsilon_0, \varepsilon_1)\, \mathrm{d}\varepsilon_0\, \mathrm{d}\varepsilon_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{h(\varepsilon_0,\varepsilon_1)}\, \mathrm{d}\varepsilon_0\, \mathrm{d}\varepsilon_1,$$

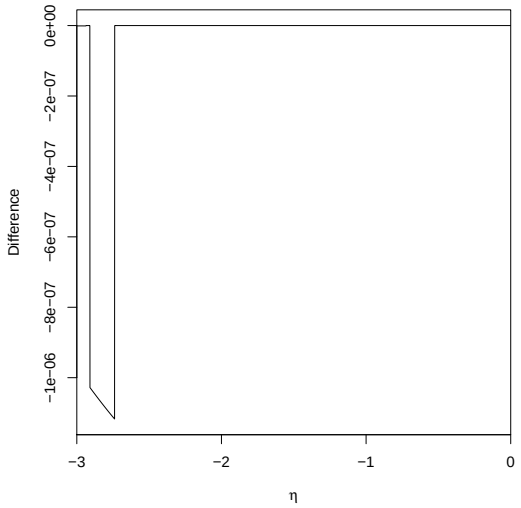
where

$$h(\varepsilon_0, \varepsilon_1) = -\log\left(2\pi\sqrt{\tau_0^2\tau_1^2 - \tau_{01}^2}\right) - \frac{\varepsilon_0^2\tau_1^2 - 2\varepsilon_0\varepsilon_1\tau_{01} + \varepsilon_1^2\tau_0^2}{2\left(\tau_0^2\tau_1^2 - \tau_{01}^2\right)} - \log\left(1 + e^{-\eta-\varepsilon_0-\varepsilon_1 x_1}\right),$$

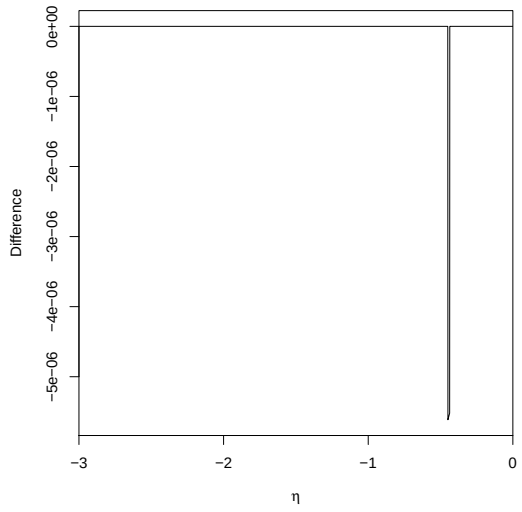
given particular values of η (`eta`), τ_0^2 (`tau0s`), τ_1^2 (`tau1s`), τ_{01} (`tau01`), and x_1 (`x1`).



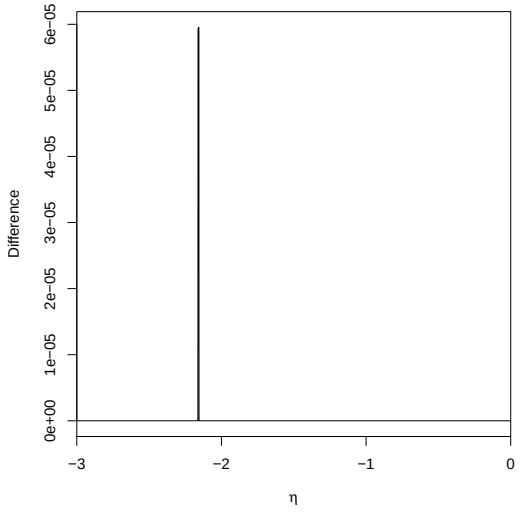
(a) $\tau^2 = 0.1$



(b) $\tau^2 = 0.2$



(c) $\tau^2 = 0.5$



(d) $\tau^2 = 1$

Figure B.1: Differences between AQ-DT and AQ-LT estimates of a single integral for a range of values of η and various values of τ^2.

```
double_integral = function(eta, tau0s, tau1s, tau01, x1)
{
  vec  = function(M)  t(t(as.vector(as.matrix(M))))
  vecT = function(M)  t(vec(M))

  TOL    = 1e-12  # absolute and relative tolerances for AQ-LT
  MAXSUB = 1000   # maximum number of subintervals for AQ-LT

  detVCV = tau0s * tau1s - tau01 ^ 2

  h = function(eps0, eps1)
      - log(2 * pi * sqrt(detVCV)) -
        ((eps0 ^ 2) * tau1s - 2 * eps0 * eps1 * tau01 + (eps1 ^ 2) * tau0s) /
        (2 * detVCV) - log(1 + exp(-eta - eps0 - eps1 * x1))

  f = function(eps0, eps1) exp(h(eps0, eps1))

  x2 = x1 ^ 2
  x3 = x1 ^ 3
  x4 = x1 ^ 4
  x5 = x1 ^ 5
  x6 = x1 ^ 6
  x7 = x1 ^ 7
  x8 = x1 ^ 8

  # Implement Newton's method for nonlinear systems to locate the mode of the
  # integrand using the tolerance of TOL.

  m0_0 = m1_0 = 0  # initial estimate of the mode
  ITER = 1e9       # maximum iterations allowed (never actually reached)

  repeat
  {
    EXP = exp(eta + m0_0 + m1_0 * x1)
    INCEXP = 1 + EXP
    INCEXP2 = INCEXP ^ 2

    H10 = - (m0_0 * tau1s - m1_0 * tau01) / detVCV + 1 / INCEXP
    H01 = - (m1_0 * tau0s - m0_0 * tau01) / detVCV + x1 / INCEXP

    H20 = - tau1s / detVCV - EXP / INCEXP2
    H11 = tau01 / detVCV - x1 * EXP / INCEXP2
    H02 = - tau0s / detVCV - x2 * EXP / INCEXP2

    detJ = H20 * H02 - H11 ^ 2
    m0 = m0_0 - (H02 * H10 - H11 * H01) / detJ
    m1 = m1_0 - (H20 * H01 - H11 * H10) / detJ

    if (sqrt( (m0 - m0_0) ^ 2 + (m1 - m1_0) ^ 2 ) < TOL) break()

    ITER = ITER - 1
    if (!ITER)
    {
      cat("Unable to find the mode - maximum iterations exceeded\n")
      return(Inf)
    }

    m0_0 = m0
    m1_0 = m1
```

```
}

EXP = exp(eta + m0 + m1 * x1)
EXP2 = EXP ^ 2
EXP3 = EXP ^ 3
EXP4 = EXP ^ 4
EXP5 = EXP ^ 5
EXP6 = EXP ^ 6

INCEXP = 1 + EXP
INCEXP2 = INCEXP ^ 2

H20 = - tau1s / detVCV - EXP / INCEXP2
H11 = tau01 / detVCV - x1 * EXP / INCEXP2
H02 = - tau0s / detVCV - x2 * EXP / INCEXP2
H2m = matrix(c(H20, H11, H11, H02), ncol = 2)

V = solve(- H2m)

H30 = (EXP - 1) * EXP / (INCEXP ^ 3)
H21 = x1 * H30
H12 = x2 * H30
H03 = x3 * H30
H3m = matrix(c(H30, H21, H21, H12, H21, H12, H12, H03), ncol = 2)

H40 = (4 * EXP - EXP2 - 1) * EXP / (INCEXP ^ 4)
H31 = x1 * H40
H22 = x2 * H40
H13 = x3 * H40
H04 = x4 * H40
H4m = matrix(c(H40, H31, H31, H22, H31, H22, H22, H13, H31, H22, H22, H13, H22,
               H13, H13, H04), ncol = 2)

H50 = (11 * EXP - 11 * EXP2 + EXP3 - 1) * EXP / (INCEXP ^ 5)
H41 = x1 * H50
H32 = x2 * H50
H23 = x3 * H50
H14 = x4 * H50
H05 = x5 * H50
H5m = matrix(c(H50, H41, H41, H32, H41, H32, H32, H23, H41, H32, H32, H23, H32,
               H23, H23, H14, H41, H32, H32, H23, H32, H23, H23, H14, H32, H23,
               H23, H14, H23, H14, H14, H05), ncol = 2)

H60 = (26 * EXP - 66 * EXP2 + 26 * EXP3 - EXP4 - 1) * EXP / (INCEXP ^ 6)
H51 = x1 * H60
H42 = x2 * H60
H33 = x3 * H60
H24 = x4 * H60
H15 = x5 * H60
H06 = x6 * H60
H6m = matrix(c(H60, H51, H51, H42, H51, H42, H42, H33, H51, H42, H42, H33, H42,
               H33, H33, H24, H51, H42, H42, H33, H42, H33, H33, H24, H42, H33,
               H33, H24, H33, H24, H24, H15, H51, H42, H42, H33, H42, H33, H33,
               H24, H42, H33, H33, H24, H33, H24, H24, H15, H42, H33, H33, H24,
               H33, H24, H24, H15, H33, H24, H24, H15, H24, H15, H15, H06),
             ncol = 2)

H70 = (57 * EXP - 302 * EXP2 + 302 * EXP3 - 57 * EXP4 + EXP5 - 1) *
      EXP / (INCEXP ^ 7)
```

```
H61 = x1 * H70
H52 = x2 * H70
H43 = x3 * H70
H34 = x4 * H70
H25 = x5 * H70
H16 = x6 * H70
H07 = x7 * H70
H7m = matrix(c(H70, H61, H61, H52, H61, H52, H52, H43, H61, H52, H52, H43, H52,
               H43, H43, H34, H61, H52, H52, H43, H52, H43, H43, H34, H52, H43,
               H43, H34, H43, H34, H34, H25, H61, H52, H52, H43, H52, H43, H43,
               H34, H52, H43, H43, H34, H43, H34, H34, H25, H52, H43, H43, H34,
               H43, H34, H34, H25, H43, H34, H34, H25, H34, H25, H25, H16, H61,
               H52, H52, H43, H52, H43, H43, H34, H52, H43, H43, H34, H43, H34,
               H34, H25, H52, H43, H43, H34, H43, H34, H34, H25, H43, H34, H34,
               H25, H34, H25, H25, H16, H52, H43, H43, H34, H43, H34, H34, H25,
               H43, H34, H34, H25, H34, H25, H25, H16, H43, H34, H34, H25, H34,
               H25, H25, H16, H34, H25, H25, H16, H25, H16, H16, H07), ncol = 2)

H80 = (120 * EXP - 1191 * EXP2 + 2416 * EXP3 - 1191 * EXP4 +
       120 * EXP5 - EXP6 - 1) * EXP / (INCEXP ^ 8)
H71 = x1 * H80
H62 = x2 * H80
H53 = x3 * H80
H44 = x4 * H80
H35 = x5 * H80
H26 = x6 * H80
H17 = x7 * H80
H08 = x8 * H80
H8m = matrix(c(H80, H71, H71, H62, H71, H62, H62, H53, H71, H62, H62, H53, H62,
               H53, H53, H44, H71, H62, H62, H53, H62, H53, H53, H44, H62, H53,
               H53, H44, H53, H44, H44, H35, H71, H62, H62, H53, H62, H53, H53,
               H44, H62, H53, H53, H44, H53, H44, H44, H35, H62, H53, H53, H44,
               H53, H44, H44, H35, H53, H44, H44, H35, H44, H35, H35, H26, H71,
               H62, H62, H53, H62, H53, H53, H44, H62, H53, H53, H44, H53, H44,
               H44, H35, H62, H53, H53, H44, H53, H44, H44, H35, H53, H44, H44,
               H35, H44, H35, H35, H26, H62, H53, H53, H44, H53, H44, H44, H35,
               H53, H44, H44, H35, H44, H35, H35, H26, H53, H44, H44, H35, H44,
               H35, H35, H26, H44, H35, H35, H26, H35, H26, H26, H17, H71, H62,
               H62, H53, H62, H53, H53, H44, H62, H53, H53, H44, H53, H44, H44,
               H35, H62, H53, H53, H44, H53, H44, H44, H35, H53, H44, H44, H35,
               H44, H35, H35, H26, H62, H53, H53, H44, H53, H44, H44, H35, H53,
               H44, H44, H35, H44, H35, H35, H26, H53, H44, H44, H35, H44, H35,
               H35, H26, H44, H35, H35, H26, H35, H26, H26, H17, H62, H53, H53,
               H44, H53, H44, H44, H35, H53, H44, H44, H35, H44, H35, H35, H26,
               H53, H44, H44, H35, H44, H35, H35, H26, H44, H35, H35, H26, H35,
               H26, H26, H17, H53, H44, H44, H35, H44, H35, H35, H26, H44, H35,
               H35, H26, H35, H26, H26, H17, H44, H35, H35, H26, H35, H26, H26,
               H17, H35, H26, H26, H17, H26, H17, H17, H08), ncol = 2)

V2 = kronecker(V, V)
V3 = kronecker(V2, V)
V4 = kronecker(V3, V)
V5 = kronecker(V4, V)
V6 = kronecker(V5, V)
V7 = kronecker(V6, V)
V8 = kronecker(V7, V)
V9 = kronecker(V8, V)

H3m2 = kronecker(H3m, H3m)
```

```
  H3m3 = kronecker(H3m2, H3m)
  H3m4 = kronecker(H3m2, H3m2)
  H3m6 = kronecker(H3m3, H3m3)
  H4m2 = kronecker(H4m, H4m)
  H4m3 = kronecker(H4m2, H4m)
  H5m2 = kronecker(H5m, H5m)

  a1 = (5/24) * vecT(V3) %*% vec(H3m2) + (1/8) * vecT(V2) %*% vec(H4m)

  a2 = (385/1152) * vecT(V6) %*% vec(H3m4) +
       (35/64) * vecT(V5) %*% vec(kronecker(H3m2, H4m)) +
       vecT(V4) %*% ((7/48) * vec(kronecker(H3m, H5m)) + (35/384) * vec(H4m2)) +
       (1/48) * vecT(V3) %*% vec(H6m)

  a3 = (85085/82944) * vecT(V9) %*% vec(H3m6) +
       (25025/9216) * vecT(V8) %*% vec(kronecker(H3m4, H4m)) +
       vecT(V7) %*% ((5005/3072) * vec(kronecker(H3m2, H4m2)) +
                     (1001/1152) * vec(kronecker(H3m3, H5m))) +
       vecT(V6) %*% ((77/128) * vec(kronecker(H3m, kronecker(H4m, H5m))) +
                     (77/384) * vec(kronecker(H3m2, H6m)) +
                     (385/3072) * vec(H4m3)) +
       vecT(V5) %*% ((7/128) * vec(kronecker(H4m, H6m)) +
                     (21/640) * vec(H5m2) + (1/32) * vec(kronecker(H3m, H7m))) +
       (1/384) * vecT(V4) %*% vec(H8m)

  Laplace_I   = 2 * pi * sqrt(det(V)) * exp(h(m0, m1))
  Laplace_II  = Laplace_I * (1 + a1)
  Laplace_III = Laplace_I * (1 + a1 + a2)
  Laplace_IV  = Laplace_I * (1 + a1 + a2 + a3)
  Laplace6    = Laplace_I * (1 + a1 + (1/48) * vecT(V3) %*% vec(H6m))
  Laplace8    = Laplace_I * (1 + a1 +
                             vecT(V4) %*% ((7/48) * vec(kronecker(H3m, H5m)) +
                                           (35/384) * vec(H4m2)) +
                             (1/48) * vecT(V3) %*% vec(H6m) +
                             (1/384) * vecT(V4) %*% vec(H8m))

  AQ_DT = integrate(function(eps1) {
            sapply(eps1, function(eps1) {
              integrate(function(eps0) f(eps0, eps1), -Inf, Inf)$value
            })
          }, -Inf, Inf)$value

  AQ_LT = integrate(function(eps1) {
            sapply(eps1, function(eps1) {
              integrate(function(eps0) f(eps0, eps1), -Inf, Inf,
                        subdivisions = MAXSUB, rel.tol = TOL, abs.tol = TOL)$value
            })
          }, -Inf, Inf, subdivisions = MAXSUB, rel.tol = TOL, abs.tol = TOL)$value

  c(AQ_DT, AQ_LT,
    Laplace_I, Laplace_II, Laplace_III, Laplace_IV, Laplace6, Laplace8)
}
```

The following R code was used for part of an investigation into the differences between the AQ-DT and AQ-LT estimates of I for various values of τ_0^2, τ_1^2, and τ_{01} in matrix

$$\Upsilon = \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix}.$$

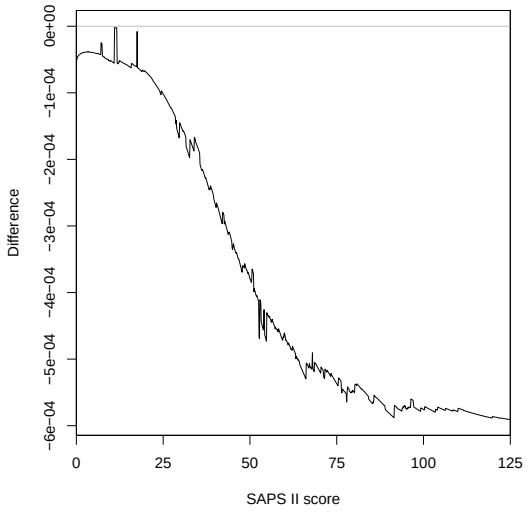
This investigation retains the use of the estimated values for the coefficients of the hierarchical model in Glance et al. (2003), so

$$\begin{aligned} \eta &= \mu + \alpha_1 x_1 + \alpha_2 x_2 \\ &= -8.100 + 1.094 x_1 + 0.063 x_2, \end{aligned}$$

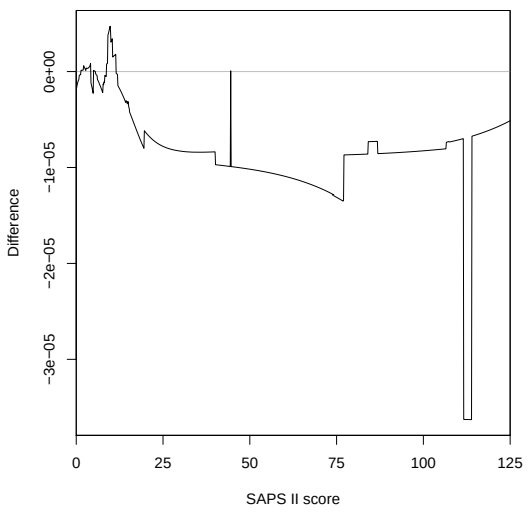
where $x_1 = \log(x_2 + 1)$ and x_2 represents the Simplified Acute Physiology Score (SAPS) II, an integer-valued predictor of mortality. The entire range of SAPS II scores in Glance et al. (2003), from 0 to 125, and intermediate, non-integer values were considered in order to search for seemingly unexpected results which occasionally arise. Figure B.2 displays the plots (ordered by increasing $|\Upsilon|$) of the differences between the corresponding AQ estimates. All parameter estimates (including those in matrix Υ) from Glance et al. (2003) are used for Figure B.2a, so this plot enables a more detailed examination of the differences between the AQ estimates than Figure 3.3a; notably, the newer plot displays a spike between the SAPS II scores of 17 and 18. Figures B.2b to B.2d were produced after selecting different values for the elements of matrix Υ. Each plot reveals spikes or regions where the differences between AQ estimates had a much smaller or larger magnitude at particular SAPS II scores. The salient features in each plot generally occurred around different scores, though closer scrutiny of the plots indicates that a rather large spike occurred at the score of 44.5 in Figure B.2b and just above this score, at 44.625, in Figure B.2d. It is also noted that of all four plots in Figure B.2, the absolute differences between estimates were of largest magnitude in Figure B.2a, apart from the early spikes in this plot. However, as expected, the differences between estimates never exceeded the requested tolerance levels.

```
PlotDifferences = function(tau0s, tau1s, tau01)
{
  TOL    = 1e-12      # absolute and relative tolerances for AQ-LT
  MAXSUB = 1000       # maximum number of subintervals for AQ-LT
  differences = NULL  # stores differences between AQ-DT and AQ-LT

  # Following the work of Glance et al. (2003), this investigation uses
  # - their estimates of the coefficients for the hierarchical model
  # - the same range of SAPS II scores, but with intermediate scores too

  mu     = -8.100
  alpha1 =  1.094
  alpha2 =  0.063
  SAPS2  = seq(0, 125, 0.125)

  detVCV = tau0s * tau1s - tau01 ^ 2

  for (saps2 in SAPS2)
  {
    x1 = log(saps2 + 1)
    x2 = saps2
    eta = mu + alpha1 * x1 + alpha2 * x2

    f = function(eps0, eps1)
        exp(-((eps0 ^ 2) * tau1s - 2 * eps0 * eps1 * tau01 + (eps1 ^ 2) * tau0s) /
        (2 * detVCV) - log(1 + exp(-eta - eps0 - eps1 * x1))) /
        (2 * pi * sqrt(detVCV))

    AQ_DT = integrate(function(eps1) {
```

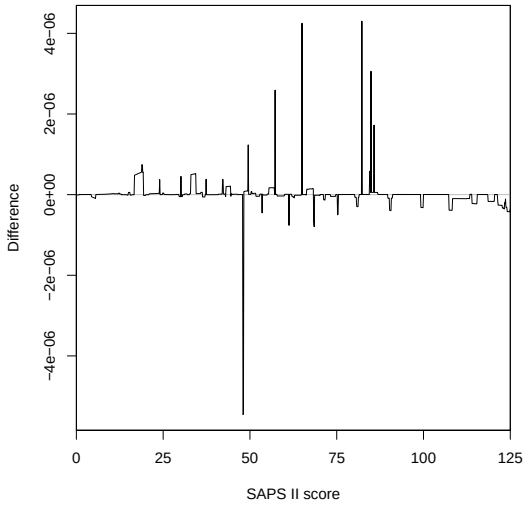
```
              sapply(eps1, function(eps1) {
                integrate(function(eps0) f(eps0, eps1), -Inf, Inf)$value
              })
            }, -Inf, Inf)$value

    AQ_LT = integrate(function(eps1) {
              sapply(eps1, function(eps1) {
                integrate(function(eps0) f(eps0, eps1), -Inf, Inf,
                          subdivisions = MAXSUB, rel.tol = TOL, abs.tol = TOL)$value
              })
            }, -Inf, Inf, subdivisions = MAXSUB, rel.tol = TOL, abs.tol = TOL)$
            value

    differences = c(differences, AQ_DT - AQ_LT)
  }

  plot(NA, xlab = "SAPS II score", ylab = "Difference",
       xlim = range(SAPS2), ylim = range(0, differences),
       xaxs = "i", xaxt = "n", type = "b")
  abline(h = 0, col = "grey")
  box()
  axis(1, (0, 125, 25))
  lines(SAPS2, differences)
}

# Using Glance et al.'s (2003) estimated values for tau0s, tau1s, and tau01,
# followed by various other values for those parameters.

PlotDifferences(2.412, 0.196, -0.680)  # detVCV = 0.010352
PlotDifferences(1.3, 0.5, -0.77)       # detVCV = 0.0571
PlotDifferences(1.5, 0.1, -0.1)        # detVCV = 0.14
PlotDifferences(3.5, 0.2, -0.43)       # detVCV = 0.5151
```



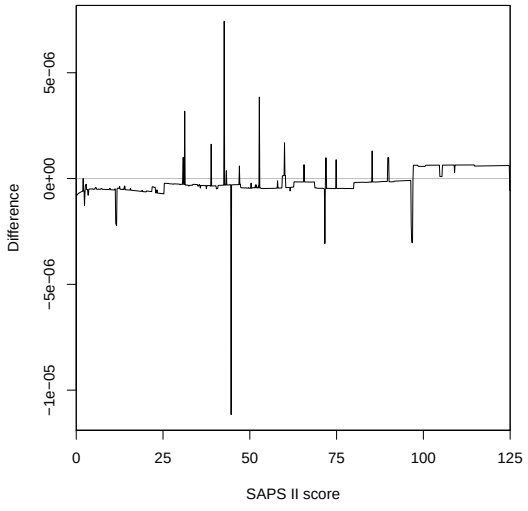
(a) $\tau_0^2 = 2.412$, $\tau_1^2 = 0.196$, $\tau_{01} = -0.680$ from Glance et al. (2003) hierarchical model



(b) $\tau_0^2 = 1.3$, $\tau_1^2 = 0.5$, $\tau_{01} = -0.77$



(c) $\tau_0^2 = 1.5$, $\tau_1^2 = 0.1$, $\tau_{01} = -0.1$



(d) $\tau_0^2 = 3.5$, $\tau_1^2 = 0.2$, $\tau_{01} = -0.43$

Figure B.2: Differences between AQ-DT and AQ-LT estimates of a double integral across the entire range of SAPS II scores in Glance et al. (2003), and interpolated (non-integer) scores, based on various values of τ_0^2, τ_1^2, and τ_{01}.

Appendix C

Further details of the ANZICS application

This appendix presents additional estimates and results for the application in Chapter 4, which used a subset of the data collected by the Australian and New Zealand Intensive Care Society (ANZICS). Matters pertaining to the selected model are presented first, and followed by the alternative, base model.

C.1 Selected model

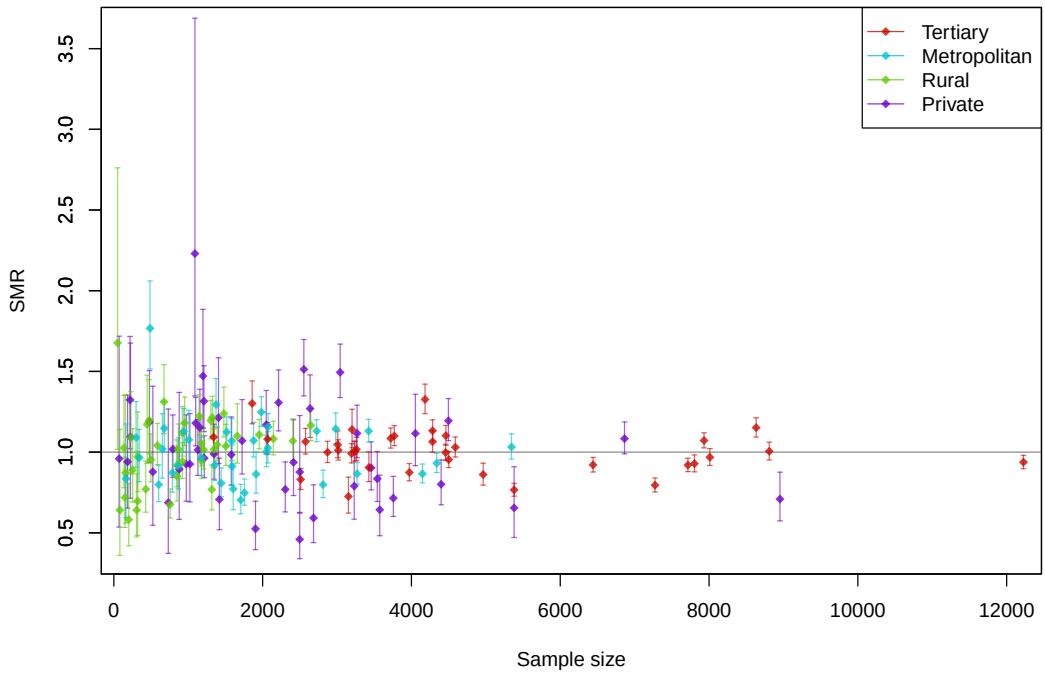
Table C.1 displays the parameter estimates of the fitted model that was selected for the ANZICS application. This model was used to estimate the standardised mortality ratios (SMRs) and corresponding 95% confidence intervals (CIs) for the intensive care units (ICUs). Figures C.1 and C.2 presents these results, by each SMR method, in an alternative manner to the graphical displays that were used in Section 4.4.2.

Table C.1: Parameter estimates of the fitted model that was selected for the ANZICS application. This model was used to predict the log odds of in-hospital mortality for each patient. Age and APACHE III score (abbreviated to APIII) were treated as numeric (continuous) variables: the values of each variable were centred (subtracting the variable's mean), then scaled by dividing the resultant values by ten. The categorical variables, with their reference levels (assigned zero values) in parentheses, were sex (male), ventilation status (not ventilated), ICU source (no transfer), diagnostic category (cardiovascular non-surgical), ICU admission year (2015), and hospital type (tertiary).

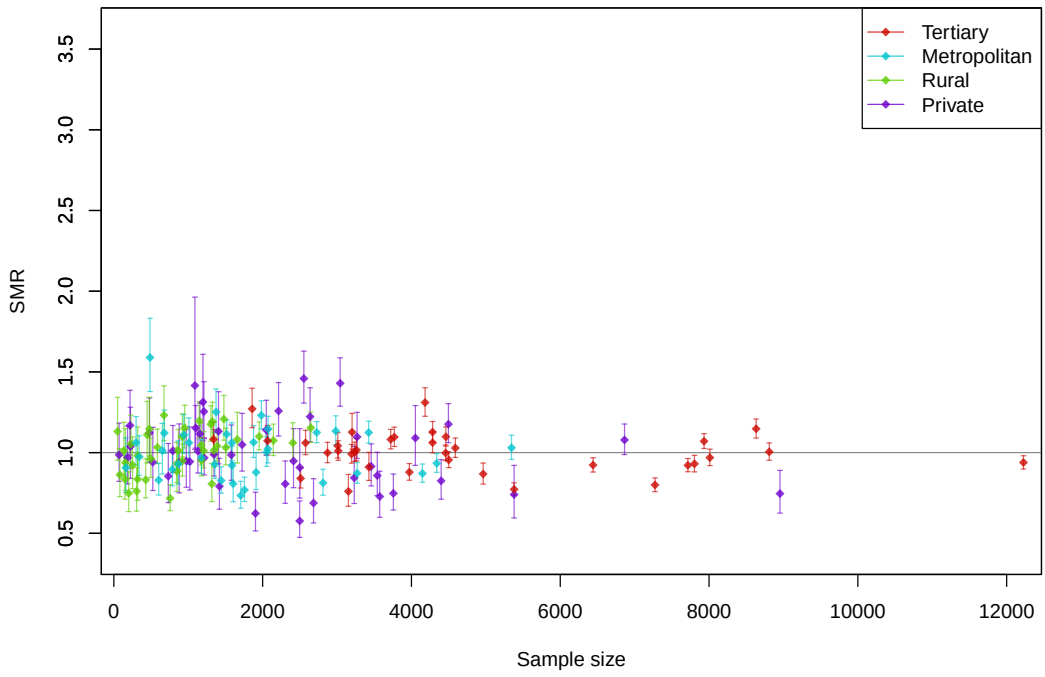
Parameter	**Estimate**	**SE**	**z**	**P > |z|**	**95% CI**
constant	-2.3647	0.0552	-42.8259	< 0.0001	[-2.4729, -2.2564]
APIII	0.6004	0.0083	72.1442	< 0.0001	[0.5841, 0.6167]
APIII squared	-0.0103	0.0008	-13.1459	< 0.0001	[-0.0118, -0.0087]
Age	0.1963	0.0081	24.3073	< 0.0001	[0.1805, 0.2121]
Age squared	0.0083	0.0020	4.1491	< 0.0001	[0.0044, 0.0123]
Sex					
Female	-0.1172	0.0213	-5.5092	< 0.0001	[-0.1588, -0.0755]
Ventilation status					
Ventilated	0.1707	0.0251	6.8122	< 0.0001	[0.1216, 0.2198]

ICU source					
Hospital transfer	-0.1851	0.0299	-6.1877	<0.0001	[-0.2437, -0.1264]
Diagnostic category					
Non-surgical					
Respiratory	0.2967	0.0321	9.2466	<0.0001	[0.2338, 0.3596]
Liver/Gastrointestinal	-0.1695	0.0497	-3.4122	0.0006	[-0.2669, -0.0721]
Central Nervous System	0.4833	0.0375	12.9020	<0.0001	[0.4099, 0.5567]
Sepsis	-0.3723	0.0403	-9.2345	<0.0001	[-0.4514, -0.2933]
Trauma	-0.4429	0.0554	-7.9950	<0.0001	[-0.5515, -0.3343]
Metabolic Hormonal	-1.8120	0.0758	-23.9120	<0.0001	[-1.9605, -1.6634]
Haematologic	0.3845	0.1066	3.6086	0.0003	[0.1757, 0.5934]
Renal/Genitourinary	-0.7532	0.0831	-9.0602	<0.0001	[-0.9161, -0.5902]
Other medical disorders	-0.5986	0.1477	-4.0537	0.0001	[-0.8880, -0.3092]
Musculoskeletal/Skin	-0.3239	0.1772	-1.8280	0.0676	[-0.6712, 0.0234]
Emergency surgical					
Cardiovascular	-0.8139	0.0654	-12.4455	<0.0001	[-0.9421, -0.6857]
Thoracic	-0.7245	0.0924	-7.8409	<0.0001	[-0.9055, -0.5434]
Gastrointestinal	-0.7082	0.0432	-16.4081	<0.0001	[-0.7928, -0.6236]
Central Nervous System	0.2898	0.0538	5.3846	<0.0001	[0.1843, 0.3953]
Traumatic/Orthopaedic	-0.4912	0.0733	-6.6980	<0.0001	[-0.6349, -0.3475]
Renal/Genitourinary	-1.8633	0.1842	-10.1144	<0.0001	[-2.2244, -1.5022]
Gynaecological	-2.2302	0.3195	-6.9805	<0.0001	[-2.8564, -1.6040]
Musculoskeletal/Skin	-1.0383	0.0877	-11.8361	<0.0001	[-1.2103, -0.8664]
Haematologic/Metabolic	-1.4511	0.3664	-3.9606	0.0001	[-2.1692, -0.7330]
Elective surgical					
Cardiovascular	-2.1613	0.0539	-40.1260	<0.0001	[-2.2669, -2.0557]
Thoracic	-1.2413	0.0778	-15.9532	<0.0001	[-1.3938, -1.0888]
Gastrointestinal	-1.3022	0.0534	-24.3794	<0.0001	[-1.4069, -1.1975]
Central Nervous System	-1.1732	0.0712	-16.4805	<0.0001	[-1.3127, -1.0337]
Traumatic/Orthopaedic	-0.7907	0.2071	-3.8181	0.0001	[-1.1966, -0.3848]
Renal/Genitourinary	-2.0561	0.1454	-14.1453	<0.0001	[-2.3410, -1.7712]
Gynaecological	-2.1568	0.2695	-8.0039	<0.0001	[-2.6850, -1.6287]
Musculoskeletal/Skin	-1.7245	0.0921	-18.7163	<0.0001	[-1.9051, -1.5439]
Haematologic/Metabolic	-2.0285	0.2944	-6.8915	<0.0001	[-2.6054, -1.4516]
ICU admission year					
2011	0.0640	0.0205	3.1194	0.0018	[0.0238, 0.1041]
2012	0.0146	0.0203	0.7223	0.4701	[-0.0251, 0.0544]
2013	0.0374	0.0204	1.8321	0.0669	[-0.0026, 0.0773]
2014	0.0201	0.0201	1.0026	0.3161	[-0.0192, 0.0594]
Hospital type					
Metropolitan	-0.2240	0.0653	-3.4332	0.0006	[-0.3519, -0.0961]
Rural	-0.2639	0.0652	-4.0460	0.0001	[-0.3917, -0.1361]
Private	-0.3155	0.0638	-4.9488	<0.0001	[-0.4405, -0.1906]
Diagnostic category × APIII					
Non-surgical					
Respiratory × APIII	-0.1117	0.0080	-13.8887	<0.0001	[-0.1274, -0.0959]
Liver/Gastrointestinal × APIII	0.0205	0.0120	1.7122	0.0869	[-0.0030, 0.0440]
Central Nervous System × APIII	-0.0329	0.0097	-3.3979	0.0007	[-0.0519, -0.0139]
Sepsis × APIII	0.0103	0.0088	1.1696	0.2422	[-0.0070, 0.0276]
Trauma × APIII	0.1495	0.0158	9.4840	<0.0001	[0.1186, 0.1804]
Metabolic Hormonal × APIII	0.0279	0.0159	1.7516	0.0799	[-0.0033, 0.0590]
Haematologic × APIII	-0.0699	0.0262	-2.6674	0.0076	[-0.1213, -0.0185]
Renal/Genitourinary × APIII	-0.0558	0.0201	-2.7845	0.0054	[-0.0951, -0.0165]
Other medical disorders × APIII	-0.0198	0.0405	-0.4898	0.6243	[-0.0992, 0.0595]
Musculoskeletal/Skin × APIII	-0.0629	0.0511	-1.2302	0.2186	[-0.1630, 0.0373]
Emergency surgical					
Cardiovascular × APIII	0.0552	0.0169	3.2624	0.0011	[0.0220, 0.0884]

Thoracic × APIII	-0.0332	0.0296	-1.1244	0.2609	[-0.0911, 0.0247]
Gastrointestinal × APIII	0.0047	0.0104	0.4537	0.6500	[-0.0157, 0.0252]
Central Nervous System × APIII	0.0034	0.0171	0.1980	0.8431	[-0.0302, 0.0370]
Traumatic/Orthopaedic × APIII	0.0833	0.0203	4.0975	<0.0001	[0.0434, 0.1231]
Renal/Genitourinary × APIII	0.0627	0.0453	1.3845	0.1662	[-0.0261, 0.1514]
Gynaecological × APIII	0.0340	0.0755	0.4500	0.6527	[-0.1140, 0.1820]
Musculoskeletal/Skin × APIII	0.0382	0.0234	1.6344	0.1022	[-0.0076, 0.0840]
Haematologic/Metabolic × APIII	0.1289	0.1030	1.2512	0.2108	[-0.0730, 0.3309]
Elective surgical					
Cardiovascular × APIII	0.2087	0.0178	11.7496	<0.0001	[0.1739, 0.2435]
Thoracic × APIII	0.0426	0.0378	1.1267	0.2599	[-0.0315, 0.1167]
Gastrointestinal × APIII	0.0533	0.0188	2.8313	0.0046	[0.0164, 0.0903]
Central Nervous System × APIII	0.0437	0.0336	1.3012	0.1932	[-0.0221, 0.1096]
Traumatic/Orthopaedic × APIII	0.0765	0.0815	0.9380	0.3482	[-0.0833, 0.2363]
Renal/Genitourinary × APIII	-0.0230	0.0568	-0.4041	0.6862	[-0.1344, 0.0884]
Gynaecological × APIII	0.1697	0.1074	1.5796	0.1142	[-0.0409, 0.3802]
Musculoskeletal/Skin × APIII	0.1873	0.0370	5.0655	<0.0001	[0.1148, 0.2598]
Haematologic/Metabolic × APIII	0.0767	0.1791	0.4284	0.6684	[-0.2743, 0.4277]
APIII × Age	-0.0218	0.0017	-13.0919	<0.0001	[-0.0251, -0.0186]
ICU source × APIII					
Hospital transfer × APIII	-0.0227	0.0070	-3.2337	0.0012	[-0.0364, -0.0089]
Ventilation status × APIII					
Ventilated × APIII	-0.1111	0.0067	-16.6781	<0.0001	[-0.1242, -0.0981]
Ventilation status × Age					
Ventilated × Age	0.0207	0.0098	2.1057	0.0352	[0.0014, 0.0400]
Ventilation status × Sex					
Ventilated × Female	0.1271	0.0270	4.7155	<0.0001	[0.0743, 0.1799]
Random effects: ICU site (variance)	0.0678	0.0097			[0.0512, 0.0898]

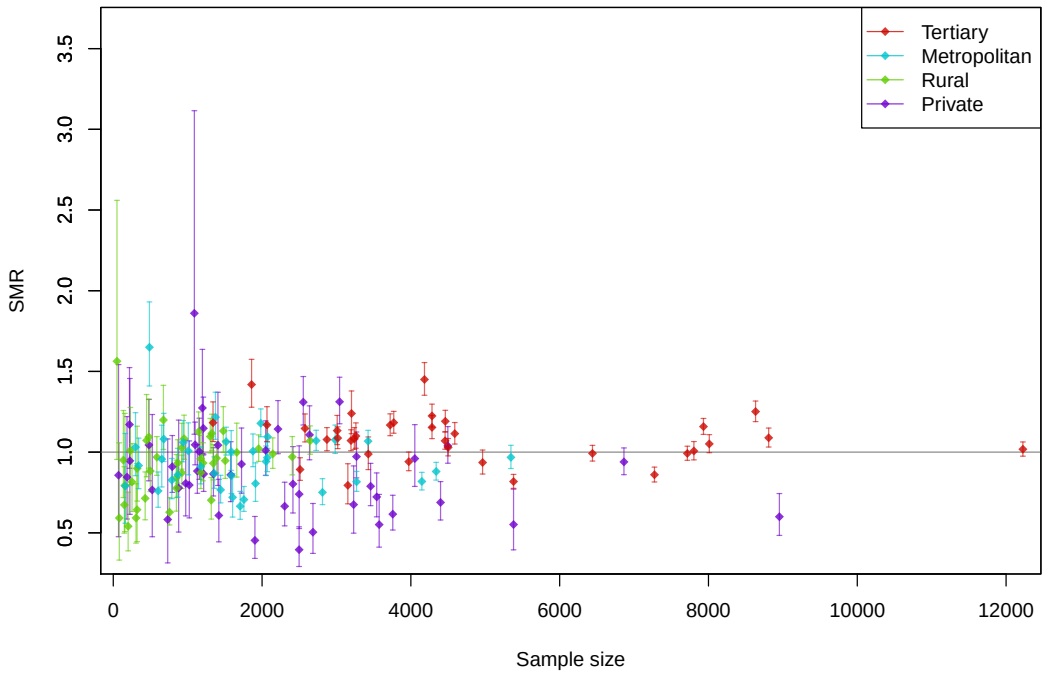


(a) Standard approach (Method 1)

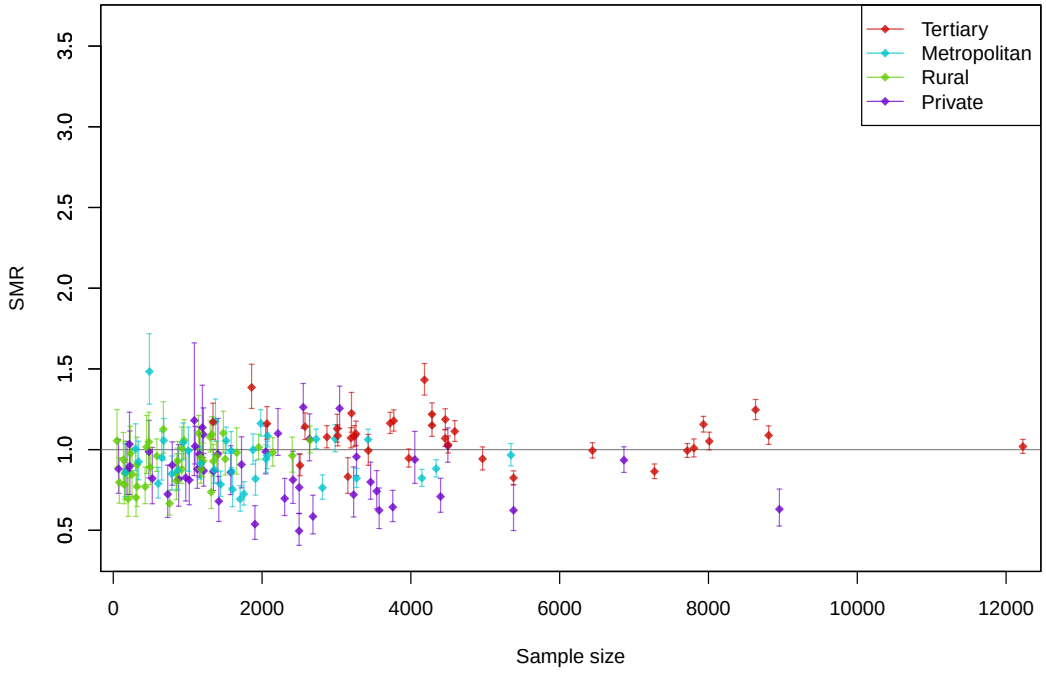


(b) Smoothed approach (Method 2)

Figure C.1: Estimated SMRs and approximate 95% CIs by actual comparators.



(a) Standard approach (Method 3)



(b) Smoothed approach (Method 4)

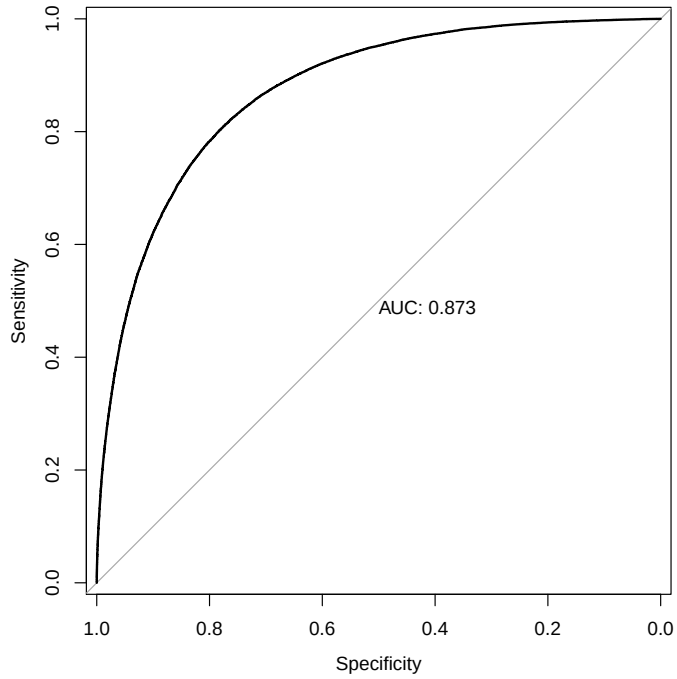
Figure C.2: Estimated SMRs and approximate 95% CIs by average comparators.

C.2 Base model

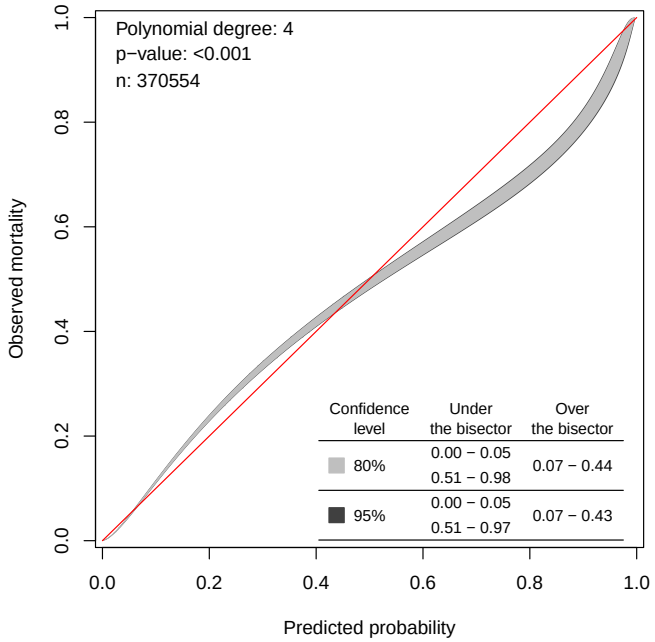
Table C.2 displays the parameter estimates of the random intercept model which limited the predictors to (linear) Acute Physiology and Chronic Health Evaluation (APACHE) III score. This was briefly referred to in Section 4.3.2. Examination of the receiver operating characteristic (ROC) curve (Figure C.3a) revealed that APACHE III scores were excellent at discriminating between patient outcomes. Indeed the area under the curve (AUC) of 0.873 was nearly as large as that for the selected model (0.894). However, the calibration belts (Figure C.3b) revealed that APACHE III scores were poorly calibrated. Mortality rates were significantly overestimated at very low (below 0.05) and moderate to very high (between 0.51 and 0.97) predicted probabilities. Furthermore, mortality rates were significantly underestimated for low to moderate predicted probabilities (between 0.07 and 0.43).

Table C.2: Parameter estimates of the base model (not selected for the ANZICS application). If the model was adequate, it would have been used to predict the log odds of in-hospital mortality for each patient. As before, APACHE III score (abbreviated to APIII) was treated as a numeric (continuous) variable whose values were centred (subtracting the variable's mean), then scaled by dividing the resultant values by ten.

Parameter	**Estimate**	**SE**	**z**	**P > |z|**	**95% CI**
constant	-2.9603	0.0374	-79.1848	< 0.0001	[-3.0336, -2.8870]
APIII	0.4943	0.0024	206.8021	< 0.0001	[0.4896, 0.4990]
Random effects: ICU site (variance)	0.2023	0.0257			[0.1578, 0.2595]



(a) Receiver operating characteristic (ROC) curve where the cut-off point being varied is the predicted mortality probability



(b) Calibration belt plot displaying the relationship between the predicted mortality probability and the observed proportion of mortalities

Figure C.3: Diagnostic plots for assessing the adequacy of the base model by its discrimination and calibration.

Bibliography

ABS (2019), *Australian Demographic Statistics,* 3101.0, Australian Bureau of Statistics, Canberra. Accessed: 14 November 2019.
URL: *https://www.abs.gov.au/AUSSTATS/subscriber.nsf/log?openagent&31010_mar%202019.pdf&3101.0&Publication&ACFD01ECA5FD8A88CA25848800154862&&Mar%202019&04.10.2019&Latest*

AIHW (2018), "Myhospitals overview", Australian Institute of Health and Welfare, Canberra. Accessed: 19 September 2019.
URL: *https://www.myhospitals.gov.au/about-myhospitals/overview*

AIHW (2019), *Australia's welfare 2019: in brief*, Australian Institute of Health and Welfare, Canberra. Accessed: 1 June 2020.
URL: *https://www.aihw.gov.au/getmedia/795385cc-6493-45c9-b341-7ddf6006d518/aihw-aus-227.pdf*

ANZICS (2019), "Adult Patient Database (APD)", Australian and New Zealand Intensive Care Society, Melbourne. Accessed: 29 September 2019.
URL: *https://www.anzics.com.au/adult-patient-database-apd/*

ANZICS (2020), "Organisation & Vision", Australian and New Zealand Intensive Care Society, Melbourne. Accessed: 24 July 2020.
URL: *https://www.anzics.com.au/organisation-vision/*

ANZICS CORE (2017), *ANZICS CORE Annual Report 2017*, Australian and New Zealand Intensive Care Society Centre for Outcome and Resource Evaluation, Melbourne. Accessed: 28 September 2019.
URL: *https://www.anzics.com.au/wp-content/uploads/2018/10/ANZICS-CORE-Annual-Report-2017.pdf*

ANZICS CORE (2018), *APD Data Dictionary for Software Programmers, Version 5.7*, Australian and New Zealand Intensive Care Society Centre for Outcome and Resource Evaluation, Melbourne. Accessed: 29 September 2018.
URL: *https://www.anzics.com.au/wp-content/uploads/2018/08/ANZICS-APD-Programmers-Data-Dictionary.pdf*

ANZICS CORE (2019), *ANZICS CORE Adult Patient Database (APD) Activity Report 2017/2018*, Australian and New Zealand Intensive Care Society Centre for Outcome and Resource Evaluation, Melbourne. Accessed: 28 September 2019.
URL: *https://www.anzics.com.au/wp-content/uploads/2019/07/ANZICS-CORE-APD-Activity-Report-2017-18.pdf*

Ash, A.S., Fienberg, S.E., Louis, T.A., Normand, S.L.T., Stukel, T.A. and Utts, J. (2012), *Statistical Issues in Assessing Hospital Performance*, Commissioned by

the Committee of Presidents of Statistical Societies (COPSS) for the Centers for Medicare and Medicaid Services (CMS). Accessed: 19 October 2015.
URL: *http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Downloads/Statistical-Issues-in-Assessing-Hospital-Performance.pdf*

Austin, P.C., Alter, D.A. and Tu, J.V. (2003), "The use of fixed- and random-effects models for classifying hospitals as mortality outliers: a Monte Carlo assessment", *Medical Decision Making* **23**(6), 526–539.

Azevedo-Filho, A. and Shachter, R.D. (1994), "Laplace's method approximations for probabilistic inference in belief networks with continuous variables", *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence* pp. 28–36.

Benjamini, Y. and Hochberg, Y. (1995), "Controlling the false discovery rate: a practical and powerful approach to multiple testing", *Journal of the Royal Statistical Society. Series B* **57**(1), 289–300.

Berthelsen, P.G. and Cronqvist, M. (2003), "The first intensive care unit in the world: Copenhagen 1953", *Acta Anaesthesiologica Scandinavica* **47**(10), 1190–1195.

Bottle, A. and Aylin, P. (2016), *Statistical Methods for Healthcare Performance Monitoring*, Chapman & Hall/CRC, Boca Raton, FL.

Breslow, N.E. and Day, N.E. (1987), *Statistical methods in cancer research. Volume II – The design and analysis of cohort studies*, Oxford University Press, Oxford.

Brinkman, S., Abu-Hanna, A., Jonge, E. and Keizer, N.F. (2013), "Prediction of long-term mortality in ICU patients: model validation and assessing the effect of using in-hospital versus long-term mortality on benchmarking", *Intensive Care Medicine* **39**(11), 1925–1931.

Burden, R.L. and Faires, J.D. (1993), *Numerical Analysis*, 5th edn, PWS-Kent Pub. Co., Boston, MA.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm", *The American Statistician* **49**(4), 327–335.

CIHI (2019), "Your Health System", Canadian Institute for Health Information, Ottawa, ON. Accessed: 19 September 2019.
URL: *https://yourhealthsystem.cihi.ca/hsp/indepth?lang=en#/*

CMS (2019*a*), "Measure Methodology", Centers for Medicare and Medicaid Services, Baltimore, MD. Accessed: 13 February 2019.
URL: *https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html*

CMS (2019*b*), "Medicare Hospital Quality Chartbook", Centers for Medicare and Medicaid Services, New Haven, CT. Accessed: 19 September 2019.
URL: *https://www.cmshospitalchartbook.com/*

Codman, E.A. (1918), *A Study in Hospital Efficiency: As Demonstrated by the Case Report of the First Five Years of a Private Hospital*, Thomas Todd Company, Boston, MA.

Cox, D.R. and Solomon, P.J. (2002), *Components of Variance*, Chapman & Hall/CRC, Boca Raton, FL.

Davison, A.C. (2003), *Statistical Models*, Cambridge University Press, Cambridge.

DeLong, E.R., Peterson, E.D., DeLong, D.M., Muhlbaier, L.H., Hackett, S. and Mark, D.B. (1997), "Comparing risk-adjustment methods for provider profiling", *Statistics in Medicine* **16**(23), 2645–2664.

Duckett, S., Cuddihy, M. and Newnham, H. (2016), *Targeting zero: supporting the Victorian hospital system to eliminate avoidable harm and strengthen quality of care*, Department of Health and Human Services, Victorian Government, Melbourne. Accessed: 23 July 2020.
URL: *www2.health.vic.gov.au/hospitals-and-health-services/quality-safety-service/hospital-safety-and-quality-review*

Dyer, C. (2001), "Bristol inquiry", *BMJ* **323**(7306), 181.

Dyer, C. (2015), "Mid Staffordshire trust pleads guilty to health and safety breaches", *BMJ* **351**(h5973).

Efron, B. and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall, New York, NY.

Farr, W. (1861), *Report of the proceedings of the fourth session of the International Statistical Congress: held in London July 16th, 1860, and the five following days*, HM Stationery Office, London.

Forbes, C., Evans, M., Hastings, N. and Peacock, B. (2011), *Statistical Distributions*, 4th edn, Wiley, Hoboken, NJ.

Glance, L., Dick, A., Osler, T. and Mukamel, D. (2003), "Using hierarchical modeling to measure ICU quality", *Intensive Care Medicine* **29**(12), 2223–2229.

Goldstein, H. (2010), *Multilevel Statistical Models*, 4th edn, Wiley, Hoboken, NJ.

Goldstein, H. and Spiegelhalter, D. (1996), "League tables and their limitations: statistical issues in comparisons of institutional performance", *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **159**(3), 385–409.

Gordon, A.C. and Russell, J.A. (2018), "Innovation and safety in critical care: should we collaborate with the industry? Pro", *Intensive Care Medicine* **44**(12), 2276–2278.

Halm, E.A., Lee, C. and Chassin, M.R. (2002), "Is volume related to outcome in health care? A systematic review and methodologic critique of the literature", *Annals of Internal Medicine* **137**(6), 511–520.

Hart, G.K. (2008), "The ANZICS CORE: an evolution in registry activities for intensive care in Australia and New Zealand", *Critical Care and Resuscitation* **10**(2), 83–88.

Hartigan, J.A. (1983), *Bayes Theory*, Springer-Verlag, New York, NY.

Heijink, R., Koolman, X., Pieter, D., van der Veen, A., Jarman, B. and Westert, G. (2008), "Measuring and explaining mortality in Dutch hospitals; The Hospital Standardized Mortality Rate between 2003 and 2005", *BMC Health Services Research* **8**(73).

Holmes, B. (1905), "The most ancient medical practice laws. The code of Hammurabi, 2200 B.C.", *Journal of the American Medical Association* **XLIV**(4), 293–294.

Hosmer, D.W. and Lemeshow, S. (1995), "Confidence interval estimates of an index of quality performance based on logistic regression models", *Statistics in Medicine* **14**(19), 2161–2172.

Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013), *Applied Logistic Regression*, 3rd edn, Wiley, Hoboken, NJ.

Huang, D.T. (2004), "Clinical review: Impact of emergency department care on intensive care unit costs", *Critical Care* **8**(6), 498–502.

ICNARC (2019), "Online Reports", Intensive Care National Audit and Research Centre, London. Accessed: 1 October 2019.
URL: *https://onlinereports.icnarc.org/*

Jha, A.K. and Zaslavsky, A.M. (2014), "Quality reporting that addresses disparities in health care", *JAMA* **312**(3), 225–226.

Kalbfleisch, J. and Wolfe, R. (2013), "On Monitoring Outcomes of Medical Providers", *Statistics in Biosciences* **5**(2), 286–302.

Kasza, J., Moran, J.L. and Solomon, P.J. (2013), "Evaluating the performance of Australian and New Zealand intensive care units in 2009 and 2010", *Statistics in Medicine* **32**(21), 3720–3736.

Kasza, J., Polkinghorne, K.R., Wolfe, R., McDonald, S.P. and Marshall, M.R. (2018), "Comparing dialysis centre mortality outcomes across Australia and New Zealand: identifying unusually performing centres 2008–2013", *BMC Health Services Research* **18**(1).

Kelly, F.E., Fong, K., Hirsch, N. and Nolan, J.P. (2014), "Intensive care medicine is 60 years old: the history and future of the intensive care unit", *Clinical Medicine* **14**(4), 376–379.

Knaus, W.A., Draper, E.A., Wagner, D.P. and Zimmerman, J.E. (1985), "APACHE II: a severity of disease classification system", *Critical Care Medicine* **13**(10), 818–829.

Knaus, W.A., Wagner, D.P., Draper, E.A., Zimmerman, J.E., Bergner, M., Bastos, P.G., Sirio, C.A., Murphy, D.J., Lotring, T., Damiano, A. and Harrell, F.E. (1991), "The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalised adults", *Chest* **100**(6), 1619–1636.

Knaus, W.A., Zimmerman, J.E., Wagner, D.P., Draper, E.A. and Lawrence, D.E. (1981), "APACHE – acute physiology and chronic health evaluation: a physiologically based classification system", *Critical Care Medicine* **9**(8), 591–597.

Kramer, A.A. and Zimmerman, J.E. (2008), "Predicting outcomes for cardiac surgery patients after intensive care unit admission", *Seminars in Cardiothoracic and Vascular Anesthesia* **12**(3), 175–183.

Laplace, P.S. (1986), "Memoir on the probability of the causes of events", *Statistical Science* **1**(3), 364–378.

Le Gall, J.R., Lemeshow, S. and Saulnier, F. (1993), "A new Simplified Acute Physiology Score (SAPS II) based on a European/North American Multicenter Study", *JAMA* **270**(24), 2957–2963.

Lindmark, A., van Rompaye, B., Goetghebeur, E., Glader, E.L. and Eriksson, M. (2016), "The importance of integrating clinical relevance and statistical significance in the assessment of quality of care – illustrated using the Swedish Stroke Register", *PLoS ONE* **11**(4).

Luft, H.S., Bunker, J.P. and Enthoven, A.C. (1979), "Should operations be regionalized? The empirical relation between surgical volume and mortality", *The New England Journal of Medicine* **301**(25), 1364–1369.

Magnus, J.R. and Neudecker, H. (2007), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd edn, Wiley, Chichester.

Manktelow, B.N., Evans, T.A. and Draper, E.S. (2014), "Differences in case-mix can influence the comparison of standardised mortality ratios even with optimal risk adjustment: an analysis of data from paediatric intensive care", *BMJ Quality & Safety* **23**(9), 782–788.

McClean, K., Mullany, D., Huckson, S., van Lint, A., Chavan, S., Hicks, P., Hart, G., Paul, E. and Pilcher, D. (2017), "Identification and assessment of potentially high-mortality intensive care units using the ANZICS Centre for Outcome and Resource Evaluation clinical registry", *Critical Care and Resuscitation* **19**(3), 230–238.

McGain, F., Durie, M.L., Bates, S., Polmear, C.M., Meyer, J. and French, C.J. (2018), "Smoking cessation therapy in Australian and New Zealand intensive care units: a multicentre point prevalence study", *Critical Care and Resuscitation* **20**(1), 68–73.

MedicineNet (2020), "Critical care". Accessed: 26 July 2020.
URL: *https://www.medicinenet.com/script/main/art.asp?articlekey=24812*

Mersmann, O. (2018), *microbenchmark: Accurate Timing Functions*. R package version 1.4-4.
URL: *https://CRAN.R-project.org/package=microbenchmark*

Mohammed, M.A., Deeks, J.J., Girling, A., Rudge, G., Carmalt, M., Stevens, A.J. and Lilford, R.J. (2009), "Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals", *BMJ* **338**(b780).

Mohammed, M.A., Manktelow, B.N. and Hofer, T.P. (2016), "Comparison of four methods for deriving hospital standardised mortality ratios from a single hierarchical logistic regression model", *Statistical Methods in Medical Research* **25**(2), 706–715.

Moran, J.L. and Solomon, P.J. (2012), "Mortality and intensive care volume in ventilated patients from 1995 to 2009 in the Australian and New Zealand binational adult patient intensive care database", *Critical Care Medicine* **40**(3), 800–812.

Moran, J.L. and Solomon, P.J. (2014), "Fixed effects modelling for provider mortality outcomes: analysis of the Australia and New Zealand Intensive Care Society (ANZICS) Adult Patient Data-Base", *PLoS ONE* **9**(7), e102297.

Narasimhan, B. and Johnson, S.G. (2017), *cubature: Adaptive Multivariate Integration over Hypercubes*. R package version 1.3-11.
URL: *https://CRAN.R-project.org/package=cubature*

Nattino, G., Finazzi, S. and Bertolini, G. (2016), "A new test and graphical tool to assess the goodness of fit of logistic regression models", *Statistics in Medicine* **35**(5), 709–720.

Nattino, G., Finazzi, S., Bertolini, G., Rossi, C. and Carrara, G. (2017), *givitiR: The GiViTI Calibration Test and Belt*. R package version 1.3.
URL: *https://CRAN.R-project.org/package=givitiR*

Neal, R.M. (2014), "Inaccurate results from microbenchmark", *Radford Neal's blog*. Accessed: 28 April 2018.
URL: *https://radfordneal.wordpress.com/2014/02/02/inaccurate-results-from-microbenchmark/*

Neuhauser, D. (2002), "Ernest Amory Codman MD", *BMJ Quality & Safety* **11**(1).

NHS Digital (2019), "Summary Hospital-level Mortality Indicator (SHMI) – Deaths associated with hospitalisation", National Health Service Digital, Leeds. Accessed: 19 September 2019.
URL: *https://digital.nhs.uk/data-and-information/publications/clinical-indicators/shmi*

Nightingale, F. (1863), *Notes on Hospitals*, 3rd edn, Longman, Green, Longman, Roberts, and Green, London.

Normand, S.L.T., Ash, A.S., Fienberg, S.E., Stukel, T.A., Utts, J. and Louis, T.A. (2016), "League tables for hospital comparisons", *Annual Review of Statistics and Its Application* **3**, 21–50.

Normand, S.L.T., Glickman, M.E. and Gatsonis, C.A. (1997), "Statistical methods for profiling providers of medical care: Issues and applications", *Journal of the American Statistical Association* **92**(439), 803–814.

NZ Ministry of Health (2018), "Socioeconomic indicators", New Zealand Ministry of Health, Wellington. Accessed: 1 June 2020.
URL: *https://www.health.govt.nz/our-work/populations/maori-health/tatau-kahukura-maori-health-statistics/nga-awe-o-te-hauora-socioeconomic-determinants-health/socioeconomic-indicators*

Paul, E., Bailey, M., van Lint, A.L. and Pilcher, D.V. (2012), "Performance of APACHE III over time in Australia and New Zealand: a retrospective cohort study", *Anaesthesia and Intensive Care* **40**(6), 980–994.

Paul, P., Pennell, M.L. and Lemeshow, S. (2013), "Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets", *Statistics in Medicine* **32**(1), 67–80.

Perner, A., Laake, J.H. and van der Horst, I.C.C. (2018), "Innovation and safety in critical care: should we collaborate with the industry? Con", *Intensive Care Medicine* **44**(12), 2279–2281.

Phillips, B., Daniels, J., Woodward, A., Blakely, T., Taylor, R. and Morrell, S. (2017), "Mortality trends in Australian Aboriginal peoples and New Zealand Māori", *Population Health Metrics* **15**(1), 1–12.

Pouw, M.E., Peelen, L.M., Lingsma, H.F., Pieter, D., Steyerberg, E., Kalkman, C.J. and Moons, K.G.M. (2013), "Hospital standardized mortality ratio: consequences of adjusting hospital mortality with indirect standardization", *PLoS ONE* **8**(4), e59160.

R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Rabe-Hesketh, S. and Skrondal, A. (2012), *Multilevel and Longitudinal Modeling Using Stata*, 3rd edn, Stata Press, College Station, TX.

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F. and Congdon, R. (2019), *HLM 8 for Windows*, Scientific Software International Inc., Skokie, IL.

Raudenbush, S.W., Yang, M.L. and Yosef, M. (2000), "Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation", *Journal of Computational and Graphical Statistics* **9**(1), 141–157.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C. and Müller, M. (2011), "pROC: an open-source package for R and S+ to analyze and compare ROC curves", *BMC Bioinformatics* **12**, 77.

Ruli, E., Sartori, N. and Ventura, L. (2016), "Improved Laplace approximation for marginal likelihoods", *Electronic Journal of Statistics* **10**(2), 3986–4009.

Searle, S.R. (1982), *Matrix Algebra Useful for Statistics*, Wiley, New York, NY.

Shahian, D.M., Torchiana, D.F., Shemin, R.J., Rawn, J.D. and Normand, S.L.T. (2005), "Massachusetts cardiac surgery report card: implications of statistical methodology", *Annals of Thoracic Surgery* **80**(6), 2106–2113.

Silber, J.H., Satopää, V.A., Mukherjee, N., Rockova, V., Wang, W., Hill, A.S., Even-Shoshan, O., Rosenbaum, P.R. and George, E.I. (2016), "Improving Medicare's Hospital Compare Mortality Model", *Health Services Research* **51**, 1229–1247.

Skrondal, A. and Rabe-Hesketh, S. (2009), "Prediction in multilevel generalized linear models", *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**(3), 659–687.

Snijders, T.A.B. and Bosker, R.J. (1999), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, Sage Publications, London.

Solomon, P.J., Kasza, J. and Moran, J.L. (2014), "Identifying unusual performance in Australian and New Zealand intensive care units from 2000 to 2010", *BMC Medical Research Methodology* **14**(53).

Spiegel, A.D. and Springer, C.R. (1997), "Babylonian medicine, managed care and Codex Hammurabi, circa 1700 B.C.", *Journal of Community Health* **22**(1), 69–89.

Spiegelhalter, D.J. (1999), "Surgical audit: statistical lessons from nightingale and codman", *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **162**(1), 45–58.

Spiegelhalter, D.J. (2005), "Funnel plots for comparing institutional performance", *Statistics in Medicine* **24**(8), 1185–1202.

StataCorp (2015), *Stata Statistical Software: Release 14*, StataCorp LP, College Station, TX.

Stow, P.J., Hart, G.K., Higlett, T., George, C., Herkes, R., Mcwilliam, D. and Bellomo, R. (2006), "Development and implementation of a high-quality clinical database: the Australian and New Zealand Intensive Care Society Adult Patient Database", *Journal of Critical Care* **21**(2), 133–141.

Stringhini, S., Carmeli, C., Jokela, M., Avendaiño, M., Muennig, P., Guida, F., Ricceri, F., d'Errico, A., Barros, H., Bochud, M., Chadeau-Hyam, M., Clavel-Chapelon, F., Costa, G., Delpierre, C., Fraga, S., Goldberg, M., Giles, G.G., Krogh, V., Kelly-Irving, M., Layte, R., Lasserre, A.M., Marmot, M.G., Preisig, M., Shipley, M.J., Vollenweider, P., Zins, M., Kawachi, I., Steptoe, A., Mackenbach, J.P., Vineis, P. and Kivimäki, M. (2017), "Socioeconomic status and the 25×25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1.7 million men and women", *Lancet* **389**(10075), 1229–1237.

Tierney, L. and Kadane, J.B. (1986), "Accurate approximations for posterior moments and marginal densities", *Journal of the American Statistical Association* **81**(393), 82–86.

Tobin, M. and Manthous, C. (2017), "Mechanical ventilation", *American Journal of Respiratory and Critical Care Medicine* **196**(2), 3–4.

van der Leeden, R., Meijer, E. and Busing, F.M.T.A. (2008), Resampling Multilevel Models, *in* "Handbook of Multilevel Analysis", Springer, New York, NY, pp. 401–433.

van Smeden, M., de Groot, J.A.H., Moons, K.G.M., Collins, G.S., Altman, D.G., Eijkemans, M.J.C. and Reitsma, J.B. (2016), "No rationale for 1 variable per 10 events criterion for binary logistic regression analysis", *BMC Medical Research Methodology* **16**(163).

Wakefield, J. (2013), *Bayesian and Frequentist Regression Methods*, Springer, New York, NY.

Wickham, H. (2015), *Advanced R*, Chapman & Hall/CRC, Boca Raton, FL.

Wolfram Research Inc. (2017), *Mathematica, Version 11.2*. Champaign, IL.

Zimmerman, J.E., Kramer, A.A. and Knaus, W.A. (2013), "Changes in hospital mortality for United States intensive care unit admissions from 1988 to 2012", *Critical Care* **17**(2), R81.

Zimmerman, J.E., Kramer, A.A., McNair, D.S. and Malila, F.M. (2006), "Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients", *Critical Care Medicine* **34**(5), 1297–1310.