Mathematical Institute

Master Thesis

Statistical Science for the Life and Behavioural Sciences

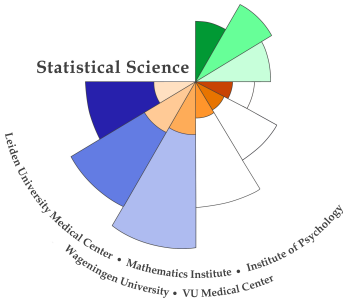
Measuring Hospital Performance: Modelling and Visualising Bivariate Outcomes

Author:
Shane Colm Ó Meachair

Supervisor:
Dr. Marta Fiocco
Leiden Univiersity

January 2016

Universiteit Leiden
The Netherlands



Abstract

Quantitative indicators are used to objectively asses the performance of large institutions such as hospitals. However, different methods for accounting for structural and random variation can lead to different conclusions on relative performance. Using data from 92 hospitals and 24,000 colo-rectal cancer patients we show how ranking methods can give substantially different results depending on the casemix correction used and how individual hospital-level effects are accounted for.

The funnel plot is a visualisation method used for plotting a performance indicator relative to the size or volume of a hospital. We extend the standard funnel plot to three dimensions to display the simultaneous performance of institutions on two outcomes relative to their volume of patients. An interactive 3D funnel plot is implemented and made freely available from the browser.

We then propose a conditional bivariate logistic regression for modelling two particular outcomes relating to colo-rectal cancer surgery; complication and failure to rescue. The model allows the correlation between the two outcomes to be estimated after casemix correction. Finally the standard random effects model is extended by the use of a Bayesian semi-parametric model for modelling non-normally distributed centre effects. We conclude that the choice of how to display and model performance indicators has important policy implications on performance assessment and that scope remains for refinement of such techniques.

Contents

1 Introduction **7**

2 Background and Literature Review **9**
2.1 Introduction . . . 9
2.2 Performance Indicators and Ranking . . . 9
2.3 Ranking . . . 11
2.4 Empirical Bayes . . . 13
2.4.1 Rankability . . . 14
2.5 Multivariate GLMMs . . . 15

3 Data Application **16**
3.1 Introduction . . . 16
3.2 Data source . . . 16
3.3 Descriptive statistics . . . 16

4 Funnel Plots for Institutional Performance Ranking **25**
4.1 Introduction . . . 25
4.2 The two-dimensional funnel plot . . . 25
4.2.1 Overdispersion . . . 27
4.3 Moving to three-dimensions . . . 30
4.4 3D Funnel: A three-dimensional funnel plot . . . 30
4.5 3D Funnel web app . . . 40

5 Joint model for failure to rescue and complications **43**
5.1 Introduction . . . 43
5.2 A conditional multivariate logistic regression . . . 43
5.3 Two-Stage Approach . . . 46
5.4 Estimating standard errors for the two-stage approach . . . 47
5.5 Results . . . 48
5.6 A brief simulation study . . . 49

6 Bayesian semi-parametric random effects **52**
6.1 Introduction . . . 52
6.2 Dirichlet process . . . 52
6.3 Model specification . . . 54
6.4 Results . . . 54
6.4.1 Complications . . . 54
6.4.2 Mortality . . . 58

7 Discussion **61**

8 Appendix: R Code **66**
8.1 R code for Chapter 4 66
8.2 R code for Chapter 5 73
8.3 R code for Chapter 6 81

List of Figures

1 Left panel shows unadjusted counts of post-operative deaths against total number of patients treated in each centre. A LOESS smoother and corresponding interval is shown. Right panel shows histogram of mortality rates, mean values is indicated in red and density estimator overlaid. 17
2 Left panel shows unadjusted counts of post-operative complications against total number of patients treated in each centre. A LOESS smoother and corresponding interval is shown. Right panel shows histogram of complication rates, mean values is indicated in red and density estimator overlaid. 18
3 Scatterplot of mortality rates and complications, with LOESS smoother and estimated interval. 19
4 Hospitals ranked by mortality rates using unadjusted fixed effects model. 22
5 Hospitals ranked by mortality rates using unadjusted random effects model. 23
6 Hospitals ranked by mortality rates using adjusted random effects model. 24
7 Risk-adjusted complication rates following operation for 92 hospitals. The target is the overall complication rate of 23%. 29
8 Risk-adjusted Failure to Rescue (FTR) rates following operation for 92 hospitals. The target is the overall FTR rate of 4.01%. . . . 29
9 Horizontal view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers. The target is 23% complications and 4% mortality rate. 33
10 Vertical view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers. The target is 23% complications and 4% mortality rate. 35
11 Horizontal view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers with 95% limit. The target is 23% complications and 4% for Failure to Rescue. 37
12 Horizontal view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers with 95% limit. The target is 23% complications and 4% for Failure to Rescue. 38
13 Vertical view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers with 98% limit. The target is 23% complications and 4% for Failure to Rescue. 39
14 Screenshots of 3D funnel web app 41
15 Screenshots of 3D funnel web app 42

16 Results from Bootstrap algorithm 1 50
17 Trace and density plots for complication parameters α, p_1, p_N and ρ_1 55
18 Histogram and density plot for hospital complication effects 56
19 Centre complication rate means and smoothed pointwise 95% credible bands 57
20 Trace and density plots for mortality parameters α, p_1, p_N and ρ_1 58
21 Histogram and density plot for hospital random effects mortality rate means 59
22 Centre mortality rate means and smoothed pointwise 95% credible intervals 60

List of Tables

1 Descriptive statistics of casemix variables 20
2 Results from the two-stage bootstrap. Parameters associated with complication are indicated by (2) 49
3 Results from two simulations to estimate ρ 51

1 Introduction

Cancer of the colon and rectum is the second most frequently encountered cancer in Europe. Colorectal cancer remains the second leading cause of European cancer-related deaths. In the Netherlands, approximately 10,000 surgical resections are performed each year to remove parts of the colon where serious cancerous lesions are situated. Of these procedures, approximately 7% will result in death (Henneman et al. 2013). In recent years a trend has emerged internationally toward outcomes-monitoring in public institutions and data-driven decision making at a policy level as well as information dissemination to the public. In principal, this is so that informed decisions can be made at an individual level regarding important choices between treatment options and where to receive them.

It is of great importance for decision makers, policy makers, clinicians, health professionals and the public to understand the relation between procedure-related mortality and complications, and how these are related to individual patient characteristics as well as hospital-related factors. In the Netherlands, the Dutch Surgical Colorectal Audit (DSCA) was established in 2006 to record data on the results of treatment for colorectal pathologies across 92 primary Dutch hospitals (*Dutch Colorectal Cancer Group* 2015).

Two issues that have been examined in the research emerging from the data collected by the DSCA have been surgical complications and failure to rescue (FTR). FTR refers to deaths arising from life-threatening surgical complications which can not be resolved by clinicians' efforts. To date research has looked at these factors in isolation (Henneman et al. 2013), however the relation between these two outcomes has not been examined. The aim of this thesis is to explore this relation and to develop the statistical methodology necessary to do so. This research is also situated within the context of measuring relative hospital performance in a statistically sound manner.

Chapter 2 will give an introduction to the use of performance indicators and hospital rankings. This sets the background for the applied context in which this work takes place. Chapter 3 describes the data that motivates this work. Chapter 4 deals with methods for visualising and exploring performance indicator data. The funnel plot is a graphical device particularly suited to the display of a univariate performance indicator. A novel extension of the funnel plot to bivariate indicators is introduced, the 3D funnel plot. Chapter 5 looks at modelling the relationship between complication and FTR and a set of important clinical variables. Particular interest lies in correctly accounting for the correlation between the two outcomes and also the hierarchical structure of the data.

Chapter 6 outlines an alternative framework for random effects modelling using a semi-parametric Bayesian method, and offers suggestions on how the previous work may be extended using this paradigm. Chapter 7 offers some discussion and conclusion from the research. The R code used for generating the 3D funnel plots and statistical models are included in Appendix A.

2 Background and Literature Review

2.1 Introduction

This thesis draws on topics from likelihood inference for Generalised Linear Mixed Models, multivariate models, visualisation and Performance Indicators and Ranking.

2.2 Performance Indicators and Ranking

The presentation of data on hospital outcomes in an informative way poses a number of issues. In the popular work on decision making and cognitive biases by Nobel-winning psychologist Daniel Kahneman, the central conclusion of his research is that humans, by nature, are poor statisticians. In particular he has noted how domain experts, such as clinicians, are no exception to this rule despite the fact that many of their judgements are often based on unacknowledged statistical information and statistical reasoning. Acknowledging underlying baseline rates, correcting for influential variables and correctly accounting for uncertainty are fundamental statistical concepts for which the human brain has not developed a functional intuition (Kahneman 2011). With this knowledge in mind, there is an onus on statisticians to present data in such a way that is as digestible and informative as possible. This is particularly the case with data on healthcare, which is an especially emotive issue for most people. The public places a large degree of faith and trust in clinicians and health institutions to acquire appropriate knowledge, to act skilfully and make the right decisions for them. The presentation of data on clinical outcomes must be carried out in a cautious and thoughtful way which avoids spurious inferences. In the healthcare there has been increase in public ranking of health institutions in 'league-tables'. However, to date, such ranking endeavours have often been carried out in a naïve ad hoc manner without sufficient adherence to statistical rigour. Recent high-profile controversy in the United Kingdom shows the perils of lax statistical methodology in such situations, which can be detrimental to the careers of healthcare providers and which are highly emotive issues among the public (*Medics' NHS league table mortality figures mired in confusion* 28/06/2013).

A performance indicator is defined by Goldstein and Spiegelhalter (1996) as some statistical quantity which is used to express the quality of the subject under investigation. They note that the choice of a certain indicator must be thoroughly considered so that the variability of the indicator is strongly correlated with the underlying quality being measured. Two essential qualities required of the input data are appropriateness and integrity. Appropriateness involves some subjec-

tive judgement, and the extent to which a certain indicator is appropriate may be cause for debate. Integrity encompasses completeness and correctness, which while also up for discussion, can be more objectively assessed. Decisions about which data to use, what to exclude, what transformations to perform etc., can be explicitly stated.

Below, we will give an example of a performance indicator which will be used later in this thesis, the complication rate during colorectal resection. The raw data used is the number of complications occurring in each hospital. We make adjustment for the size of the centre by dividing by the number of patients undergoing surgery. The performance indicator is defined as the proportion of patients undergoing this specific type of surgery who suffer a complication. The complication indicator in an individual centre (hospital) i is given by

$$p_i = \frac{x_i}{n_i}.$$

While the overall complication rate can be calculated as

$$p = \frac{\Sigma_{i=1}^{N} x_i}{\Sigma_{i=1}^{N} n_i} = \frac{\#complications}{\#patients},$$

where N is the total number of centres. A useful way to check if a single centre deviates extremely from the overall mean is to define an interval around the grand mean which is a function of the size of a given centre. Making the assumption that complications follow a binomial distribution with parameters p and n_i, the approximate 95% interval is calculated by

$$p - 1.96\sqrt{\frac{p(1-p)}{n_i}} \leq p_i \leq p + 1.96\sqrt{\frac{p(1-p)}{n_i}}.$$

Such adjustment is not sufficient however. Centres with smaller numbers of patients undergoing surgery will have a greater degree of uncertainty as to the 'true' rate of complication. Uncertainty and its quantification are at the core of statistical inference, whether frequentist or Bayesian. However, when statistical summaries make the transition from the statistician's analysis to the end user, measurements of uncertainty are often discarded and point estimates are accepted as truth. Interval estimation around any point estimates is crucial to convey the uncertainty in numbers estimated from imperfect data, sampled imperfectly from the population of interest. In addition the real interest will lie in estimates after adjusting for the casemix of the hospital cohorts, to allow for fair comparison, and also highlight controllable factors and areas for intervention.

2.3 Ranking

When performance indicators are used to produce league tables, it is known that the rankings of such tables are highly sensitive to sampling variability and that there has been no straightforward way to place intervals around those ranks. Naïve ranking is an inherently fraught exercise, as even if there is no statistically real difference between centres (i.e. the differences result only from random variation) and rank order can still be derived where one centre is the best and another is the worst. The league table approach will give the illusion of significant differences between the best and worst.

The general ranking problem can be describe as follows: N centres need to be ranked based on the values of a set on unknown parameters called θ_i, $i = 1, \ldots, I$. Each θ_i could represent for example a centre, clinician, or any unit of interest to be ranked, and R_i denotes the rank of θ_i. Laird and Louis (1989) define the simple ranking function:

$$R_i = rank(\theta_i) = \sum_{j=1}^{N} I(\theta_i \leq \theta_i) = \sum_{j=1}^{N} I_{ji}$$

where $I(\cdot)$ is the indicator function and $I_{ji} = I(\theta_i \leq \theta_j)$.

The expected rank is given by

$$E(R_i) = \Sigma_{j=1}^{N} P_{ji}$$

where $P_{ii} = 1$ and $P_{ji} = P(\theta_i \leq \theta_j)$ for all $j \neq i$.

Assuming the performance indicators to be independently normally distributed with mean μ_i and variance τ_i, the probability of each rank is given by

$$P_{ji} = P(\theta_i - \theta_j \leq 0) = \Phi[(\mu_i - \mu_j)\sqrt{\tau_i^2 + \tau_j^2}].$$

Note that if the standardised differences $(\mu_i - \mu_j)/\sqrt{\tau_i^2 - \tau_j^2}$ are very close to 0 the expected ranks are all close to the mid-rank $(n + 1)/2$.

Spearman's rank correlation coefficient is a modification of the Pearson correlation coefficient where the raw data of two random variables X_i and Y_i are converted to sets of ranks x_i and y_i. It measures the degree of similarity between two vectors of ranks. The correlation ρ is calculated by

$$\rho = \frac{\Sigma_i (x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\Sigma_i (x_i - \hat{x})^2 \Sigma_i (y_i - \hat{y})^2}}.$$

An alternative measure is τ (Kendall 1938) which computes the number of pairs from two sets of ranks whose relative ordering in the two rank sets agrees. Let π_1 and π_2 be two vectors of ranks. τ is computed using a heaveside function $H(x)$ such that

$$H(x) = \begin{cases} 0, \text{if x} < 0 \\ 1, \text{if x} \geq 0. \end{cases}$$

The term τ_{ij} is thus defined as

$$\tau_{ij} = H((\pi_1(i) - \pi_1(j))(\pi_2(i) - \pi_2(j)))$$

where $\tau_{ij} = 1$ when the relative ranks agree and 0 when they disagree. Kendall's τ for the pair of full rank vectors is

$$\tau(\pi_1, \pi_2) = \frac{1}{\binom{N}{2}} \Sigma_{i<j} \tau_{ij}(\pi_1, \pi_2).$$

Kendall's τ can be interpreted as the probability that the two rankings agree on a given pair of ranks. Zuk, Ein-Dor, and Domany (2007) test the reliability of such a measure under the assumption that the noise in observed rankings is Gaussian. The 'noisy' observed ranking is defined as $s = r + Z$ where r is the true ranking and $Z_i \sim N(0, \sigma^2)$. An estimate of the noise could be derived for example by bootstrapping the entire data set in one subsampling, or by centre, and computing the centre ranking in each sample and comparing them using Kendall's τ.

One possible approach is using random sampling such as bootstrapping or Monte Carlo methods to produce many datasets with random variation which should follow the same sampling distribution as the dataset at hand. The variability in the rankings can then be quantified. Goldstein and Spiegelhalter (1996) outline two procedures for producing confidence intervals. One, with an interval around a mean for each institutions and a test based on intervals not overlapping. The alternative is to provide intervals around ranks, which they say is more interpretable to practitioners.

Adjustment for initial state or case mix is another consideration when comparing institutions. In general, some adjustment to the raw numbers should be made, with careful consideration given to the degree of stratification. One adjustment which is particularly suited to institutional comparisons is for the size of the institution so that larger and smaller hospitals can be assessed simultaneously, as we used for the complication performance indicator.

The authors emphasise caution in any rankings system, recommending that they should be used as suggestions rather than definitive placings. In addition, many models may fit the data equally well yet give different institutional rankings. They conclude that any rankings or league tables must be provided with an official warning as to their use. Despite such warnings, public demand for ranking systems insists that they will be produced, statistically correct or not.

2.4 Empirical Bayes

Some attempt at producing league tables which are as good as they can be is therefore required. Laird and Louis (1989) proposed such a method based upon Empirical Bayes estimation. Often referred to as 'Bayes light', Empirical Bayes utilises the techniques of Bayesian inference, but it is not fully Bayesian as no external prior knowledge is introduced into the analysis. Rather the priors are estimated from the data itself. The empirical Bayes approach considers the centre effects θ_k to be random with distribution G. The simplest model is $G \sim N(\mu, \tau^2)$. To estimate μ and τ^2 the same approach as DerSimonian and Laird (1986) can be used.

We use a normal-normal model for inference on θ:

$$\hat{\theta}_i \sim N(\theta_i, \sigma_i^2); \theta_i \sim N(\mu, \tau^2).$$

Inference about centre effects is based on the posterior distribution of θ_i

$$\theta_i | \hat{\theta}_i \sim N(pm_i, pv_i), \quad i = 1, \ldots, N,$$

where

$$pm_i = \mu + \frac{\tau^2}{\tau^2 + \sigma_i^2}(\hat{\theta}_i - \mu)$$

and the posterior variance is given as:

$$pv_i = \frac{\tau^2 \sigma_i^2}{\tau^2 + \sigma_i^2}.$$

Here σ_i^2 represents the variance of the measurement error.

To estimate the posterior mean as an estimate of θ_i, the measurement error variance σ^2, μ and τ^2 need to be estimated.

The posterior mean is also called the empirical Bayes estimate (EBE) of the centre effect. If $\tau^2 = 0$, $EBE_i = \mu$. Note that the the posterior mean pm_i is a weighted average of τ and $\hat{\theta}_i$. So the posterior mean is closer to μ than $\hat{\theta}_i$ and it

is shrunken toward μ. This phenomenon is called shrinkage and the shrinkage is more severe if σ_i^2 is large. That means that shrinkage is more severe for smaller centres.

Laird and Louis firstly note the importance of correcting for characteristic (case mix) variables in a school-comparison setting. In that article the authors emphasises the positive aspects of ranks, such as identifying subjects for further case studies. For example the highest or lowest performing centres can be further studied to compare influential features, or clusters can be examined. Under the Empirical Bayes assumption, the log odds of an event, such as complication, occurring are distributed normally with mean μ and variance τ. Then μ and τ can be estimated directly from the data. The use of Empirical Bayes methods allow shrinking of estimates toward an overall mean while increasing the posterior variance. The shrinkage effect has been found in practice to lead to better prediction accuracy when using statistical inference for prediction. The adjusted variance also gives a better representation of the uncertainty in estimated ranks.

2.4.1 Rankability

Laird and Louis (1989) introduce the concept of Expected Rank (ER) as a measure of relative performance which accounts for the variability in the estimates used to order alternatives. The ER is defined as

$$ER_i = 1 + \sum_{j \neq i} \Phi((EB_i - EB_j)/\sqrt{pv_i + pv_j}).$$

Houwelingen, Brand, and Louis (2004) notes that if the posterior variances are large the ER's will all be around 50%, in other words the ranking signal will be low relative to the noisy data. The expected ranks can be turned into percentiles;

$$PCER_i = \frac{100(ER_i - 0.5)}{n}.$$

This can be interpreted as the probability of a randomly chosen centre j is better than centre i.

An overall assessment of the rankability of a modelling procedure on a given data set can be estimated from the following formula:

$$\rho = \frac{\tau^2}{\tau^2 + median(s_i^2)}.$$

The resulting percentage gives an estimate of the amount of variation in ranks explained by the fixed effects estimates as opposed to random variation.

2.5 Multivariate GLMMs

McCulloch, Searle, and Neuhaus (2011) give a comprehensive treatment of GLMMs and estimation methods. Berridge and Crouchley (2011) provide a formulation for multivariate GLMMs and implemented an accompanying R package which used Gaussian quadrature to estimate the models. However, this package has subsequently been removed from CRAN. Hadfield (2010) use MCMC for estimation and inference of multi-response GLMMs, while Gueorguieva (2001) derive an EM algorithm for estimating the model parameters. These topics will be expanded upon later in Chapter 5.

3 Data Application

3.1 Introduction

This chapter will summarise the particular dataset that is used in this thesis and which motivates some of the methods. Medical context and descriptive statistics will be explored and its relevance to policy and practice.

3.2 Data source

The data were obtained from the Dutch Surgical Colorectal Audit (DSCA), a continuous national quality improvement project which maintains a registry of data on variables concerning patient characteristics, disease, diagnostics, treatment, and outcomes collected on a prospective basis (Henneman et al. 2013). The registry contains data from all 92 Dutch hospitals which perform colorectal cancer surgery in the Netherlands. The data has a case-ascertainment rate of approximately 95% compared to the Netherlands Cancer registry. Patient and hospital data are anonymised to remove any potential for bias during the analysis. Records from 24,667 patient surgeries were collected across hospitals over the period of 2009-2011.

3.3 Descriptive statistics

The average complication rate across all patients was 22.9% and post operative mortality was 4%. One centre had no deaths while the maximum observed mortality rate was 9%. The minimum observed complication rate was 3% while the highest rate was 43%. Figures 1, left panel, shows the relationship between hospital size, defined as number of patients undergoing surgery and the raw number of deaths in that centre. As might be expected a priori, the number of incidents increases with size. A LOESS smoother is overlaid on the graph and local confidence intervals drive from t-distributions are fitted along the curve. It can be seen that many centres have counts outside of this crude interval region. The right panel shows a histogram of mortality rates, defined as the number of deaths divided by number of patients treated, with the mean indicated as a vertical red line. Considerable variability can be seen in the distribution of these rates as well as a possible multimodal shape.

Figures 2 shows a similar display for complications, again some multimodality may be observed in the distribution of post-operative complication rates.

Figure 1: Left panel shows unadjusted counts of post-operative deaths against total number of patients treated in each centre. A LOESS smoother and corresponding interval is shown. Right panel shows histogram of mortality rates, mean values is indicated in red and density estimator overlaid.

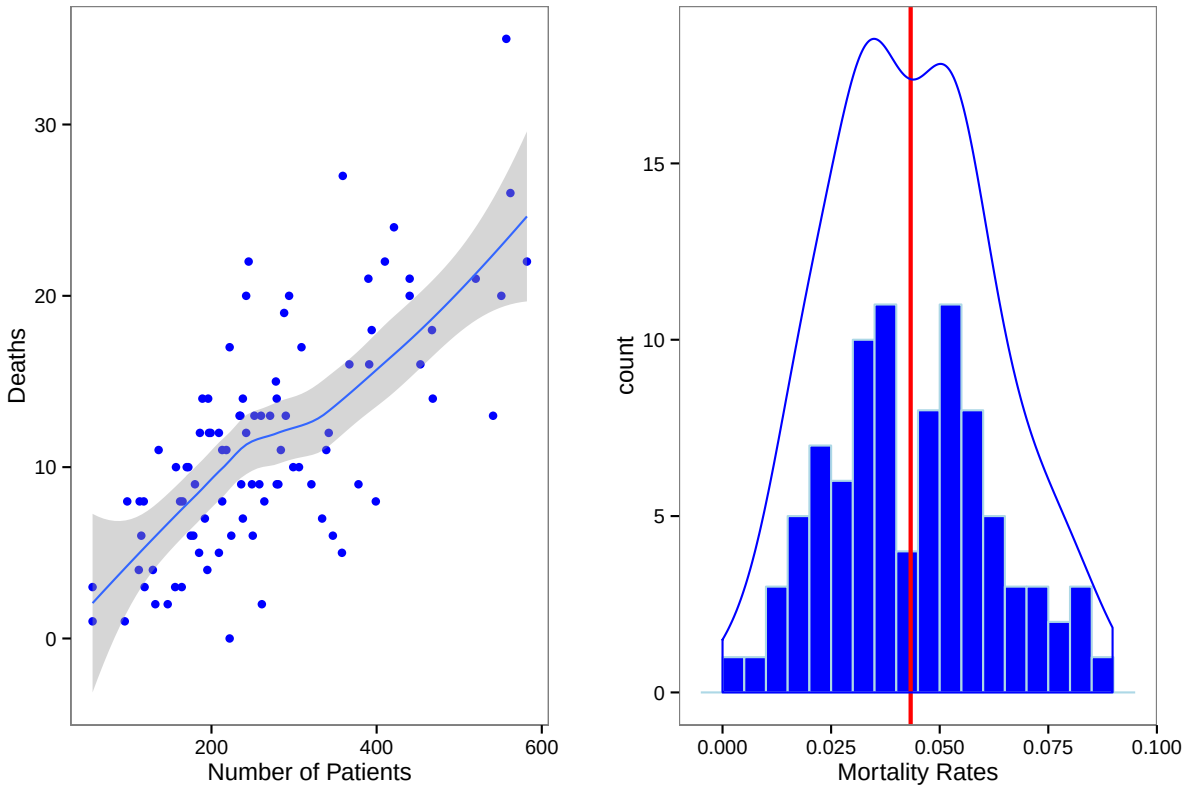
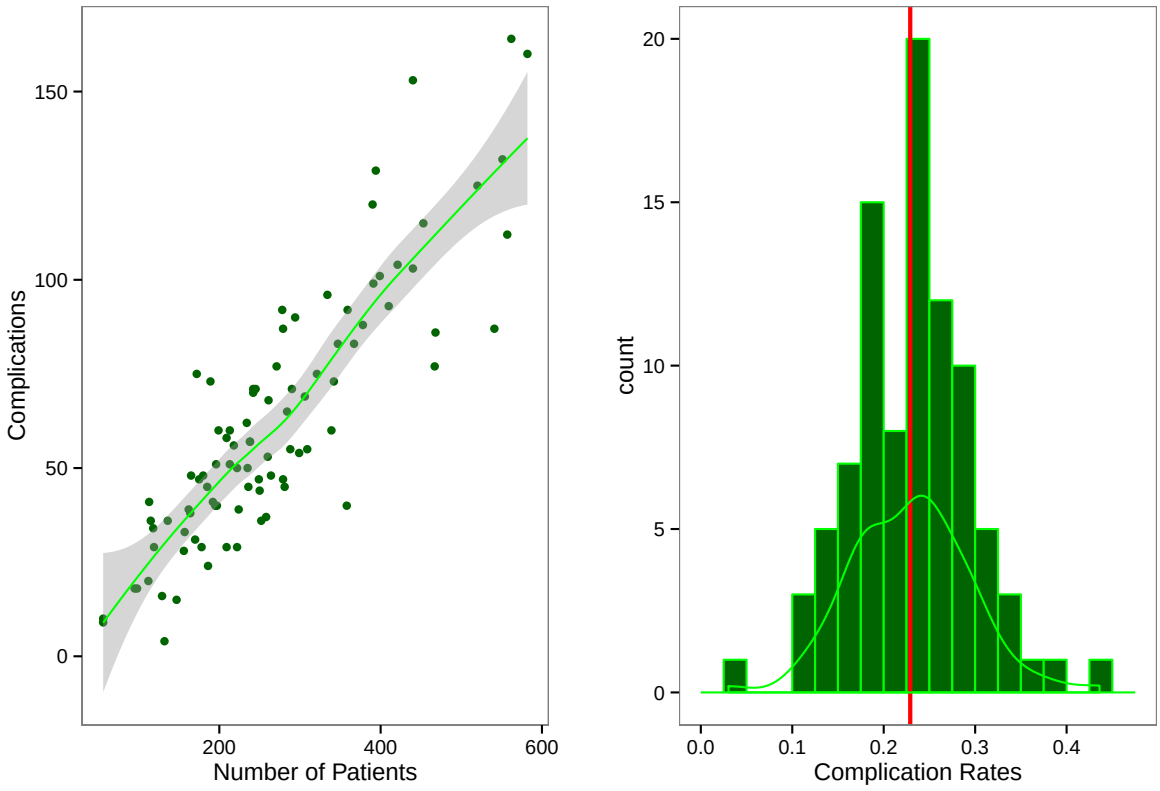


Figure 3 shows the relationship between unadjusted mortality and complication rates. A LOESS smoother is applied with a span of $\frac{2}{3}$. An overall positive relationship can be seen between the two outcome measures.The pearson correlation coefficient is 0.46.

Table 1 shows a summary of the casemix variables. Variables were chosen during discussions with a colorectal cancer surgeon who collaborated on assembling the data. Almost 37% of patients were over 75 years of age and thus at higher risk of complication or death. The Charlson index is a measure of co-morbidities used to predict the ten-year mortality of a given patient. A score of 1 is an indicator of co-morbidities with reasonable survival prognosis such as heart disease, ulcers, dementia, or chronic liver disease. Scores of 2 or above indicate more serious ailments such as leukaemia, sever liver disease, malignant tumours

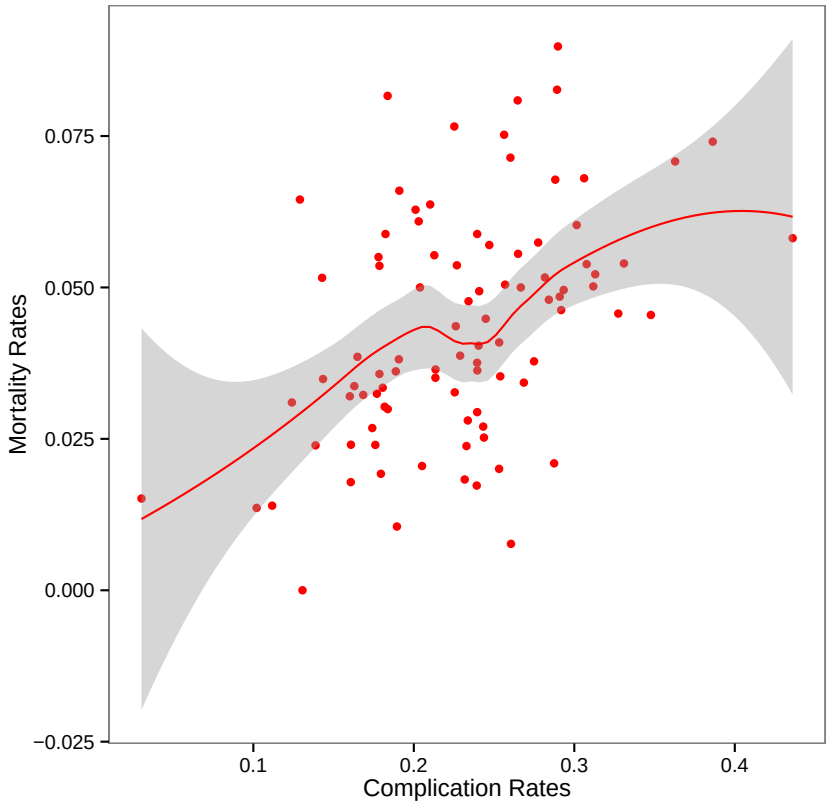
Figure 2: Left panel shows unadjusted counts of post-operative complications against total number of patients treated in each centre. A LOESS smoother and corresponding interval is shown. Right panel shows histogram of complication rates, mean values is indicated in red and density estimator overlaid.



or metastatic cancer. The ASA (American Society of Anaesthesiologists) score is an indicator of a patient's overall state of health prior to undergoing surgery. Scores are ascribed as below (Daabiss 2011):

1. Patient is a completely healthy fit patient.
2. Patient has mild systemic disease.
3. Patient has severe systemic disease that is not incapacitating.
4. Patient has incapacitating disease that is a constant threat to life.
5. A moribund patient who is not expected to live more than 24 hours with or without surgery.

Figure 3: Scatterplot of mortality rates and complications, with LOESS smoother and estimated interval.



The stage variable indicates cancer staging using the pathological TNM stage, aggregating T stage (0-4), N stage (0-2) and M stage (0-1) into stages 0 through 4. The majority (66%) of cancers are in the mid stages of 2-3. Preoperative complications were present in 22.9% of surgeries while the majority of surgeries (85%) were elective rather than urgent.

Table 1: Descriptive statistics of casemix variables

Variable	**Levels**	**n**	%
Gender	Female	13513	54.8
	Male	11154	45.2
	All	24667	100.0
Age Group	¡75	15405	62.5
	75+	9236	37.5
	All	24641	100.0
BMI	Unknown	8019	32.5
	¡25	6616	26.8
	25 - 29	6884	27.9
	30+	3148	12.8
	All	24667	100.0
Charlson index	0 - 1	19077	77.3
	2+	5590	22.7
	All	24667	100.0
ASA score	1	4997	20.8
	2	13444	55.9
	3	5164	21.4
	4	468	1.9
	All	24073	100.0
Resection	None	22304	90.4
	Extensive	1083	4.4
	Limited	1280	5.2
	All	24667	100.0
Pre-op Complication	None	19019	77.1
	Present	5648	22.9
	All	24667	100.0
Urgency	Elective	20865	84.8
	Urgent	3741	15.2
	All	24606	100.0

Using the methodology described in Section 2.3, the hospitals can be ranked according to their proportion of deaths or complications. In Figure 4, centre effects are fitted as fixed effects and no covariates are included, 95% intervals are shown. Figure 5 shows the rankings using a random effects model for centres. The shrinkage effect of the random effects model is clear, and the 95% credible (or probability) intervals for extreme centres are narrowed considerably showing the benefit of the random effects approach in this instance. Figure 6 shows the random effects model after casemix adjustment. Now many of the centres are even closer to each other and differences more slight, however two centres can be seen to do noticeably worse than the others.

The remainder of this thesis will look at visualising and quantifying the simultaneous performance of hospitals by mortality, complications and failure to rescue (death resulting from operative complications), while accounting for the casemix variables described above as well estimating the correlation between the two outcomes across centres.

Figure 4: Hospitals ranked by mortality rates using unadjusted fixed effects model.

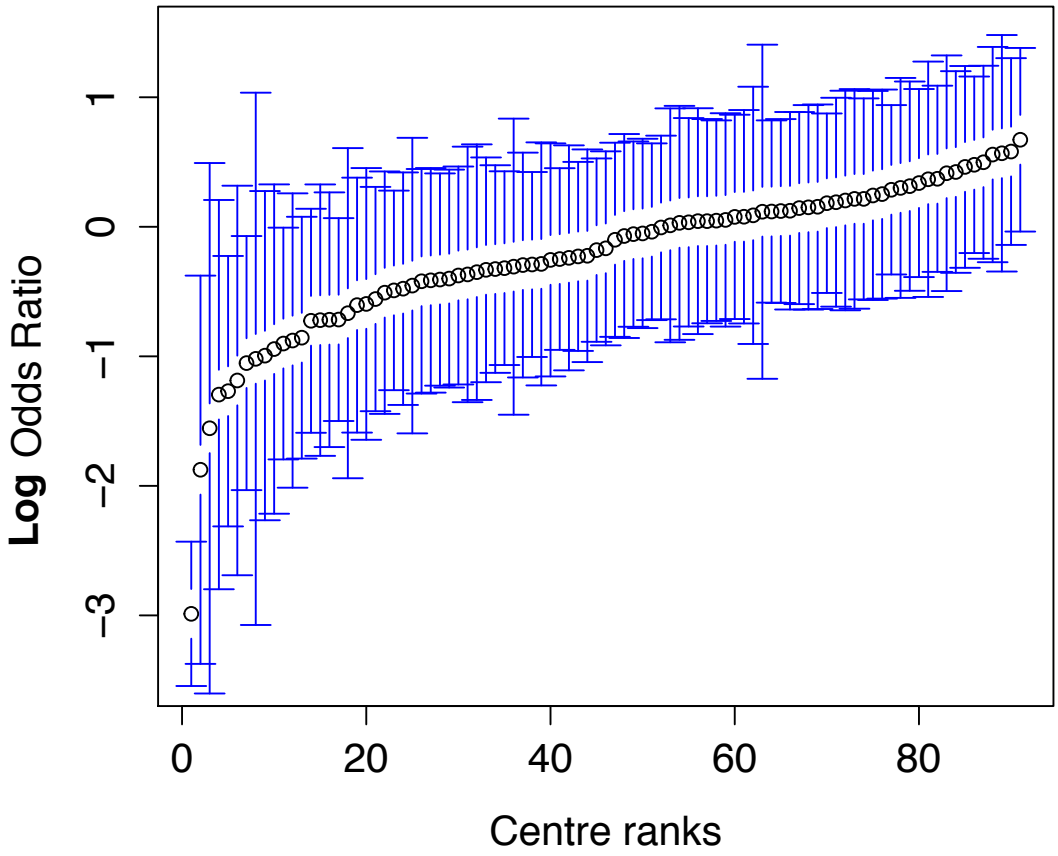


Figure 5: Hospitals ranked by mortality rates using unadjusted random effects model.

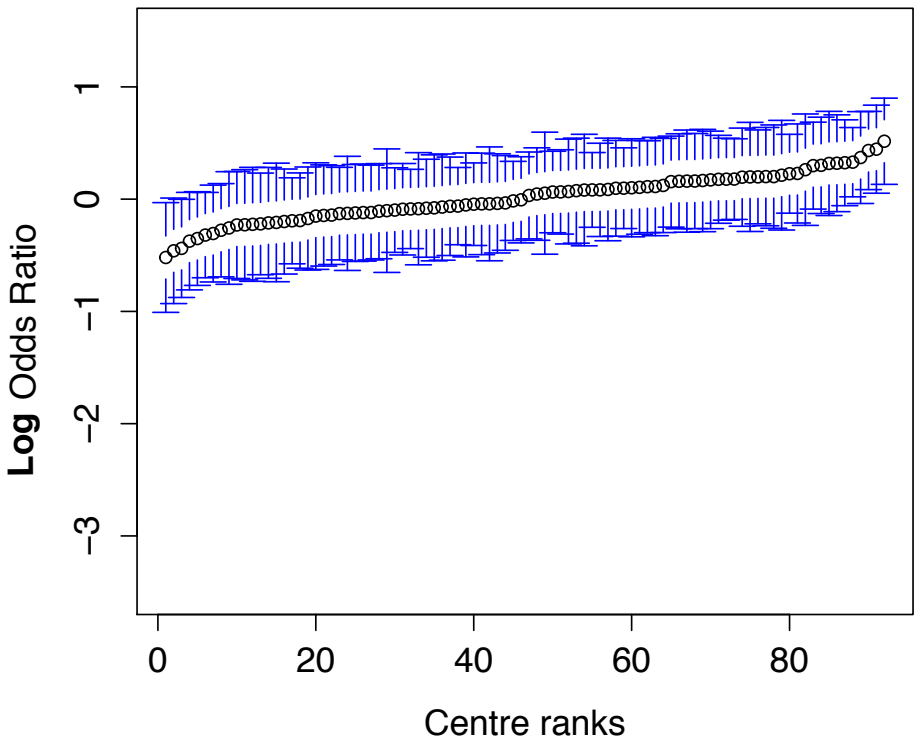
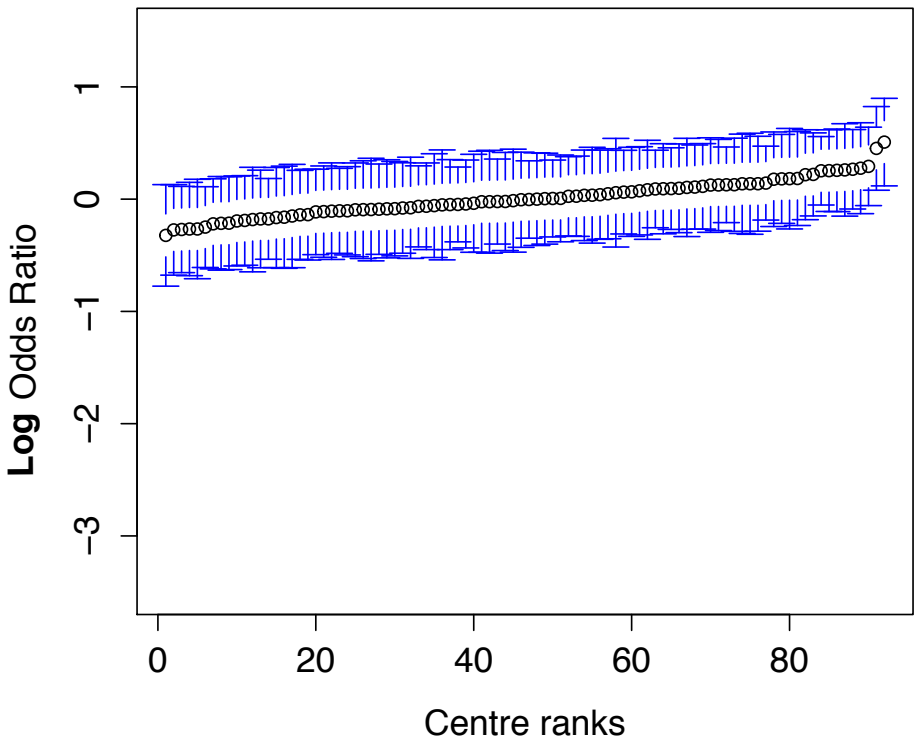


Figure 6: Hospitals ranked by mortality rates using adjusted random effects model.



4 Funnel Plots for Institutional Performance Ranking

4.1 Introduction

This chapter introduces the funnel plot as used for visualising the relative performance of units such as hospitals, schools, industrial installations etc. Similar plots are more commonly used for assessing publication bias in meta-analyses. Funnel plots allow for the simultaneous display of the relation between a primary outcome variable and some measure of volume such as patients within a hospital, along with confidence limits around those measurements. The chapter first looks at the standard two-dimensional funnel plot for one outcome variable. Methods for calculating confidence limits are described and compared. As the primary methodological interest of this paper involves bivariate measure, we move on to introduce a novel three dimensional funnel plot.

4.2 The two-dimensional funnel plot

The funnel plot was originally developed as a tool for visualising the results of meta-analyses to visualise the extent of publication bias in the component studies (Light and Pillemer 1984). Spiegelhalter (2005) modified and extended the funnel plot for use in institutional performance ranking, as a form of 'control plot'. Control plots specify a control region in which institution are assumed to be subject to in-control variation which is common to all institutions. In the language of random-effects models, this can be referred to as the intra-institution random effect, or the institution component of variance. Institutions which fall outside of the control region can be examined for unique causes of their outlier status. Commonly a threshold of 3 standard deviations is used to specify the control region.

The funnel plot extends the notion of the control-region to incorporate a measure of precision for each institution or observational unit. The control region then varies with the precision of the estimate, which, when all institutions and their control region are plotted simultaneously, forms a funnel shape as the precision increases. Imprecise estimates (possibly due to small sample sizes) will have a wider control region which takes into account the greater degree of uncertainty in the estimate. In the performance indicator setting, if all institutions fall within the control region, there may be no basis for ranking institutions at all. This is because differences in performance are assumed to be as result of the common-cause variation across institutions (i.e. the institution random effect)

and not institute-specific factors.

A funnel plot has 4 components listed here below:

1. An indicator variable X,
2. A target variable T_0. This is the expectation given that the institution is in the control region so that $E(X|T_0) = T_0$,
3. A precision parameter α. For an institution in the control region, the null distribution for X is equivalent to the joint distribution of α and X, ie. $p(X|T_0, \alpha)$. A common form for α is to take it as proportional to the inverse variance of X, so that for some function g,

$$\alpha = \frac{g(T_0)}{V(X|T_0)}. \tag{1}$$

The choice of α is quite flexible and should be chosen to have some direct interpretation. In the present context, the indicators are discrete incidence rates within hospitals. We standardise these indicators by dividing by the number of patients being treated so the indicators are proportions with expectation T_0 and variance $V = T_0(1 - T_0)/n$. Using formula (1) and setting

$$g(x) = T_0(1 - T_0),$$

an interpretable measure of the precision is the sample size and therefore $\alpha = n$. Control limits, $X_p(T_0, \alpha)$, which specify the boundary of the control region, are based on a p-value p. The p-value specifies the probability p of an institution which is in control exceeding the control limits. Typically the p-value is chosen to correspond to values exceeding 2 or 3 standard deviations from the target T_0.

The funnel plot draws the target, T_0, as a straight line and the upper and lower control limits are plotted as a function of α, which form curved lines either side of T_0. For each institution the pair (x_i, p_i) are plotted as coordinates on the 2-dimensional Cartesian plane, where x_i is the hospital volume and y_i is the complication rate.

For most applications with no further knowledge of the distribution of the indicator variables and approximate normal distribution is assumed so that

$$X|T_0, \alpha \sim N(T_0, g(T_0)/\alpha).$$

The control limits are plotted at

$$y_p(T_0, \alpha) = T_0 \pm z_p \sqrt{\frac{g(T_0)}{\alpha}},$$

where z_p is the quantile value corresponding to $\phi(z_p) = P(Z \leq z_p) = p$ and ϕ is the CDF of the standard normal distribution. Thus z_i is equivalent to the standardised Pearson residual

$$z_i = \frac{y_i - T_0}{\sqrt{V(X|T_0)}}, \tag{2}$$

referred to as the naive z-score as there is no correction for other explanatory variables. The funnel plot can easily be extended to visualise risk-adjusted data.

4.2.1 Overdispersion

The funnel plot as explained above is not limited to the implicit assumption that the variability of the in-control observations is fully captured in the null distribution. The schema can be extended to account for overdispersion, where actual 'in-control' units exceed the control limits, due to other factors.

Spiegelhalter (2005) outlines two approaches to accounting for overdispersion, the first using the standard overdispersion parameter familiar from generalised linear models, and a second 'random-effects' approach. Under the first approach, a multiplicative overdispersion parameter, ϕ, is used to inflate the null variance to a degree proportional to the mean, leading to

$$V(X|T_0, \alpha, \phi) = \phi V_0(X|t_0, \alpha) = \frac{\phi g(T_0)}{\alpha}.$$

The overdispersion parameter can be estimated from the following formula where I is the sample size of in-control units;

$$\hat{\phi} = \frac{1}{I}\Sigma_i \frac{(x_i - T_0)^2 \alpha_i}{g(T_0)} = \frac{1}{I}\Sigma_i z_i^2,$$

where z_i has been defined in (2).

The uncorrected control limits are then multiplied by a factor of $\sqrt{\hat{\phi}}$ so that the new control limits become

$$x_p(T_0, \alpha) = T_0 + z_p\sqrt{\hat{\phi} g(T_0)/\alpha}.$$

when the approximation with a normal control limits is used.

A number of methods are available for identifying in-control observations to plug in to this formulation.

The second approach to modelling possible over dispersion uses a method of moments estimator is

$$\hat{\tau}^2 = \frac{I\hat{\phi} - (I-1)}{\Sigma_i w_i - \Sigma_k w_i^2 / \Sigma_i w_i},$$

with $w_i = 1/s_j^2$ (DerSimonian and Laird 1986). Here, $\hat{\phi}$ is a test for homogeneity, so that if $\hat{\phi} < (I-1)/I$, then $\hat{\tau}^2 = 0$. The control limits are then plotted at

$$T_0 \pm z_p \sqrt{V(X|T_0, \alpha) + \tau^2}.$$

This form of over-dispersion modelling is used by the Healthcare Commission for England and Wales. Spiegelhalter warns that such statistical modelling of over-dispersion should be seen only as a temporary fix. Overdispersion should thus be considered as a sign that other variables should be included to account for the unexplained excess variability.

Funnel plots for complications and Failure to Rescue are shown in Figures 7 and 8 respectively. The funnel plot in Figure 7 identifies high complication low-volume hospital, as well as borderline low-complication high-volume hospital. There is a suggestion of an outcome volume association but this is beyond the scope of this thesis. The majority of hospitals follow the pattern delineated by the funnel but there are 21 who are outside the lower 95% limit and 17 who are above the upper 95% limit. Centres which are outside the funnel are on the upper limit should be inspected more closely to detect reasons for elevated rates. Centres falling below the lower limit may also provide evidence of factors which lead to lower incidence rates.

Figure 7: Risk-adjusted complication rates following operation for 92 hospitals. The target is the overall complication rate of 23%.

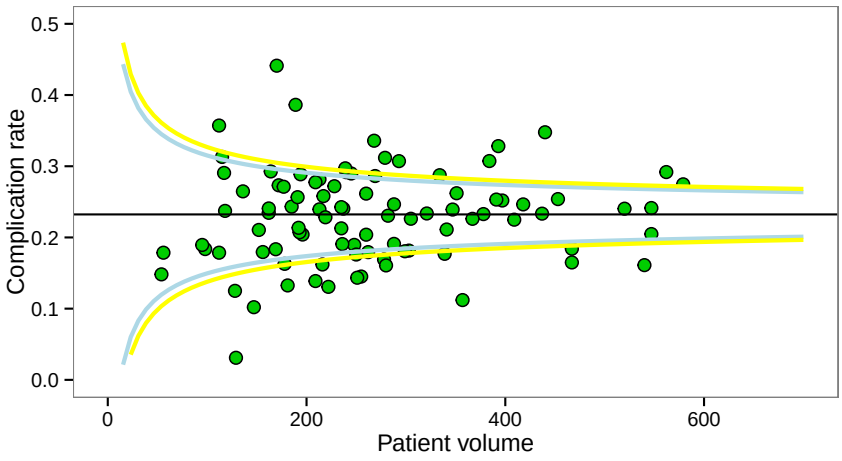
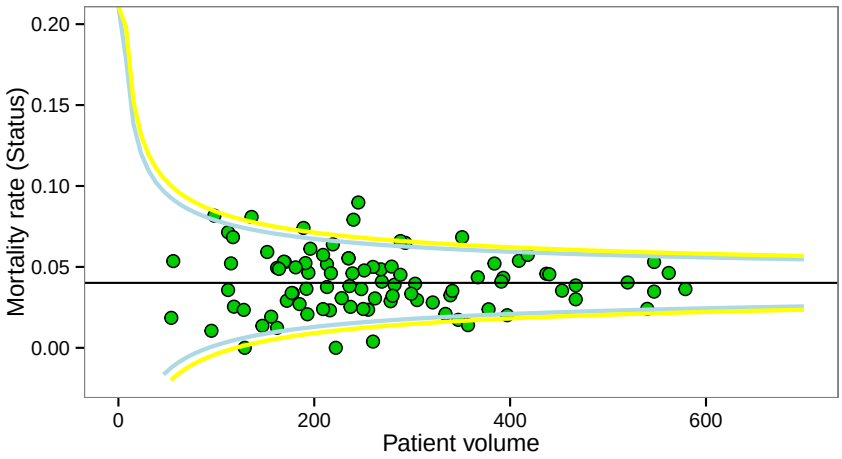


Figure 8: Risk-adjusted Failure to Rescue (FTR) rates following operation for 92 hospitals. The target is the overall FTR rate of 4.01%.



4.3 Moving to three-dimensions

In many applied analyses of medical data, multiple outcomes are of interest. This could be a clinical trial which has multiple primary outcomes such as survival and disease progression. In practice, often separate analyses are run which do not take into account the correlation between the outcome variables. Similarly in the performance indicator setting, it is typical that indicators are analysed and visualised separately. Separate two-dimensional funnel plots display the relationship between the indicator variable and the precision variable, but cannot easily convey the joint relationship between the two outcome variables. Where two outcomes are of interest, we have a unique opportunity to model the joint outcomes using a two-dimensional representation of three-dimensional space.

Three-dimensional data visualisation is most often used to show the relationship between variables in a regression context. For example, in a multiple linear regression of two explanatory or independent variables $X = X_1, X_2$ on an independent variable Y. So far there has been no attempt in the literature to extrapolate the funnel plot to three-dimensions in order to display a bivariate outcome set along with the control-regions of the funnel plot.

As will be shown below, a three-dimensional funnel plot allows the user to view a large amount of information simultaneously, and to get an intuitive sense for the data and the relationships between variables. In addition, with the use of interactive graphical devices that are freely available in R, 3D funnel plots can be viewed in internet browsers or as applets for mobile phones or desktop computers.

4.4 3D Funnel: A three-dimensional funnel plot

The control region in the standard funnel plot is directly analogous to a confidence interval and indeed is constructed in a similar fashion. Where a confidence interval is required for two random variables, rather than draw separate confidence intervals, it is often preferable to define a joint $100(1-\alpha)\%$ region for the two variables. This can be visualised using an elliptical region, or confidence ellipse. Batschelet (1981) distinguished between a standard ellipse and a confidence ellipse, sometimes also referred to as Hotelling's confidence region. The standard ellipse is used as a data descriptive tool, while the confidence ellipse is a confidence region for a bivariate point estimate, such as the means of a bivariate jointly distributed random variable. Batschelet notes that the former does not depend on the sample size yet the latter does. In the present case, the ellipse is a data descriptive tool but is dependent on sample size. Shewhart (1931) also made

reference to a 'control ellipse' for monitoring quality of manufactured products though subsequently this does not seem to have been widely used (Alexandersson 2004).

The 3D funnel plot is composed as follows. For each level of the precision variable (in the present case, N, the number of patients in each hospital) a control ellipse is drawn for the two indicator variables. When these regions are plotted simultaneously in a three dimensional region, a 3D funnel emerges around the scatter plot of performance indicators.

The centre of each control ellipse is given by the Cartesian coordinates of the variable means (μ_1, μ_2). Batschelet refers to this point as the centre of the sample or the centre of its mass or gravity. For our example these are the mean proportions of Complications and Deaths related to surgery.

The equation of the standard ellipse based on that given by Batschelet is

$$\frac{1}{1-r^2}\left(\frac{(x_1-\hat{x}_1)}{s_1^2} - 2r\frac{(x_1-\hat{x}_1)(x_2-\hat{x}_2)}{s_1 s_2} + \frac{(x_2-\hat{x}_2)}{s_2^2}\right),$$

where $\hat{x}_1, \hat{x}_2$ are the sample means, s_1^2, s_2^2, are the sample variances and r is the correlation between X_1 and X_2.

The coordinates of the control ellipses can be plotted using the vectors

$$X_1(t)_i = \mu_1 + a_i\cos(\theta) - b_i\sin(\theta)$$

$$X_2(t)_i = \mu_2 + a_i\cos(\theta) - b_i\sin(\theta),$$

where

$$a_i = -1.96 * \sqrt{\frac{\mu_1 * (1-\mu_1)}{N_i}}$$

and

$$b_i = -1.96 * \sqrt{\frac{\mu_2 * (1-\mu_2)}{N_i}}$$

i.e. the lengths of the semi-major and semi-minor axes, respectively, of the ellipse which define the size of the control region, N_is are the size of the ith centre. For the 3D case here, θ is an $n \times n$ matrix of the form:

$$\theta = \begin{bmatrix} N_1 & N_1 & \cdots & N_1 \\ N_2 & N_2 & \cdots & N_2 \\ \cdots & \cdots & \cdots & \cdots \\ N_n & N_n & \cdots & N_n \end{bmatrix}$$

To rotate the ellipses in the direction of greatest variance, an angle between the major axis of the ellipse and the x-axis can be defined by

$$\phi = \arctan\frac{v_1}{v_2},$$

where v denotes the eigenvector corresponding the largest eigenvalue from an eigen decomposition of the correlation matrix of the two outcomes (or random effects matrix of the adjusted outcomes). The coordinates of the ellipses are then defined as

$$X_1(t)_i = \mu_1 + a_i\cos(\theta)\cos(\phi) - b_i\sin(\theta)\sin(\phi)$$

$$X_2(t)_i = \mu_2 + a_i\cos(\theta)\sin(\phi) - b_i\sin(\theta)\cos(\phi).$$

Figure 9 shows a screen-shot of the 3D funnel plot laid out in a similar manner to the standard two-dimensional funnel plot, with the precision variable varying across the horizontal axis, ascending from left to right. From this angle, the contribution of the Status variables (proportion of Deaths) to the position of observations around the control region is more easily visible.

Figure 9: Horizontal view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers. The target is 23% complications and 4% mortality rate.

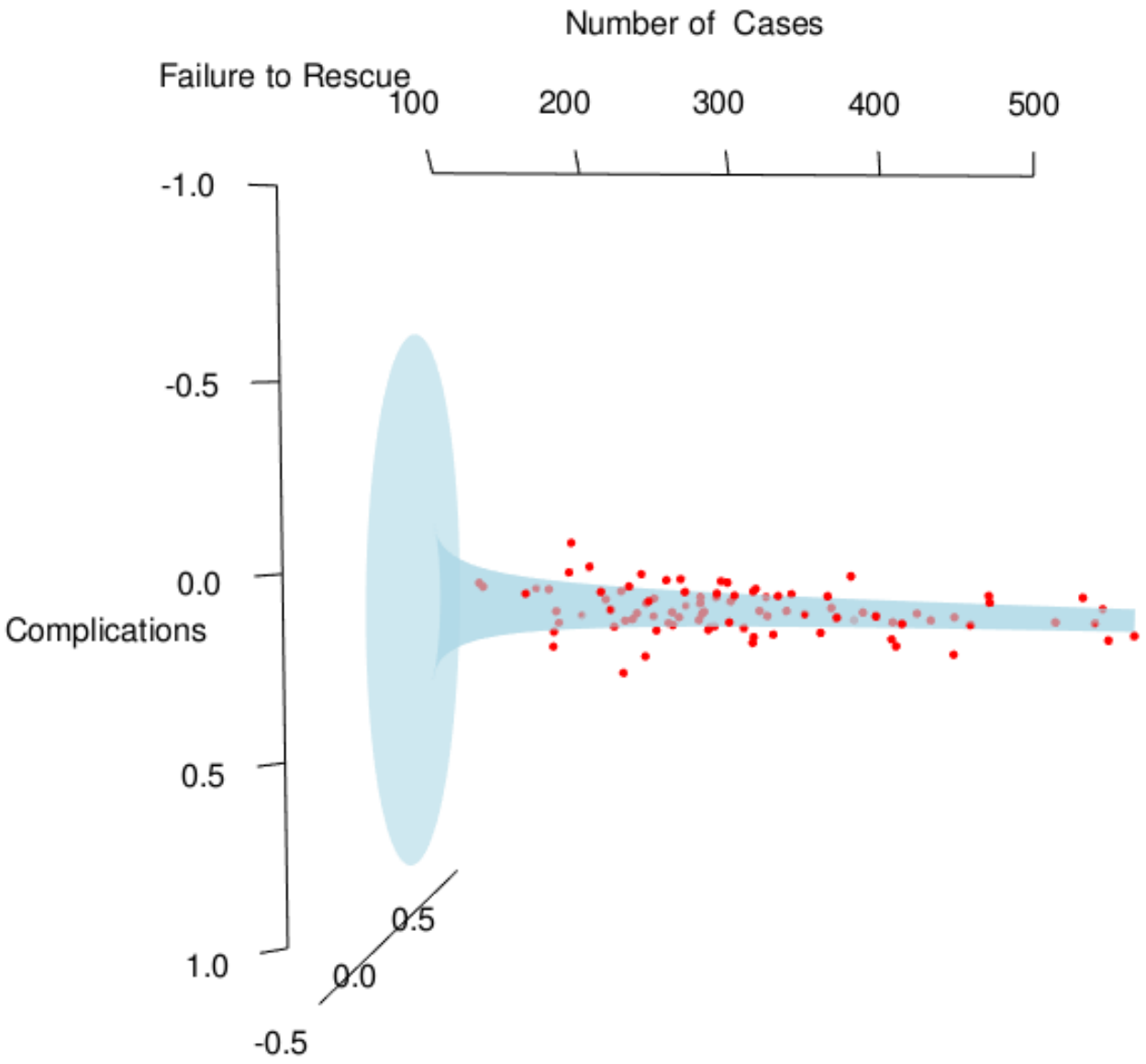


Figure 10 shows the same figure but rotated so that the precision variable varies across the vertical axis and the two indicator variables vary in the horizontal planar space. We suggest that this alignment may be more useful as the interest lies more in these variables than the precision variable, and neither of the outcome variables is considered more important than the other. Traditionally in graphical representations of data, the primary variable of interest (outcome, or dependent variable) is plotted on the Y (vertical axis), in this case as we do not want to give preference to either of the outcome variables, they are given equal prominence on the horizontal plane.

Figure 10: Vertical view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers. The target is 23% complications and 4% mortality rate.

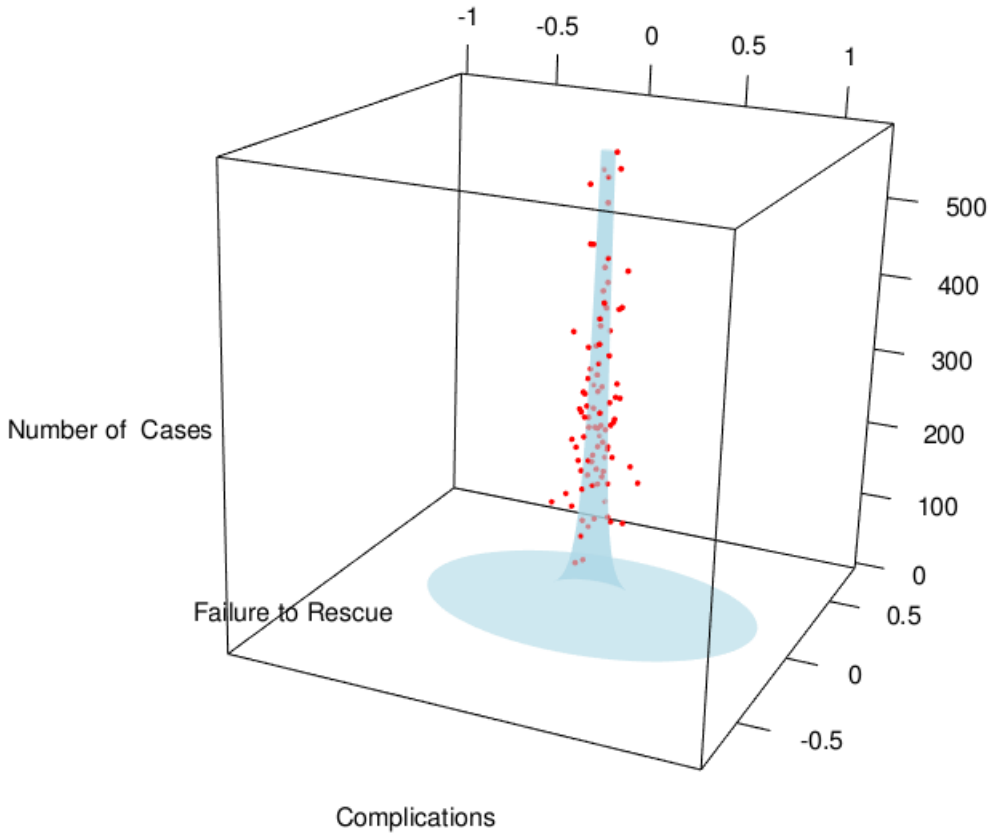


Figure 11 shows the 3d funnel plot with 95% limits and the four smallest centres removed to yield an easier to read view of the plot. From the plot we can see the joint control region more easily, however, to get the full benefit of the 3D funnel plot, the interactive features are required. These allow the user to rotate the graph in any direction and thus multiple statistical features can be extracted visually. For example, rotating the plot so that either of the major or minor axes are fully parallel to the viewing screen allows the viewer to view the same information as in either of the two 2-dimensional funnel plots. This is shown in Figure 12 and Figure 13. By rotating the plot so that the X_1, X_2 plane is parallel to the viewing surface yields a correlation plot of the two indicator variables.

Figure 11: Horizontal view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers with 95% limit. The target is 23% complications and 4% for Failure to Rescue.

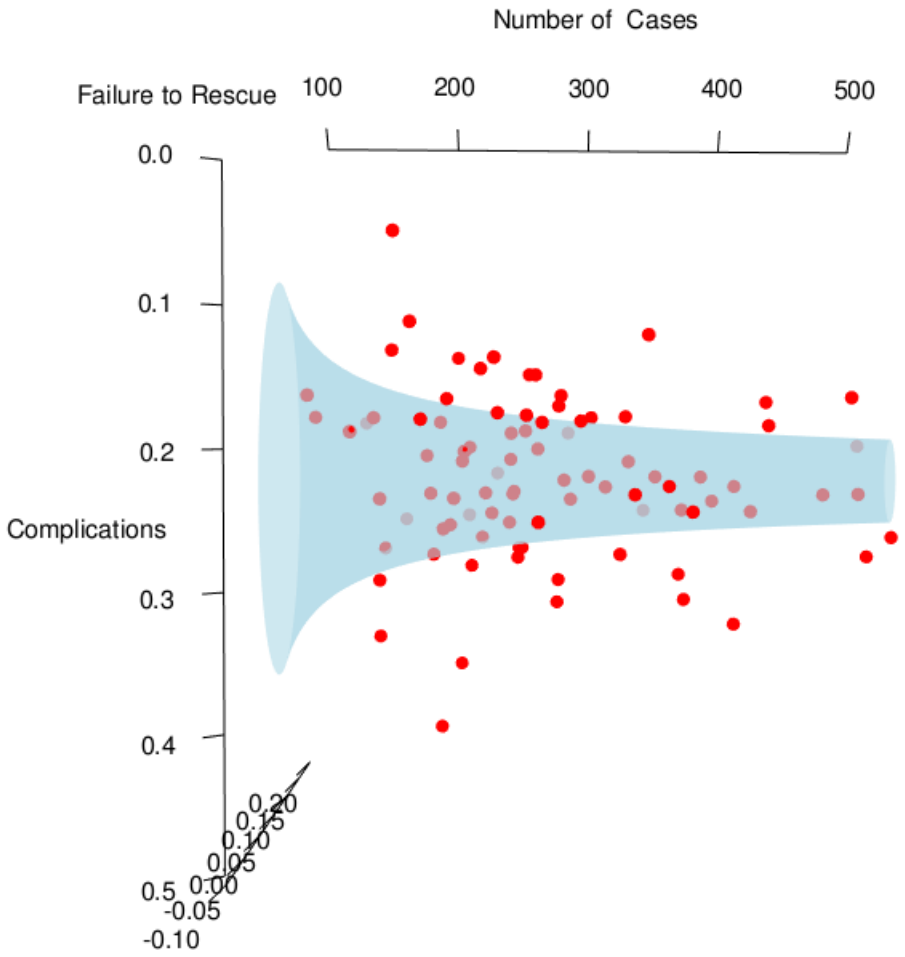


Figure 12: Horizontal view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers with 95% limit. The target is 23% complications and 4% for Failure to Rescue.

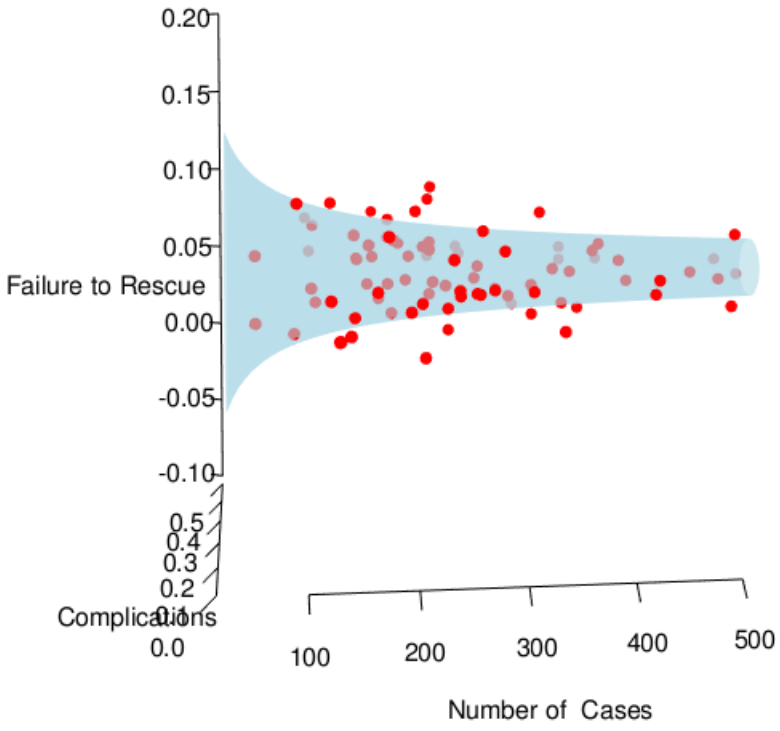
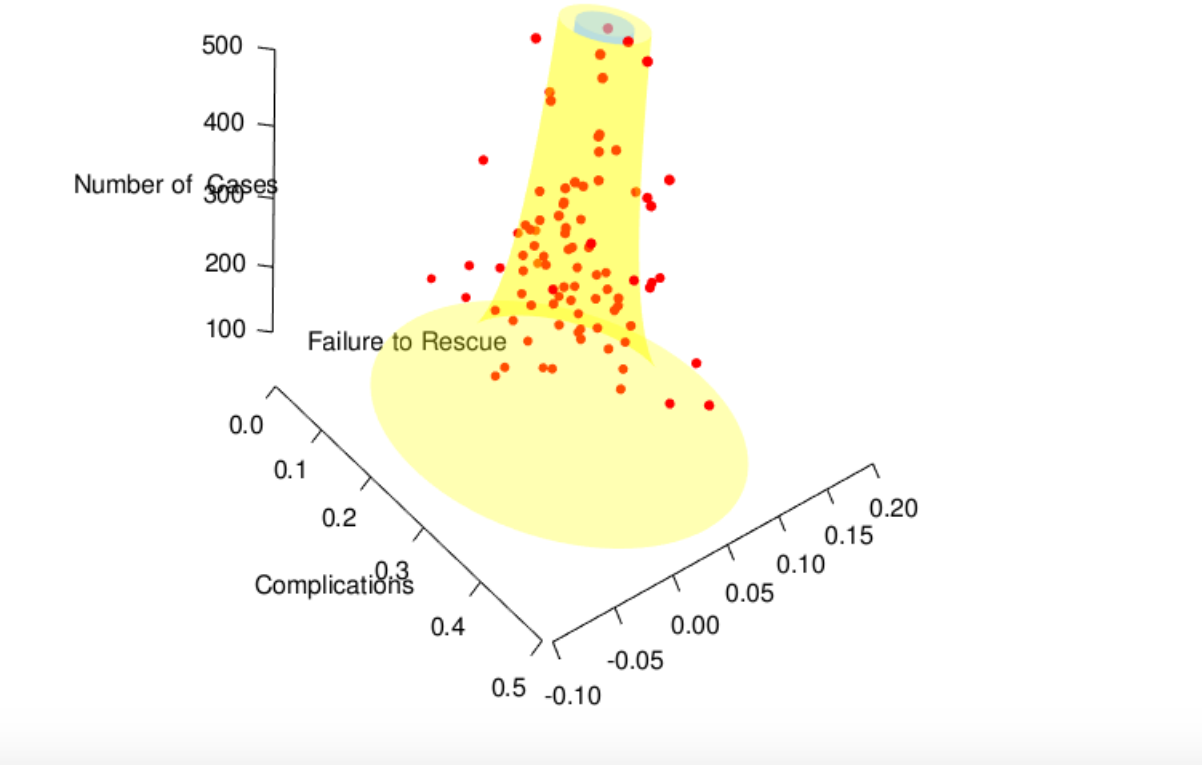


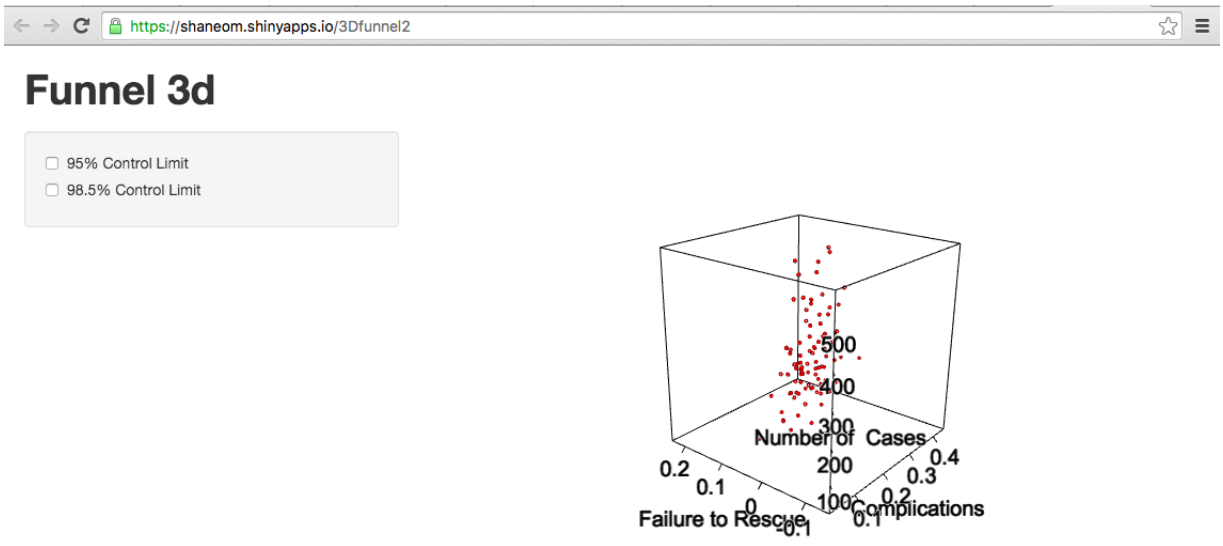
Figure 13: Vertical view of 3D Funnel plot of 30-day mortality and complication rates following surgery in 92 centers with 98% limit. The target is 23% complications and 4% for Failure to Rescue.



4.5 3D Funnel web app

A webapp displaying the 3d funnel plot and data is currently available at `https://shaneom.shinyapps.io/3Dfunnel2`, implemented using packages available for R. RStudio's *shiny* package provides a R wrapper for javascript, html and CSS in order to produce interactive web pages. The user specifies a *ui* (user interface) script and a *server* script. The *ui* script specifies the appearance of the web page as well as the features which allow the user to provide some form of input. The *server* script runs R code on the user input and provides an output back to the *ui* for rendering. Figures 14 and 15 show screenshots of the web app in use. By interacting with the plot, the users acquires an intuitive understanding of the data, as well as specific quantitative information on which centres are in or outside of the control regions. As the functionality is extended users will be able to upload data in an appropriate format and view a 3D funnel plot with user-specified confidence limits.

Figure 14: Screenshots of 3D funnel web app



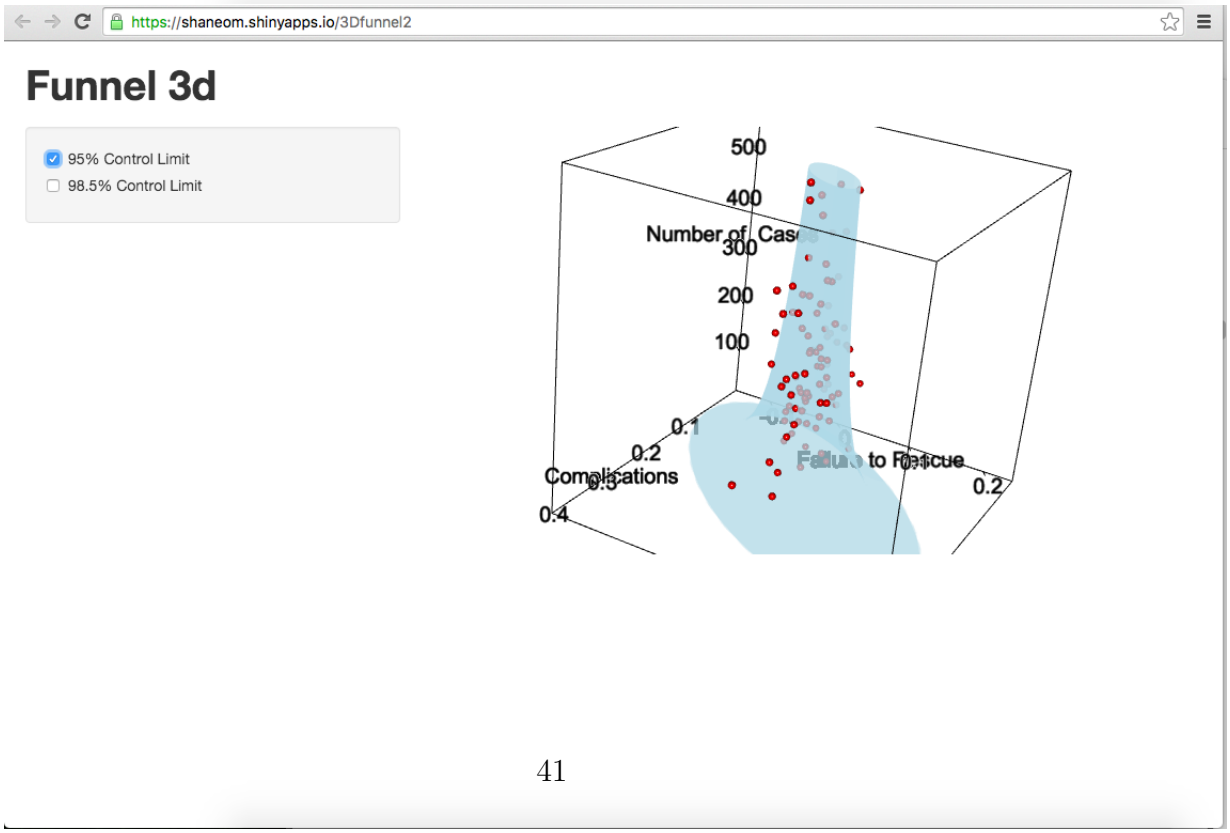
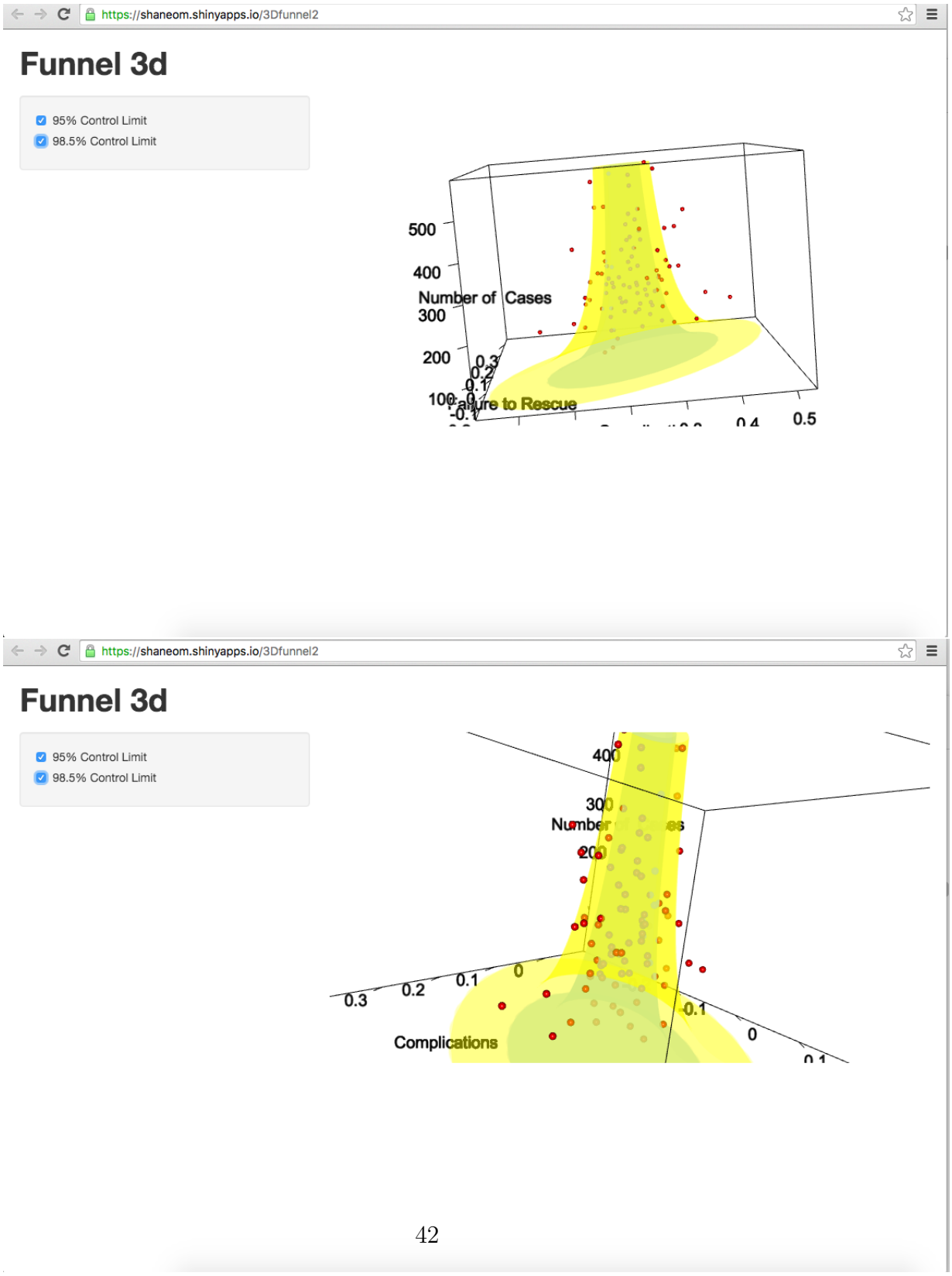


Figure 15: Screenshots of 3D funnel web app



5 Joint model for failure to rescue and complications

5.1 Introduction

As the two outcome variables of interest are a priori likely to be correlated with each other, our interest lies in estimating a joint model for these outcomes. The two outcomes, failure to rescue (FTR) and complication, were modelled jointly using a conditional bivariate Generalised Linear Mixed Model (GLMM). Following discussion with the clinician leading the DSCA, five casemix variables were selected for inclusion in the linear predictor. In addition, a random intercept was fitted for each hospital on each of the two outcome variables to account for unobserved random variation across hospitals that cannot be accounted for by the observed casemix variables.

5.2 A conditional multivariate logistic regression

The two outcomes, complication and FTR, can be modelled separately using Generalised Linear Mixed Models for binary data, to obtain estimates of the fixed effects (casemix) and random effects (residual hospital-level variation) for each outcome. However, the nature of the problem suggests that occurrence of FTR will be correlated with the frequency of complications. A raw correlation for discrete data could be calculated such as the tetrachoric correlation (Bonett and Price 2005). However this is a crude measure of correlation as it does not adjust for casemix and hospital level variation. Joint modelling of FTR and complications would allow for the correct inference to be made on the correlation between the two outcomes after adjustment. It may also lead to more precise estimates of variation and standard errors around fixed effects. Joint models exist for other types of data such as longitudinal responses and survival data (Rizopoulos 2010). However standard models do not capture one particular aspect of the DSCA data, which is the conditional nature of the data, where FTR occurrence is conditional on the occurrence of a complication.

The probability of observing complication, FTR, or neither in an individual patient were each modelled as a logistic function of a linear combination of the casemix variables. To take into account the correlation between patients within the same hospital, a random intercept term for the hospital in which the patient was treated was included in the probability model. This formulation of the probabilities corresponds to a Generalised Linear Mixed Model (GLMM). As we wish to explore not only the effect of the casemix variables on the outcomes but also the relation between FTR and surgical complication, a conditional bivariate

model is employed, where we explore the probability of FTR conditional on a surgical complication having occurred. The model and notation are described below. Let m_i be the number of patients in centre i with $i = 1, \ldots, N$ where N denotes the total number of centres. There are three possible joint outcomes for patient j in hospital i denoted by:

$$X_{ij} = \begin{cases} 1, \textit{ no complication } (0,0) \\ 2, \textit{ complication, no FTR } (1,0) \\ 3, \textit{ complication and FTR } (1,1) \end{cases}$$

The probability of observing each outcome are estimated conditioning on the observed hospital random effects, Θ.

$$p = p(X = x|\Theta)$$

$$logit(p) = \eta$$

where η represents the linear predictor.

For a single response variable

$$\eta = \pi^{(k)} + \beta^{(k)} Z_{ij} + \Theta_i^{(k)}$$

where $\pi^{(k)}$ is the intercept associated with variable k (where $k = 1$ stands for complication and $k = 2$ represents FTR), and Z is the $n \times p$ design matrix.

Therefore

$$p = expit(\eta).$$

The probabilities for each outcome are:

$$P(X_{ij} = 1|\Theta_i) = P(Y_{ij}^{(1)} = 0, Y_{ij}^{(2)} = 0) = 1 - expit(\pi^{(1)} + \beta^{(1)} Z_{ij} + \Theta_i^{(1)}) \tag{3}$$

$$\begin{aligned} P(X_{ij} = 2|\Theta_i) = P(Y_{ij}^{(1)} = 1, Y_{ij}^{(2)} = 0) = \\ expit(\pi^{(1)} + \beta^{(1)} Z_{ij} + \Theta_i^{(1)}) \cdot (1 - expit(\pi^{(2)} + \beta^{(2)} Z_{ij} + \Theta_i^{(2)}) \end{aligned} \tag{4}$$

$$\begin{aligned} P(X_{ij} = 3|\Theta_i) = P(Y_{ij}^{(1)} = 1, Y_{ij}^{(2)} = 1) = \\ expit(\pi^{(1)} + \beta^{(1)} Z_{ij} + \Theta_i^{(1)}) \cdot expit(\pi^{(2)} + \beta^{(2)} Z_{ij} + \Theta_i^{(2)}) \end{aligned} \tag{5}$$

To write out the likelihood, we can set up indicator variables $\delta_1, \delta_2, \delta_3$, where

$$\delta_{ij}^1 = \begin{cases} 1, \ if \ Y_{ij}^{(1)} = Y_{ij}^{(2)} = 0 \\ \\ 0, \ otherwise \end{cases}$$

$$\delta_{ij}^2 = \begin{cases} 1, \ if \ Y_{ij}^{(1)} = 1, Y_{ij}^{(2)} = 0 \\ \\ 0, \ otherwise \end{cases}$$

$$\delta_{ij}^3 = \begin{cases} 1, \ if \ Y_{ij}^{(1)} = 1, Y_{ij}^{(2)} = 1 \\ \\ 0, \ otherwise, \end{cases}$$

and $Y_{ij}^{(1)}, Y_{ij}^{(2)}$ represent the marginal models.

The full likelihood is therefore

$$L(\pi, \beta, \Sigma_\Theta | y) = \prod_i^N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_j^{m_i} p(x_{ij} = 1|\Theta_i)^{\delta_{ij}^1} p(x_{ij} = 2|\Theta_i)^{\delta_{ij}^2} p(x_{ij} = 3|\Theta_i)^{\delta_{ij}^3} f(\Theta_i) d\Theta_i \tag{6}$$

where

$$\Theta_i \sim N(0, \Sigma_\Theta)$$

with

$$\Sigma_{\Theta_i} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

and

$$f(\Theta_i) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left[-\frac{1}{2(1-\rho^2)} \left(\frac{(\Theta_i^{(1)})^2}{\sigma_1^2} + \frac{(\Theta_i^{(2)})^2}{\sigma_1^2} - \frac{2\rho\Theta_i^{(1)}\Theta_i^{(2)}}{\sigma_1\sigma_2} \right) \right]$$

where

$$\rho = cor(\Theta^{(1)}, \Theta^{(2)}).$$

5.3 Two-Stage Approach

As no further simplification can be done on the likelihood equation, closed form solutions to the Maximum likelihood Estimates (MLEs) of the parameters cannot be derived. A two-stage approach was applied to estimating the parameters of the model. In the the first stage, two separate logistic mixed models were fitted for each response. We model the mean probability of observing p^k (where $k = (1, 2)$ for complication and FTR respectively) as

$$\mu_{y_{ij}^k} = logit^{-1}(\pi^{(k)} + \beta^{(k)} Z_{ij} + \Theta_i^{(k)})$$

and the random effects Θ_i are independently distributed

$$\theta_{ij} \sim N(0, \sigma_k).$$

These models can be estimated quickly using R's *lme4* package which utilises *nlminb*'s PORT routine for optimisation, and gaussian quadrature to approximate the log likelihood. Using 3 quadrature points each model is estimated in just under one minute. This model yields estimates of the fixed effects parameters and variance components of interest, however it does not give an estimate of ρ. Therefore in the second stage, the estimates from the first stage can be plugged into the model for the full likelihood and the correlation coefficient, ρ, is estimated using these parameters.

The double integral in the log likelihood is troublesome from a computational perspective and creates a considerable burden on the numerical methods used to maximise the loglikelihood. Three different approaches to approximating the double integral in the likelihood were tested; two-dimensional cubature, making separate calls to one-dimensional cubature, and Gaussian-quadrature. There was no significant difference between the first two methods, however, Gaussian Quadrature led to significant decreases in the time taken to compute the likelihood.

Gaussian Quadrature computes the integral

$$\int_{-\infty}^{\infty} e^{-x^2} f(x) dx$$

using the approximation

$$\sum_{i=1}^{n} w_i f(x_i)$$

where the w_i represent quadrature weights.

Actually the method employed here using the R package *fastQHQuad* evaluates

$$\sqrt{2\sigma}\sum_i w_i \exp(x_i^2) g(\mu + \sqrt{2}\sigma x_i).$$

Repeated calls were made to the *fastQHQuad* package in R to evaluate the double integral. However, when many parameters are estimated a trade-off must be made between the number of quadrature points, which provides a more accurate approximation, and speed. The literature suggest that 3 to 5 points is acceptable in practice. However it is worth experimenting with different numbers of points and checking the results. For the present data set, anything less than six points gave unstable and sometimes non-sensical results, with the correlation coefficient often going toward the boundary of the parameter space, depending on the starting values used. Six points gave consistent, stable results and increasing beyond that did not substantially change the estimates.

Extra constraints were placed on the variance covariance matrix of the random effects to ensure positive semi-definiteness as illustrated below. The condition of positive definiteness is ensured if all of its eigenvalues are positive.

To ensure this condition, the spectral decomposition of the matrix was used such that

$$|\Sigma_\Theta| = UDU'$$

where D is a diagonal matrix with diagonal elements corresponding to the eigenvalues of Σ_Θ,and U is a matrix which satisfies $UU^T = I$, where I is the identity matrix.

5.4 Estimating standard errors for the two-stage approach

As analytic solutions for the parameters and standard errors are not available, other methods must be used for estimation. Bootstrapping is a simple but computationally intensive method for such estimation. Two bootstrap schemes were employed and compared. The first procedure is outlined below.

Bootstrap algorithm 1

1. Generate B bootstrap datasets, each of size n, by sampling from the original dataset with replacement.
2. Estimate the parameter of interest, ρ_b in each b bootstrap dataset, where, $b \in 1:B$.

3. Estimate the sample standard error of the test statistic from across the B bootstrap datasets, ie $SE(\rho_B)$.

The second approach takes into account the hierarchical structure of the data.

Bootstrap algorithm 2

1. For each hospital N_i of size m_i, sample from patients $1 : m_i$ with replacement, this constitutes one b dataset.
2. Estimate the parameter of interest, ρ_b in each b bootstrap dataset , where, $b \in 1 : B$.
3. Estimate the sample standard error of the test statistic from across the B bootstrap datasets, ie $SE(\rho_B)$.

5.5 Results

Table 2 shows the results from both the two-stage approach and the two separate GLMMs. The standard errors for the two-stage approach are calculated from 500 bootstrap samples.

Histogram and qqplot for the estimate of ρ for the two-stage approach are displayed in Figure 16. The two bootstrap algorithms gave almost identical estimates and the plotted results are from the first schema, not taking hospital size into account.

The fixed and random effects estimate are similar between the two-stage approach and the two univariate models as would be expected. Of primary interest for this bivariate model was the value of ρ, representing the residual correlation between peri-operative complication and FTR, after accounting for the casemix variables. The two-stage model gives an estimate of $\rho = 0.8$ (Table 2), representing a strong positive correlation. The direction of the result is what one would intuitively expect a priori, however the strength of the result may be surprising, especially given the unadjusted correlation of 0.46 discussed in Chapter 3. The result implies that after other patient-level factors are taken into account such as age, co-morbidities and pre-operative complications, the association between complication and subsequent FTR is extremely high at a hospital level.

Table 2: Results from the two-stage bootstrap. Parameters associated with complication are indicated by (2)

	GLMM	Two-Stage
Loglikelihood	-21760.88	-15462.42
Rho	NA	0.80 (0.24)
Sigma 1	0.08	0.10 (0.04)
Sigma 2	0.12	0.15 (0.02)
(Intercept)	-4.62 (0.07)	-4.71 (0.08)
ASA > 2	1.17 (0.07)	1.20 (0.08)
Age 75+	1.08 (0.07)	1.10 (0.08)
Charlson 2+	0.42 (0.07)	0.41 (0.07)
Urgency: UIrgent	0.79 (0.09)	0.78 (0.09)
Resection: extensive	0.25 (0.14)	0.23 (0.15)
Resection: limited	0.08 (0.14)	0.05 (0.15)
Pre-op complication	0.36 (0.08)	0.39 (0.09)
(Intercept) (2)	-1.62 (0.03)	-1.70 (0.03)
ASA > 2 (2)	0.57 (0.04)	0.60 (0.04)
Age 75+ (2)	0.20 (0.03)	0.23 (0.03)
Charlson 2+ (2)	0.24 (0.04)	0.22 (0.04)
Urgency: Urgent (2)	0.32 (0.05)	0.35 (0.05)
Resection: extensive (2)	0.32 (0.07)	0.28 (0.07)
Resection: limited (2)	0.13 (0.07)	0.12 (0.07)
Pre-op complication (2)	0.24 (0.04)	0.23 (0.05)

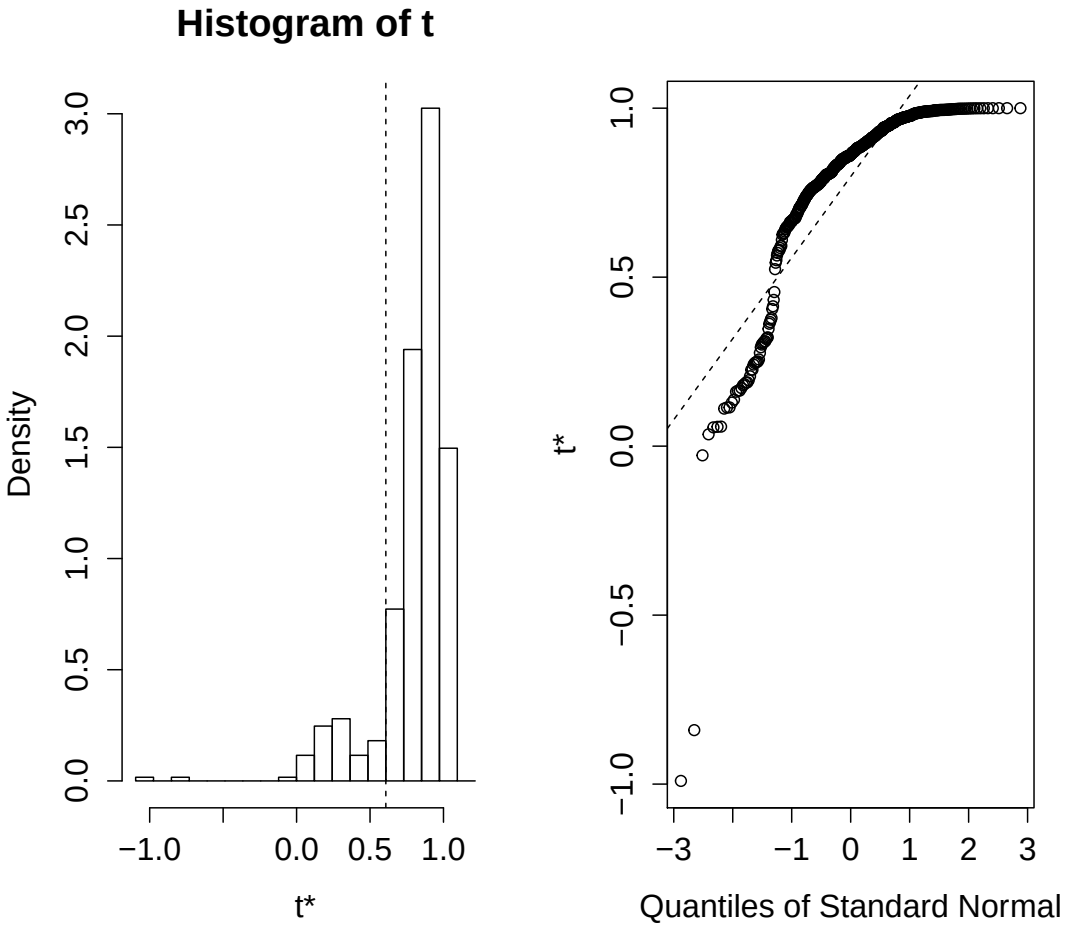
The tight relationship has potentially important policy implications in motivating hospitals to prevent against peri and post-operative complications however possible, whether through increasing surgical resources, surgical training or patient preparation. Whatever the causal mechanism, the results should prompt further investigation by clinicians and hospital managers to identify hospital specific drivers of complication and FTR.

From a methodological point of view, the result shows the benefit of taking the extra step to jointly model the two outcomes and estimate the correlation coefficient, as the estimate is considerably stronger than the naive estimate, a result which could have practical significance for decision-makers.

5.6 A brief simulation study

In order to provide some assurance that our estimate of ρ is at least approximately correct we conduct a very brief simulation study. A random sample of $20,000$

Figure 16: Results from Bootstrap algorithm 1



subjects is taken from the original dataset. A linear predictor is constructed from the original casemix variables and the fixed effects estimated from the two univariate GLMMs. Normally distributed noise, σ_ϵ, is added to each subject from a $N(0,1)$ distribution, along with a hospital level random effect, σ_h, from a multivariate normal distribution with zero means and covariance matrix given by

$$\Sigma_h = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

with different values of ρ used for each simulation.

The linear predictors thus take the form

$$\eta_1 = \beta_0^{(1)} + X\beta_1^{(1)} + \sigma_h^{(1)} + \sigma_\epsilon$$
$$\eta_2 = \beta_0^{(2)} + X\beta_1^{(2)} + \sigma_h^{(2)} + \sigma_\epsilon,$$

where superscripts denote parameters associated with each of the two outcomes.

Finally, the outcome variables $Y^{(1)}, Y^{(2)}$ are simulated from binomial distributions as follows:

$$Y^{(1)} \sim \text{Binomial}(n^{(1)}, \text{expit}(\eta_1))$$
$$Y^{(2)} \sim \text{Binomial}(n^{(2)}, \text{expit}(\eta_2)).$$

The two-stage bootstrap model is fitted to this new dataset and results for two different values of ρ are displayed below in Table 3.

Table 3: Results from two simulations to estimate ρ

	True	Estimate
Simulation 1	0.5	0.50 (0.05)
Simulation 2	-0.5	-0.42 (0.06)

The difference between the true and estimated value in the second study is larger, possibly due to the simplistic simulation setup combined with the overall positive correlation between the two outcomes. However, the results are close enough to the true values to give some reassurance as to the veracity of the direction of the results from the actual data set.

6 Bayesian semi-parametric random effects

6.1 Introduction

In practice when dealing with mixed effects models it is typical to assume a normal distribution for the random effects. Often consideration is given only to testing whether a random component is necessary, or at most testing the covariance structure of the random effects matrix. However, there may be little reason to assume a normal distribution a priori. Other parametric distributions can be fitted. Here we look at using a Dirichlet process to estimate a semi-parametric distribution for random effects within a Bayesian framework.

6.2 Dirichlet process

Ohlssen, Sharples, and Spiegelhalter (2007) outline the basic principles of a Dirichlet process. F is some unknown non-parametric distribution. This target distribution is modelled using a positive real α and some continuous 'baseline' distribution F_0. The real line is then broken down into k disjoint segments

$$(-\infty, x_1), [x_1, x_2), \ldots, [x_{k-2}, x_{k-1}), [x_{k-1}, \infty)$$

where

$$-\infty \leq x_1 \leq x_2 \leq \ldots \leq x_{k-1} \leq x_k \leq \infty$$

and

$$p_1 = F(x_1), p_2 = F(x_2) - F(x_1), p_3 = F(x_3) - F(x_2), \ldots, p_{k-1} = F(x_{k-1} - F(x_{k-2}),$$

and $p_k = 1 - F(x_{k-1})$ are the probabilities of falling in the respective segments of the target distribution. The probabilities for the baseline distribution are given by $p_{0,k-1} = F_0(x_{k-1}) - F_0(x_{k-2})$. The p's come from a Dirichlet distribution

$$(p_1, p_2, \ldots, p_k) \sim Dir(\alpha p_{0,1}, \alpha p_{0,2}, \ldots, \alpha p_{0,k}).$$

The α parameter measures the variability of F around F_0, where higher values of α indicates F is close to F_0.

Dirichlet Processes (DPs) can be used to simulate from random distribution functions. First, draw a random sequence $\theta_1, \theta_2, \ldots$ from F_0, then draw $\eta_1, \eta_2, \ldots$ from a Beta$(1, \alpha)$, so $p(\eta_i) = \alpha\eta_i^{\alpha-1}$ and $\mathrm{E}(\eta_i) = (1+\alpha)^{-1}$. Probability is assigned to the θ_i in the following manner, $p_1 = \eta_1$ to θ_1, $p_2 = (1-\eta_1)\eta_2$ to θ_2, $p_3 = (1-\eta_1)(1-\eta_2)\eta_3$ to θ_3 etc. Let $q_1 = (1-\eta_1)$, $q_2 = (1-\eta_1)(1-\eta_2)$ etc, so that

$$p_k = \eta_k \prod_{j<k}(1-\eta_j) = \eta_k q_{k-1}.$$

Each portion $1-\eta_i$ has expectation $\alpha/(1+\alpha)$, therefore after $N-1$ segments are assigned, the remainder is given by

$$E[1-\sum_{i=1}^{N-1} p_i] = E[q_{N-1}] = E[\prod_{i=1}^{N-1}(1-\eta_i)] = \left(\frac{\alpha}{1+\alpha}\right)^{N-1}.$$

Under this definition of the DP, realisations from the process can be represented as infinite-component mixtures of point masses, and the density function can be represented by

$$f(\cdot) = \sum_{k=1}^{\infty} p_k I_{\theta_k}, \theta_k \sim F_0.$$

Here $f(\cdot)$ is the density function corresponding to F and I_{θ_k}. It is implied in this definition that realisations from the DP are discrete probability distributions. However the DP can be constructed also as a mixture of continuous distributions such that

$$f(\cdot) = \sum_{k=1}^{\infty} p_k h(\cdot|\theta_k), \theta_k \sim F_0.$$

With both of these models, practical problems arise for implementation due to the infinite space of components. In practice, MCMC algorithms are used within a restricted space of N components so that

$$\sum_{k=1}^{\infty} p_k I_{\theta_k} \approx \sum_{k=1}^{N} p_k I_{\theta_k}.$$

This is referred to as truncated Dirichlet process and is denoted by $F \sim TDP(\alpha, F_0, N)$. A restriction also placed on the Nth weight so that weights sum to one, ie. $p_N = 1 - \Sigma_{k=1}^{N} p_k$. These can then be modelled as finite mixture models. Similarly the continuous model can be modelled such that

$$\sum_{k=1}^{\infty} p_k h(.|\theta_k) \approx \sum_{k=1}^{N} p_k h(.|\theta_k).$$

Spiegelhalter et al. recommend choosing a value of N such the final point p_N is small i.e. $E[p_N] \approx \epsilon$. From the definition of the segments one can calculate

$$N \approx 1 + \frac{log\epsilon}{log[\alpha/(1+\alpha)]}.$$

They further show that $N \approx 1 - \alpha log(\epsilon)$, and so if $\epsilon = 0.01$, a rough but conservative approximation of N is $5\alpha + 2$. Previous research showed that for values of α up to 10, the tail probability as represented by p_N was negligible once N is 'large', say around 50. The authors show that this rough approximation gives similar results to a more rigorous approach.

6.3 Model specification

The baseline distribution F_0 is given a form similar to a normal random effects model;

$$\theta_k \sim N(\mu_{F_0}, \sigma^2_{F_0}), \quad k = 1, \ldots, N.$$

Weak priors are used which give coverage outside of the observed range of the data, namely

$$\mu_{F_0} \sim N(0, 10^2)$$
$$\sigma_{F_0} \sim Uniform(0, 10).$$

Particularly important is the specification of α. Smaller values of α imply all centres form one big cluster, i.e. a common random effect and large values indicate support for individual normal random effects. Spiegelhalter et al. recommend a prior distribution of

$$\alpha \sim \text{Uniform}(0.3, 10)$$

with N set to 52 to handle the larger values of α.

6.4 Results

The output from the DP random effects model suggests that the distribution of random effects is not normal and thus the standard normal random effects model may not be suitable. The posterior mean of α is less than two indicating some clustering of centres.

6.4.1 Complications

Four chains were run for $200,000$ iterations, discarding the first $1,000$ burn-in and adaptive iterations. Figure 17 shows trace plots for 4 selected complication parameters; α, p_1, p_{27}, and ρ_1. Mixing is slightly poorer for the p parameters than the ρs, however diagnostic checks using the Gelman-Rubin statistic (Gelman et al. 2013) suggest stationarity.

Figure 17: Trace and density plots for complication parameters α, p_1, p_N and ρ_1

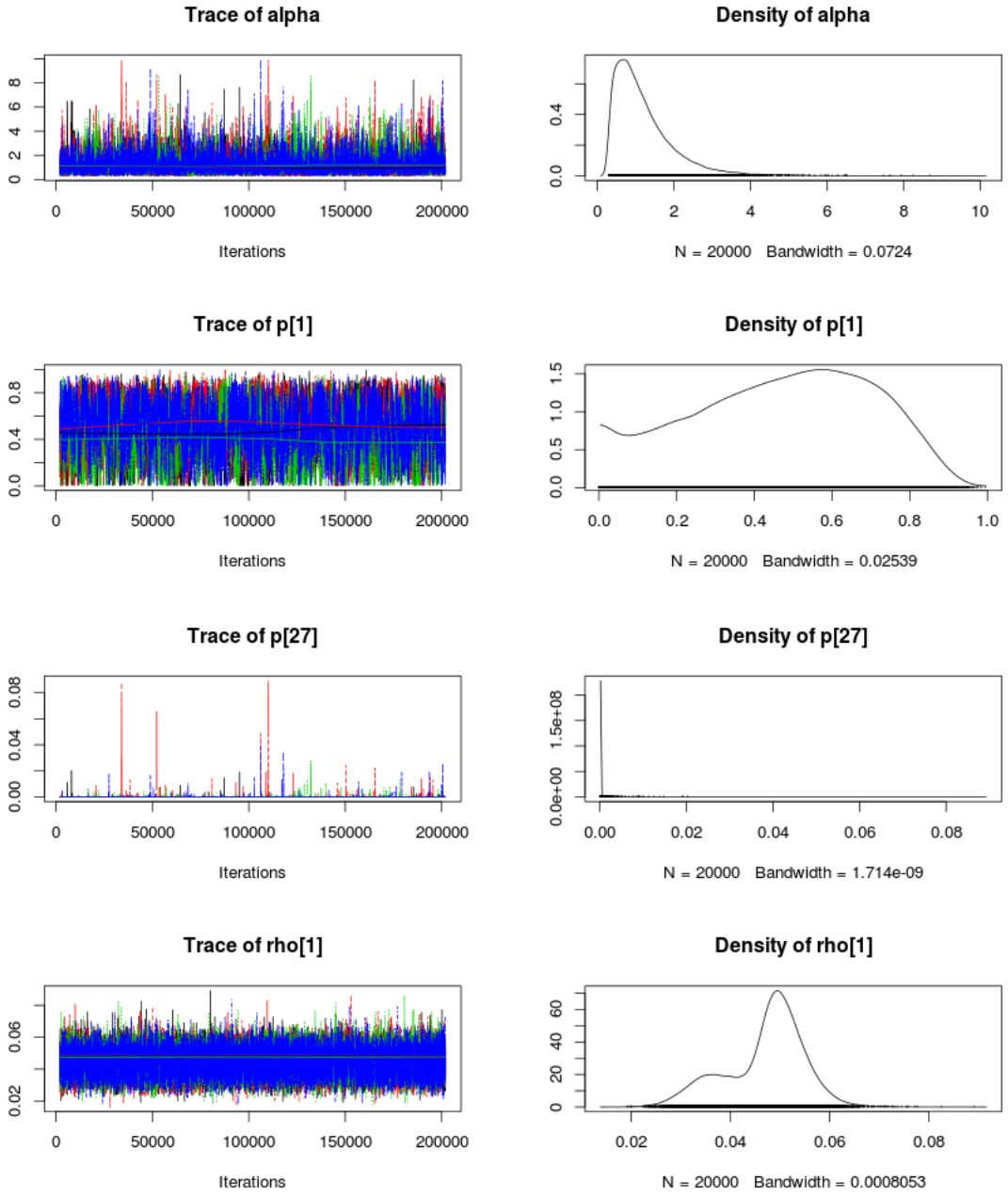
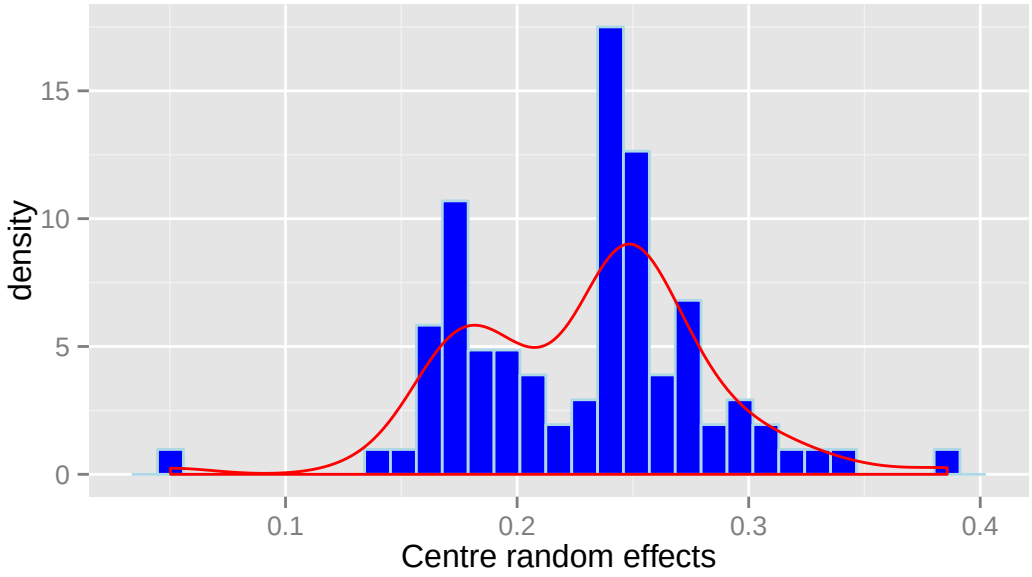


Figure 18 shows the histogram and density of hospital random effect means. The graphs suggest a skewed bi-model distribution of random effects, implying that the normal random effects model may not be suitable and so the flexible model is more appropriate.

Figure 18: Histogram and density plot for hospital complication effects



The MCMC samples can be used to construct credible intervals around the point estimates for each centre. Joining these together creates a point-wise credible band. Figure 19 shows the empirical 95% point-wise credible intervals for complications, using LOESS smoothing. The bands are not non-decreasing/increasing, but show some narrowing as the centre size increases in line with the traditional funnel plot.

Figure 19: Centre complication rate means and smoothed pointwise 95% credible bands

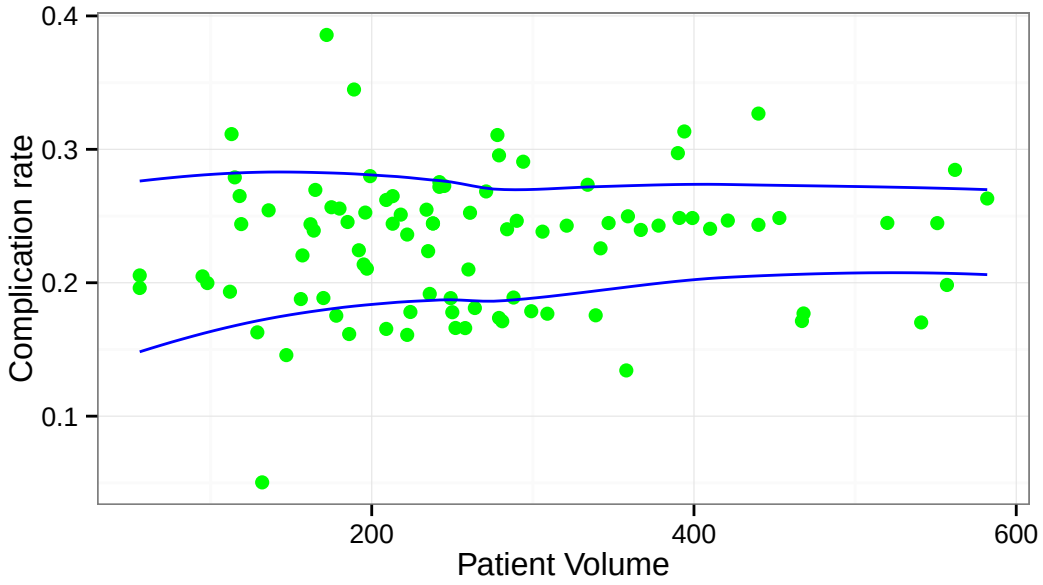
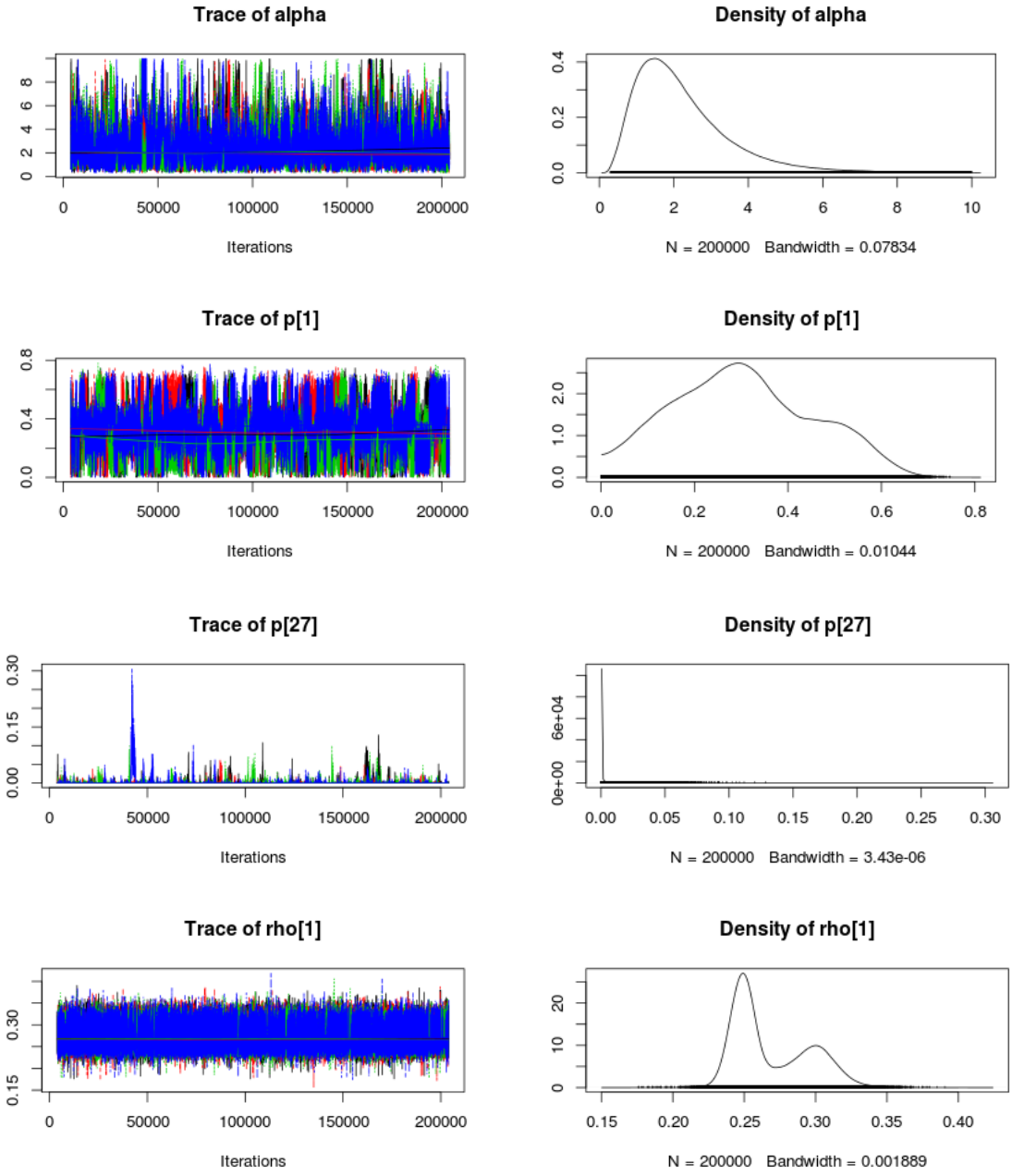


Figure 20: Trace and density plots for mortality parameters α, p_1, p_N and ρ_1



6.4.2 Mortality

Figure 20 shows trace plots for 4 selected complication parameters again after $200,000$ iterations and $1,000$ burn-in and adaptive iterations. Diagnostic checks

using the Gelman-Rubin statistic suggest stationarity.

Figure 21 shows a histogram and density plot of hospital random effect means. Here the distribution is left-skewed with some slight suggestion of bimodality.

Figure 21: Histogram and density plot for hospital random effects mortality rate means

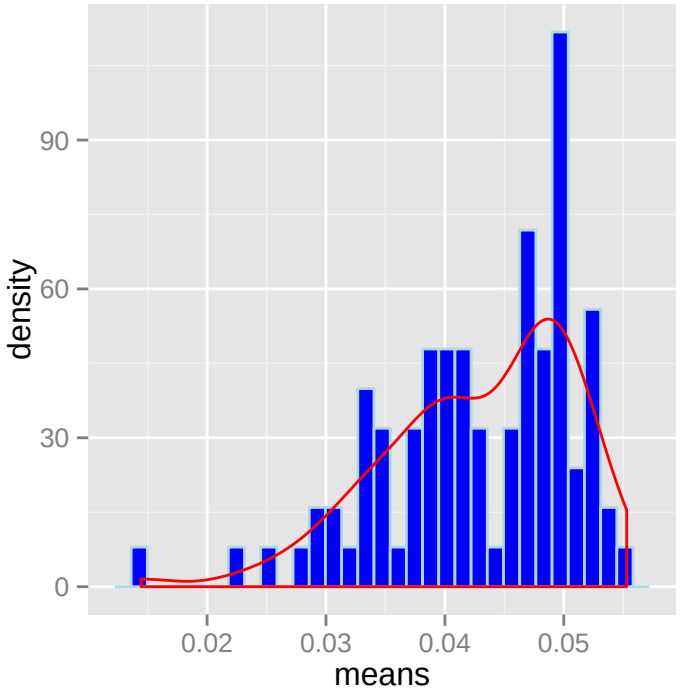
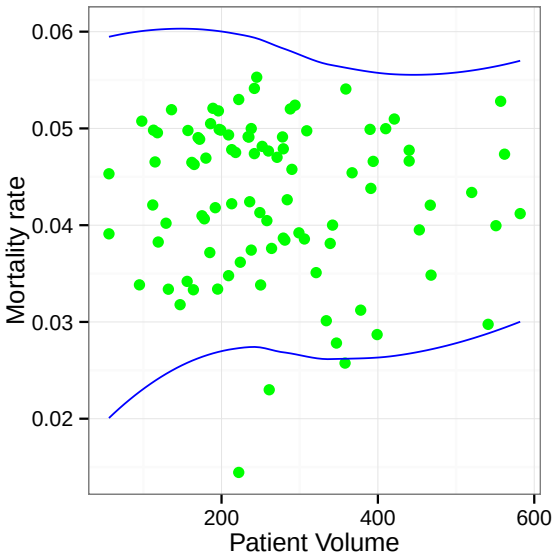


Figure 22 shows a smoothed 95% point-wise credible intervals using LOESS smoothing. In this case the smoothed bands are wider, illustrating less narrowing with volume than the traditional plots and encompassing most of the data points.

Figure 22: Centre mortality rate means and smoothed pointwise 95% credible intervals



7 Discussion

An accurate assessment of institutional performance is an essential tool in both knowledge dissemination for users and for helping institutions themselves improve. A ranking of hospitals will both allow prospective patients make an informed decision about elective surgery for example, while also giving an impetus to poorly performing hospitals to improve performance. Explicit modelling provides information on factors that may be associated with poorer or better performance. Centipede plots show institutional ranks along with uncertainty estimates for each performance metric. Chapter 3 shows that choice of modelling used in generating the centipede plot can have a large influence on the apparent differences between centres. Using Empirical Bayes (EB) methods for the same plots shrinks the differences even further, in principle reducing the effect of unusual results. The funnel plot allows for visualising institutional outcomes as a function of patient volumes and includes a test of whether centres are within some measure of 'normal' performance for their respective level of volume. Again the results are subject to modelling assumptions, but the results of funnel plots have found popular use in medical literature. Chapter 4 looked at the issue of using a funnel plot to visualise two performance indicators simultaneously. It showed that the concept of the confidence ellipse along with RStudio and the OpenGL framework for three-dimensional graphics can be used to make a three-dimensional extension of the funnel plot. Such a plot has many potential advantages, mostly in the area of conveying as much necessary information as possible while allowing the user to make intuitive simultaneous inferences about the two indicators. It remains to be seen how clinicians and other users of such information will find the device useful in practice.

Chapter 5 looked at modelling two performance indicators when the indicators are separate events but intimately connected with each other. In this case the second indicator, failure to rescue (FTR), can only be observed conditional on the first, complication. The results showed a strong correlation between the two outcomes, significantly larger than the simple unadjusted estimate. Such a result shows what can be gained from modelling the two outcomes jointly rather than as two independent outcomes. The so-called 'two-stage' approach to estimation was proposed and tested on simulated and real data. However, further research is needed into the optimal estimation method for such models.

In Chapter 6, the assumption of normally distributed hospital effects was set aside and a semi-parametric random effects model based upon a truncated Dirichlet Process was considered. The results of the model showed that the centre effects diverged notably from the normal distribution. The Bayesian approach al-

lows the construction of pointwise credible bands. The chapter showed smoothed bands based upon the pointwise credible regions which gives some inclination as to the possible shape of such a band. The width of the credible region is not necessarily symmetric or decreasing with volume. An open question remains as to the coverage probability of such bands, similar to the funnel plot bands discussed in Chapter 4. In addition such models could be extended to jointly model two or more outcomes, potentially yielding more inferential gains as per Chapter 5.

This thesis has explored statistical issues in ranking, modelling and visualising hospital performance, and made some contribution to the methodologies used. It is important for statisticians and those involved in reporting on institutional performance to chose suitable metrics and mathematical methods. The utility of the bivariate model proposed here as well as the visualisation methods should be tested in a decision-making context to determine avenues for further refinement and as well as new developments. As a practical step in this regard our plotting app has been made freely accessible online for users to inspect and offer opinions on its usability and features. Ultimately the correct choice of method is that which is accurate and which aids decision-making on the part of patients, clinicians and policy-makers. It is important therefore that feedback from such groups be taken into account when developing further methods.

References

Alexandersson, Anders (2004). "Graphing confidence ellipses: An update of ellip for Stata 8". In: *Stata Journal* 4, pp. 242–256.

Batschelet, E. (1981). *Circular Statistics in Biology*. Mathematics in biology. Academic Press. ISBN: 9780120810505.

Berridge, D.M. and R. Crouchley (2011). *Multivariate Generalized Linear Mixed Models Using R*. CRC Press. ISBN: 9781439813270.

Bonett, Douglas G and Robert M Price (2005). "Inferential methods for the tetrachoric correlation coefficient". In: *Journal of Educational and Behavioral Statistics* 30.2, pp. 213–225.

Daabiss, Mohamed (2011). "American Society of Anaesthesiologists physical status classification". In: *Indian journal of anaesthesia* 55.2, p. 111.

DerSimonian, R. and N. Laird (1986). "Meta-analysis in clinical trials". In: *Controlled Clinical Trials* 7.3, pp. 177 –188. ISSN: 0197-2456. DOI: http://dx.doi.org/10.1016/0197-2456(86)90046-2. URL: http://www.sciencedirect.com/science/article/pii/0197245686900462.

Dutch Colorectal Cancer Group (2015). http://www.dccg.nl/about-dccg/. Accessed: 2015-10-19.

Gelman, A. et al. (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis. ISBN: 9781439840955.

Goldstein, Harvey and David J Spiegelhalter (1996). "League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance". In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 159.3, pp. 385–443.

Gueorguieva, R (2001). "A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family". In: *Statistical Modelling* 1.3, pp. 177–193. DOI: 10.1177/1471082X0100100302. eprint: http://smj.sagepub.com/content/1/3/177.full.pdf+html. URL: http://smj.sagepub.com/content/1/3/177.abstract.

Hadfield, Jarrod D (2010). "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package". In: *Journal of Statistical Software* 33.2, pp. 1–22. URL: http://www.jstatsoft.org/v33/i02/.

Henneman, D. et al. (2013). "Hospital Variation in Failure to Rescue after Colorectal Cancer Surgery: Results of the Dutch Surgical Colorectal Audit". English. In: *Annals of Surgical Oncology* 20.7, pp. 2117–2123. ISSN: 1068-9265. DOI: 10.1245/s10434-013-2896-7. URL: http://dx.doi.org/10.1245/s10434-013-2896-7.

Houwelingen, Hans C van, Ronald Brand, and TA Louis (2004). "Empirical bayes methods for monitoring health care quality". In: *Bulletin of the ISI, ISI* 99, pp. 75–78.

Kahneman, Daniel (2011). *Thinking, Fast and Slow*. Macmillan. URL: /bib/kahneman/kahneman2011thinking/Thinking%2C%20Fast%20and%20Slow%20-%20Kahneman%2C%20Daniel.epub.

Kendall, M. G. (1938). "A New Measure of Rank Correlation". English. In: *Biometrika* 30.1/2, pp. 81–93. ISSN: 00063444. URL: http://www.jstor.org/stable/2332226.

Laird, Nan M and Thomas A Louis (1989). "Empirical Bayes Ranking Methods". In: *Journal of Educational Statistics* 14.1, pp. 29–46.

Light, Richard J and David B Pillemer (1984). *Summing up : the science of reviewing research*. Harvard University Press.

McCulloch, C.E., S.R. Searle, and J.M. Neuhaus (2011). *Generalized, Linear, and Mixed Models*. Wiley Series in Probability and Statistics. Wiley. ISBN: 9781118209967.

Medics' NHS league table mortality figures mired in confusion (28/06/2013). http://www.theguardian.com/society/2013/jun/28/mortality-figures-surgeons-nhs-estimates. Accessed: 2015-10-19.

Ohlssen, David I, Linda D Sharples, and David J Spiegelhalter (2007). "Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons". In: *Statistics in medicine* 26.9, pp. 2088–2112.

Rizopoulos, Dimitris (2010). "JM: An R package for the joint modelling of longitudinal and time-to-event data". In: *Journal of Statistical Software* 35.9, pp. 1–33.

Shewhart, W.A. (1931). *Economic control of quality of manufactured product*. Bell Telephone Laboratories series. D. Van Nostrand Company, Inc.

Spiegelhalter, David J. (2005). "Funnel plots for comparing institutional performance". In: *Statistics in Medicine* 24.8, pp. 1185–1202. ISSN: 1097-0258. DOI: 10.1002/sim.1970. URL: http://dx.doi.org/10.1002/sim.1970.

Zuk, Or, Liat Ein-Dor, and Eytan Domany (2007). "Ranking Under Uncertainty". In: *UAI 2007, Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence, Vancouver, BC, Canada, July 19-22, 2007*, pp. 466–473.

8 Appendix: R Code

8.1 R code for Chapter 4

```
1
2  setwd("") # set working directory
3
4  install.packages("rgl", dependencies=TRUE)
5  install.packages("rglwidget", dependencies=TRUE)
6
7  # required packages
8  library(foreign)
9  library(lme4)
10 library(rgl)
11 library(gplots)
12
13 # Data Preparation
14 hosps <- read.spss("data_FTR_Shane.sav", use.value.labels=T, to.data
       .frame=T)
15
16 hosps <- hosps[,c("id", "yyASAgrp", "xleef75", "fcharlgrp", "XTYPOK"
       , "xresadd",
17                    "XCOMPLPRE", "XBELOOP", "status")] # keep casemix
       variables
18
19 hosps$yyASAgrp <- factor(hosps$yyASAgrp)
20 hosps$id <- factor(hosps$id)
21 levels(hosps$id) <- 1:92
22 levels(hosps$status) <- c("In leven", "Overleden")
23
24 ASA2 <- ifelse(hosps$yyASAgrp == 1  | hosps$yyASAgrp == 2, 0, # ASA
       as binary variable
25                 ifelse(hosps$yyASAgrp == 3 |
26                          hosps$yyASAgrp == 4 |
27                          hosps$yyASAgrp == 9 , 1, hosps$yyASAgrp))
28 hosps$ASA2 <- factor(ASA2)
29
30
31
32 # plot deaths per hospital
33 status_table <- table(hosps$id, hosps$status)# death rates per
       hospital
34 status_proporps <- as.numeric(status_table[,2]/table(hosps$id)) #
       proportion of deaths
35
36 plot(1:92, status_proporps, pch=21, bg=11, xlab = "Centre",
37      ylab ="Proportion of Deaths",
38      main="Proportion of Deaths by Hospital", ylim=c(0, .09))
```

```
39
40
41
42 # plot complications by hospital
43 beloop_table <- table(hosps$id, hosps$XBELOOP) # complication rate
44 beloop_proporps <- as.numeric(beloop_table[,2]/table(hosps$id)) #
       complication proportion
45 plot(1:92, beloop_proporps , pch=21, bg=11,xlab = "Centre",
46      ylab ="Proportion of Complications")
47
48
49
50 ### Funnel Plots ###
51
52 #complications
53 comp <- as.numeric(hosps$XBELOOP) - 1 # convert to 0/1
54
55 p0=mean(comp)  # overall mean 0.23
56 x=tapply(comp, hosps$id, sum) # for each center compute the number
       of observed complications
57 n=tapply(comp, hosps$id,length) # number of cases in each center
58 p=x/n # complication proportions
59
60 plot(n,p, pch=21, bg=11,
61      ylab="Complication rate", xlab = "Patient volume")
62
63 vec2 = seq(0,600,1) # plotting vector for x-axis 'Volume'
64
65 # plot control limits
66 lines(vec2, p0-1.96*sqrt(p0*(1-p0)/vec2),
67       col="lightblue", lwd=1)
68 lines(vec2, p0+1.96*sqrt(p0*(1-p0)/vec2),
69       col="lightblue", lwd=1)
70 abline(a=p0, b=0, lwd=1) # target
71
72 # 97.5% limits
73 lines(vec2, p0-2.24*sqrt(p0*(1-p0)/vec2),
74       col="yellow", lwd=1)
75 lines(vec2, p0+2.24*sqrt(p0*(1-p0)/vec2),
76       col="yellow", lwd=1)
77
78 lims <- cbind(p, n, p0-1.96*sqrt(p0*(1-p0)/n), p0+1.96*sqrt(p0*(1-p0
       )/n)) #limits
79
80 # identify observations outside limits
81 outside <- apply(lims, 1, function(x) {
82   as.numeric(x[1] < x[3] | x[1] > x[4])
83 })
84
```

```
below <- apply(lims, 1, function(x) {
  as.numeric(x[1] < x[3])
})

above <- apply(lims, 1, function(x) {
  as.numeric(x[1] > x[4])
})



# Mortality
status <- as.numeric(hosps$status) - 1

p02=mean(status)  # overall mean 0.04
x2=tapply(status, hosps$id, sum) # for each center compute the
    number of observed complications
n2=tapply(status, hosps$id,length) # number of cases in each center
p2=x2/n2  # proprtions



plot(n2,p2, pch=21, bg=11,
     ylab="Mortality rate (Status)", xlab = "Patient volume")

vec2 = seq(0,600,1)

# 95% limits
lines(vec2, p02-1.96*sqrt(p02*(1-p02)/vec2),
      col="lightblue", lwd=1)
lines(vec2, p02+1.96*sqrt(p02*(1-p02)/vec2),
      col="lightblue", lwd=1)
abline(a=p02, b=0, lwd=1)

# 97.5% limits
lines(vec2, p02-2.24*sqrt(p02*(1-p02)/vec2),
      col="yellow", lwd=1)
lines(vec2, p02+2.24*sqrt(p02*(1-p02)/vec2),
      col="yellow", lwd=1)


### Random Effects Models ###

# Remove Observations not used
hosps <- hosps[!(hosps$status=="Overleden" &
                   hosps$XBELOOP=="ongecompliceerd"), ]

hosps <- hosps[complete.cases(hosps),]

# random effects model for mortality
```

```
model1 <- glmer(status ~ ASA2 + xleef75 + fcharlgrp +
                  XTYPOK +xresadd + XCOMPLPRE + (1|id),
                data=hosps,
                family=binomial(link="logit"),
                na.action=na.omit, nAGQ = 5)

# random effects model for complications
model2 <- glmer(XBELOOP ~ ASA2 + xleef75 + fcharlgrp +
                  XTYPOK +xresadd + XCOMPLPRE + (1|id),
                data=hosps,
                family=binomial(link="logit"),
                na.action=na.omit, nAGQ = 5)

tmp1 <- summary(model1)
tmp2 <- summary(model2)

# fixed effect coefficients
betas1 <- fixef(model1)
betas2 <- fixef(model2)


## Empirical Bayes estimates ##
pm1=ranef(model1,drop=T)$id # conditional ("posterior") means
pv1=attr(ranef(model1,drop=T,condVar=TRUE)$id,"postVar")  #
    conditional ("posterior") variances
names(pv1)=names(pm1)

tau1=as.numeric(VarCorr(model1))  #"conjugate prior" variance
UCI <- pm1 + 1.96 * sqrt(pv1) # upper confidence limit
LCI <- pm1 - 1.96 * sqrt(pv1) # lower upper confidence limit

matrixCI1=cbind(pm1,UCI,LCI)
matrixCI1=matrixCI1[order(matrixCI1[,1]),] #sort matrix on pm and
    sort corresponding CI along with it

#plot EB+CI
plotCI(x=matrixCI1[,1], ui=matrixCI1[,2], li=matrixCI1[,3],  col="
    black", barcol="blue",
       lwd=1, xlab="", ylab="")
#title("corrected log odds complication ratios for all centers
    sorted by posterior mean")
title("center effects with EB methodology")
plotCI(x=matrixCI1[,1], ui=matrixCI1[,2], li=matrixCI1[,3], ylim=c
    (-4,2), col="black",
       barcol="blue", lwd=1, xlab="", ylab="")
title("center effects with EB methodology scale as FE")

# variance components
sigma1 <- VarCorr(model1)$id[1,1]
```

```
sigma2 <- VarCorr(model2)$id[1,1]


### 3D funnel plot ###

##myplots
x <- table(hosps$id, hosps$XBELOOP)[,2] # for each center compute
    the number of observed complications
x <- unlist(x)


x2 <- table(hosps$id, hosps$status)[,2] # for each center compute
    the number of observed complications
x2 <- unlist(x2)


n <- as.numeric(unlist(table(hosps$id)))

p <- x/n
p2 <- x2/n


p0 <- mean(p)
p02 <- mean(p2)

vec=seq(1,max(n), length.out = 92) #vector of plotting points on the
    z axis

x1up <- function(integ) 1.96*sqrt((p0*(1-p0))/integ) # function to
    calculate "upper" boundary of funnel for first dimension
x1vals1 <- x1up(vec) # calulcate boundary value at each point on the
    axis

x2up <- function(integ) 1.96*sqrt((p02*(1-p02))/integ)  # function
    to calculate "upper" boundary of funnel for second dimension
x2vals1 <- x2up(vec)

# 98% interval
x1up2 <- function(integ) 2.3*sqrt((p0*(1-p0))/integ)
x1vals2 <- x1up2(vec)

x2up2 <- function(integ) 2.3*sqrt((p02*(1-p02))/integ)
x2vals2 <- x2up2(vec)


z <- matrix(rep(vec,each=length(vec)), length(vec), # plotting grid
    for z-axis
            length(vec), byrow=T)

```

```
piseq <- seq(0, 2*pi, length.out = 92) # points for mapping ellipse

z2 <- matrix(rep(piseq,each=length(vec)), length(piseq),
             length(piseq), byrow=T) #matrix of plotting points
theta2 <- t(z2)

# compute angle of rotaion in direction of largest eigenvalue based
    on
# random effects correlatiion matrix
crat <- matrix(c(1, rep(cor(x, x2), 2), 1), 2,2, byrow = T)
phi <- atan(eigen(cmat)$vectors[,1][1]/eigen(cmat)$vectors[,1][1]) #
     angle

#define centre of the ellipse
xc <- p0
yc <- p02

# 95% funnel
xt <- xc + x1vals1*cos(theta2)*cos(phi) - x1vals1*sin(theta2)*sin(
    phi)
yt <- yc + x2vals1*cos(theta2)*sin(phi) + x2vals1*sin(theta2)*cos(
    phi)

# 98% funnel
xt2 <- xc + x1vals2*cos(theta2)*cos(phi) - x1vals2*sin(theta2)*sin(
    phi)
yt2 <- yc + x2vals2*cos(theta2)*sin(phi) + x2vals2*sin(theta2)*cos(
    phi)


# plot using RGL
open3d()
par3d(font=1, cex=1, zoom=.82)

persp3d(xt, yt, z, col="lightblue", alpha=.6,
        ylab="Failure to Rescue", zlab="Number of  Cases",
        xlab="Complications", lit=FALSE,
        xlim = c(-1, 1.2), ylim = c(-0.8, 0.8))

plot3d(p, p2, n, add=TRUE, type="s", col="red", size=.4, lit=FALSE,
    shininess=100)

#remove first four z values for better persepctive
aspect3d("iso")
persp3d(xt[-c(1:4), ], yt[-c(1:4), ], z[-c(1:4), ], col="lightblue",
     alpha=.6,
        xlab="Complications", ylab="Failure to Rescue", zlab="Number
     of  Cases",
         lit=FALSE, xlim = c(0,0.5), ylim = c(-0.1, .2),
```

```
259          axes = FALSE)
260 axes3d(c('x--','y+-', 'z'))
261
262 plot3d(p[-c(1:4)], p2[-c(1:4)], n[-c(1:4)], add=TRUE, type="s", col=
    "red", size=.7, lit=FALSE, shininess=100)
263
264 persp3d(xt2[-c(1:4), ], yt2[-c(1:4), ], z[-c(1:4), ], col="yellow",
    alpha=.3, #98% funnel
265          add = TRUE, lit = FALSE)
266
267 # use to take still snapshot
268 rgl.postscript("3dFun.eps", "pdf" )
269
270 rgl.snapshot("Funnell3dcor.png", fmt="png", top=TRUE )
```

8.2 R code for Chapter 5

```
1
2 setwd("")
3
4 ### Packages ###
5
6 install.packages("lme4",
7                   repos=c("http://rstudio.org/_packages",
8                           "http://cran.rstudio.com"))
9
10 library(foreign)
11 library(lme4)
12 library(fastGHQuad)
13 library(boot)
14 library(xlsx)
15
16
17 # Laplaces's demon for multivariate normal distribution using
       Cholesky factoristaion
18 install.packages("LaplacesDemon") # not currently available on CRAN,
       use github
19 library(LaplacesDemon)
20
21 install.packages("Rcpp")
22 install.packages("xml2")
23 install.packages("rversions")
24
25 install.packages("devtools", dependencies = TRUE)
26 library(devtools)
27 options(download.file.method = "libcurl")
28 install_github("ecbrown/LaplacesDemon") # github
29
30 install.packages(pkgs="~/LaplacesDemon_15.03.19.tar.gz", # or zip
       binary
31                   repos=NULL, type="source")
32
33
34 ### data preparation
       ###-----------------------------------------------------
35 data <- read.spss("data_FTR_Shane.sav", use.value.labels=T, to.data.
       frame=T)
36
37 data$geslacht <- factor(data$geslacht)
38 data$id <- factor(data$id)
39 data$status <- factor(data$status)
40 data$yyASAgrp <- factor(data$yyASAgrp)
41
42 levels(data$yyASAgrp) <- c("1","2","3","4",NA)
```

```
levels(data$id) <- 1:92
levels(data$XBELOOP) <- c("no complication","complication")
levels(data$status) <- c("alive","dead")

ASA2 <- ifelse(data$yyASAgrp == 1  | data$yyASAgrp == 2, 0,
               ifelse(data$yyASAgrp == 3 |
                        data$yyASAgrp == 4 |
                        data$yyASAgrp == 9 , 1, data$yyASAgrp))
data$ASA2 <- factor(ASA2)

N <- length(levels(data$id))


# stage I and II models
    ------------------------------------------------------



model1 <- glmer(status ~ ASA2 + xleef75 + fcharlgrp +
                  XTYPOK +xresadd + XCOMPLPRE + (1|id),
                data=data, nAGQ = 2,
                family=binomial(link="logit"),
                na.action=na.omit)
summary(model1)

model2 <- glmer(XBELOOP ~ ASA2 + xleef75 + fcharlgrp +
                  XTYPOK +xresadd + XCOMPLPRE + (1|id),
                data=data,
                family=binomial(link="logit"),
                na.action=na.omit)
summary(model2)

# fixed effect coefficients
betas_start1 <- model1@beta
betas_start2 <- model2@beta

# variance components
sigma1 <- VarCorr(model1)$id[1,1]
sigma2 <- VarCorr(model2)$id[1,1]

# select observations for each of the three joint outcomes (X = 1),
# (X = 2), (X = 3)
x1 <- data[data$status=="alive" & data$XBELOOP=="no complication",]
x2 <- data[data$status=="alive" & data$XBELOOP=="complication",]
x3 <- data[data$status=="dead" & data$XBELOOP=="complication",]

# remove observations with missing data
x1 <- x1[complete.cases(x1),]
x2 <- x2[complete.cases(x2),]
```

```r
x3 <- x3[complete.cases(x3),]

# expit function
expit <- function(x) 1/(1 + exp(-x))

# parameters for random effects: bivariate normal distribution
mu = c(0,0)

# variance-covariance matrix
Sigma = function (rho, sigma1, sigma2) {
  cov12 <- rho*sigma1*sigma2
  matrix(c(sigma1^2, cov12, cov12, sigma2^2), nrow=2, byrow=T)}



### Two-stage estimation ###

loglikelihood <- function(rho, pars, x1, x2, x3){

  centre.probabilities <- NULL
  centres <- as.character(1:N)

  for(i in 1:N){

    # create model matrices with observations from centre i
    x1fix <- model.matrix(~ 1 + ASA2 + xleef75 + fcharlgrp +
                            XTYPOK +xresadd + XCOMPLPRE,
                          data = x1[x1$id==i,])
    x2fix <- model.matrix(~ 1 + ASA2 + xleef75 + fcharlgrp +
                            XTYPOK +xresadd + XCOMPLPRE,
                          data = x2[x2$id==i,])
    x3fix <- model.matrix(~ 1 + ASA2 + xleef75 + fcharlgrp +
                            XTYPOK +xresadd + XCOMPLPRE,
                          data = x3[x3$id==i,])

    # function to be integrated
    integrand <- function(rho, pars, r1, r2){
      betas1 <- pars[3:10]
      betas2 <- pars[11:18]

      # write probabilities for each joint outcome in centre i
      # where 'x' is a vector of random effects to be integrated out
      px1 <- (1 - expit((x1fix %*% betas2) + r2))
      px2 <- (1 - expit((x2fix %*% betas1) + r1)) *
        expit((x2fix %*% betas2) + r2)
      px3 <- expit((x3fix %*% betas1) + r1) *
        expit((x3fix %*% betas2) + r2)


```

```
140       # the part of the likelihood to be integrated over
141
142       U <- chol(Sigma(rho, pars[1], pars[2]))
143
144       return(prod(px1)*prod(px2)*prod(px3)*
145                dmvnc(c(r1,r2), mu, U))
146     }
147
148     # integrate using gaussian quadrature
149     rule <- gaussHermiteData(6)
150
151     integral <- ghQuad(function(y){
152       sapply(y, function(y){
153
154         ghQuad(Vectorize(integrand, vectorize.args = "r1"), r2=y,
155                rho = rho, pars=pars,
156                rule=rule)
157     })
158     }, rule=rule)
159
160     # save integral value in centre i
161     centre.probabilities[i] <- integral
162
163   }
164   # sum over centres
165   loglik <- sum(log(centre.probabilities) )
166   loglik <- ifelse(loglik == -Inf, -999999, loglik)
167   return(loglik)
168
169 }
170
171 bootlik <- function(dat, k) {
172
173   bdat <- dat[k, ]
174
175   bmodel1 <- glmer(status ~ ASA2 + xleef75 + fcharlgrp +
176                      XTYPOK +xresadd + XCOMPLPRE + (1|id),
177                    data=bdat, nAGQ = 2,
178                    family=binomial(link="logit"),
179                    na.action=na.omit)
180
181   bmodel2 <- glmer(XBELOOP ~ ASA2 + xleef75 + fcharlgrp +
182                      XTYPOK +xresadd + XCOMPLPRE + (1|id),
183                    data=bdat,
184                    family=binomial(link="logit"),
185                    na.action=na.omit)
186
187   # fixed effect coefficients
188   bbetas1 <- bmodel1@beta
```

```
  bbetas2 <- bmodel2@beta

  # variance components
  bsigma1 <- VarCorr(bmodel1)$id[1,1]
  bsigma2 <- VarCorr(bmodel2)$id[1,1]


  x1b <- bdat[bdat$status=="alive" & bdat$XBELOOP=="no complication",]
  x2b <- bdat[bdat$status=="alive" & bdat$XBELOOP=="complication",]
  x3b <- bdat[bdat$status=="dead" & bdat$XBELOOP=="complication",]

  # remove observations with missing bdat
  x1b <- x1b[complete.cases(x1b),]
  x2b <- x2b[complete.cases(x2b),]
  x3b <- x3b[complete.cases(x3b),]


  stage1_pars <- c(bsigma1, bsigma2, bbetas1, bbetas2)

  opt.par <- optimize(loglikelihood, interval = c(-1, 1), pars = stage1_pars,
                      x1 = x1b, x2 = x2b, x3 = x3b, maximum = TRUE,
                      tol = .Machine$double.eps)
  c(opt.par$max, stage1_pars)
}


system.time( # results from bootstrap estimation
  b1_rho1 <- boot(data, bootlik, R = 1000, parallel = "multicore",
                  ncpus = 4)
)

b1_rho1
plot(b1_rho1)

# save results for tables
res1 <- data.frame(Start = c(NA, start_pars2), "Joint" = c(opt.par12$value,
                                                           opt.par2$par))

tab_names <- c("Loglikelihood", "Rho", "Sigma 1", "Sigma 2",
               names(model1$coef), paste(names(model1$coef), "_2", sep=""))

rownames(res1) <- tab_names

res1[2,1] <- NA
```

```
233
234 write.csv(res1, "results1.csv")
```

```
1
2 ## simulation study
3
4 # fixed effects for linear predictor
5 model1 <- glmer(status ~ ASA2 + xleef75 + fcharlgrp +
6                    XTYPOK +xresadd + XCOMPLPRE + (1|id),
7                  data=data, nAGQ = 2,
8                  family=binomial(link="logit"),
9                  na.action=na.omit)
10 model2 <- glmer(XBELOOP ~ ASA2 + xleef75 + fcharlgrp +
11                    XTYPOK +xresadd + XCOMPLPRE + (1|id),
12                  data=data,
13                  family=binomial(link="logit"),
14                  na.action=na.omit)
15 betas_start1 <- model1@beta
16 betas_start2 <- model2@beta
17
18 set.seed(10)
19
20 # create model matrix
21 MM1 <- model.matrix( ~ ASA2 + xleef75 + fcharlgrp +
22                        XTYPOK +xresadd + XCOMPLPRE + as.numeric(id),
      data = datasub)
23
24 N <- dim(MM1)[1]
25
26 # sample N patients from original dataset
27 subsamples <- sample(1:N, 20000)
28 MM1s <- MM1[subsamples,]
29
30 #covariance matrix
31 Sigma = function (rho, sigma1, sigma2) {
32   cov12 <- rho*sigma1*sigma2
33   matrix(c(sigma1^2, cov12, cov12, sigma2^2), nrow=2, byrow=T)}
34
35 covm <- Sigma(0.5,1,1) # rho = 0.5
36 raneffs <- rmvnorm(92, mean = rep(0,2), sigma = covm) #centre random
      effects
37
38 # fixed effects protion of linear predictor
39 simfix1 <-  MM1s[, -9] %*% betas_start1
40 simfix2 <-  MM1s[, -9] %*% betas_start2
41
42 df1 <- data.frame(MM1s) # gather into data frame
43
44
45
46 M <- length(simfix1)
47 # create linear predictors with gaussian noise
```

```
48 linpred1 <- simfix1 + raneffs[df1$as.numeric.id., 1] + rnorm(M, 0,
       1)
49 linpred2 <- simfix2 + raneffs[df1$as.numeric.id., 2] + rnorm(M, 0,
       1)
50
51 # simulate new observations
52 preds1 <- sapply(inv.logit(linpred1), function (x) rbinom(1,1,x))
53 preds2 <- sapply(inv.logit(linpred2), function (x) rbinom(1,1,x))
54
55 # gather simulated dataset
56 simdat <- data.frame(status = preds1, XBELOOP = preds2,
57                      datasub[subsamples,])
58
59 #check summary stats
60 p1 <- table(simdat$id, simdat$status)[,2]/rowSums(table(simdat$id,
       simdat$status))
61 p2 <- table(simdat$id, simdat$XBELOOP)[,2]/rowSums(table(simdat$id,
       simdat$XBELOOP))
62
63 cor(p1,p2)
64
65 table(simdat$status)[2]/sum(table(simdat$status))
66 table(simdat$XBELOOP)[2]/sum(table(simdat$XBELOOP))
67
68 #estimate using two-stage algorithm
69
70 system.time(
71   sim_rho1 <- boot(simdat, bootlik, R = 10, parallel = "multicore",
72                    ncpus = 4)
73 )
74 sim_rho1
```

8.3 R code for Chapter 6

JAGS code

```
1
2  model {
3  # Random effects logistic regression part of model
4  for( i in 1: M ) {
5  logit(rho[i]) <- theta[Z[i]]
6  y[i] ~ dbin(rho[i],n[i])
7  Z[i] ~ dcat(p[]) # integer variable
8  }
9  # Constructive DPP
10 #stick-breaking prior
11 p[1] <- r[1]
12 for (j in 2:(N-1)) {p[j] <- r[j]*(1-r[j-1])*p[j-1]/r[j-1]}
13 for (k in 1:(N-1)){ r[k] ~ dbeta(1,alpha)}
14 #assumption to ensure sum p[] is 1 Ishwaran truncation
15 ps <- sum(p[1:(N-1)])
16 for(k in N){p[k]<-1-ps}
17 # Baseline distribution
18 for(k in 1:N){theta[k] ~ dnorm(basemu,basetau) }
19 basemu~dnorm(0,0.01)
20 basetau<-pow(sigmaF0,-2)
21 sigmaF0~dunif(0,10)
22 # DPP parameter prior
23 alpha~dunif(0.3,10)
24 # Programming for calculating summary statistics #
25 for(i in 1:M) {for (j in 1:N) {
26 SC[i,j] <- equals(j,Z[i])
27 } }
28 # total clusters K#
29 for (j in 1:N) {cl[j] <-step(sum(SC[,j])-1)}
30 K<-sum(cl[])
31
32
33 # Random effects distribution mean#
34 for(i in 1:N){mean2[i]<-p[i]*theta[i]}
35 poptrue<-sum(mean2[])
36 # Random effects distribution variance#
37 for(i in 1:N){mom2[i]<-p[i]*theta[i]*theta[i]}
38 mom2.true<-sum(mom2[])
39 var.true<-mom2.true-(poptrue*poptrue)
40 #Number of Centres in same cluster#
41 for(i in 1:M){
42 for(j in 1:M){
43 equalsmatrix[i,j]<-equals(rho[i],rho[j])
44 }
45 equalsres[i]<-sum(equalsmatrix[i,])
46 }
```

```
47 }
```

R code

```
1
2
3 # Packages
4
5 library(rjags)
6 library(foreign)
7 library(doMC)
8 library(dclone)
9 library(ggplot2)
10
11
12 registerDoMC(4) # set number of workers for parallel computaion
13 getDoParWorkers()
14
15 # Structure data for jags
16 data <- read.spss("data_FTR_Shane.sav", use.value.labels=T, to.data.
       frame=T)
17
18 data$id <- factor(data$id)
19 data$status <- factor(data$status)
20
21 levels(data$id) <- 1:92
22 levels(data$XBELOOP) <- c("no complication","complication")
23 levels(data$status) <- c("alive","dead")
24
25 M <- length(levels(data$id))
26 N <- 52 # number of points for DP approximation
27
28 n <- table(data$id)
29 n <- as.numeric(n)
30
31 y <- table(data$id, data$status)[, 2]
32 y2 <- table(data$id, data$XBELOOP)[, 2]
33
34
35 jagsdat1 <- list(N = N, M = M, n = n, y = y)
36 jagsdat2 <- list(N = N, M = M, n = n, y2 = y2)
37
38 # Build model in jags
39
40 jags_death <- jags.model('Model2.bug',
41                          data = list(N = N, M = M, n = n, y = y),
42                          n.chains = 4,
43                          n.adapt = 100)
44
45 update(jags_death, 100000) # jags model, takes a long time to update
       4 chains
46
```

```
47 js1 <- coda.samples(jags_death,
48                       c('alpha', 'p', "rho", "K"),
49                       100000)
50 plot(js1)
51
52 params <- c('alpha', 'p', "rho", "K") # parameters to monitor
53
54
55
56 # Structure data for jags
57 data <- read.spss("data_FTR_Shane.sav", use.value.labels=T, to.data.
       frame=T)
58
59 data$id <- factor(data$id)
60 data$status <- factor(data$status)
61
62 levels(data$id) <- 1:92
63 levels(data$XBELOOP) <- c("no complication","complication")
64 levels(data$status) <- c("alive","dead")
65
66 M <- length(levels(data$id))
67 N <- 52
68
69 n <- table(data$id)
70 n <- as.numeric(n)
71
72 y <- table(data$id, data$status)[, 2]
73 y <- table(data$id, data$status)[, 2]
74
75 jagsdat1 <- list(N = N, M = M, n = n, y = y)
76
77 # Build model in jags
78
79 params <- c('alpha', 'p', "rho", "K") # parameters to monitor
80
81 # use jags.parfit from dclone for parallel computations
82 system.time( # death
83   jp1 <- jags.parfit(4, jagsdat1, params, 'Model2.bug', n.chains =
       4,
84                      n.adapt = 1000, n.update = 1000, thin = 1, n.
       iter = 100000)
85 )
86
87 jagsdat1.2 <- list(N = 27, M = M, n = n, y = y)
88 system.time(
89   jp1.3 <- jags.parfit(4, jagsdat1.2, params, 'Model2.bug', n.chains
       = 4,
90                      n.adapt = 1000, n.update = 1000, thin = 10, n.
       iter = 200000)
```

```
)


# plot some summary statistics using coda
plot(jp1.2[, 2:3])
plot(jp1.2[, 29:30])

# check diagnostics
plot(jp1.2, ask = TRUE)
gelman.plot(jp1.2)
gelman.diag(jp1.2, multivariate = FALSE)
geweke.diag(jp1.2)
raftery.diag(jp1.2)
1 - rejectionRate(jp1.2)


summ_stats1 <- summary(jp1.2)$statistics

rownames(summ_stats1)

# descriptive plots
par(mfrow = c(1,1))
hist(summ_stats1[55:146, 1]) # centre effects
plot(density(summ_stats1))

rho_means1 <-summ_stats1[30:121, 1]
cent_mean1 <- data.frame(means = rho_means1) # centre means
quants1 <- summary(jp1.3)$quantiles[ , c(1, 5)] # 2.5th and 97.5th
    quantiles

p1 <- ggplot(cent_mean1, aes(means)) # plot with ggplot
p1 + geom_histogram(aes(y=..density..), fill = "blue", col = "
    lightblue") +
  geom_density(col = "red")


# Plot pointwise intervals
ints1 <- quants1[30:121, ]
ci_ind1 <- order(n) # order by size of centre
ci_lines1 <- ints1[ci_ind1, ]

fp1_dat <- data.frame(n, y = rho_means1,
                      lower = ints1[, 1], upper = ints1[ , 2])
fp1_dat <- fp1_dat[ci_ind1, ] # order all data
fp1 <- ggplot(fp1_dat, aes(n, y))
fp1 + geom_point(col = "green", size = 2.5) +
  geom_line(aes(n, lower), col = "blue", se=FALSE, lwd = 0.2) +
  geom_line(aes(n, upper), col = "blue", se=FALSE, lwd = 0.2) +
  theme_bw() + xlab("Patient Volume") + ylab("Mortality rate")
```

```
138
139 fp1 + geom_point(col = "green", size = 2.5) +
140   geom_smooth(aes(n, lower), col = "blue", se=FALSE, span= 0.9, lwd
      = 0.4) +
141   geom_smooth(aes(n, upper), col = "blue", se=FALSE, span= 0.9, lwd
      = 0.4) +
142   theme_bw() + xlab("Patient Volume") + ylab("Mortality rate")
143
144
145
146 ## complications
147 system.time(
148   jp2 <- jags.parfit(4, jagsdat2, params, 'Model2_complication.bug',
        n.chains = 4,
149                      n.adapt = 1000, n.update = 1000, thin = 1, n.
      iter = 100000)
150 )
151
152 jagsdat2.2 <- list(N = 27, M = M, n = n, y2 = y2)
153 system.time(
154   jp2.4 <- jags.parfit(4, jagsdat2.2, params, 'Model2_complication.
      bug', n.chains = 4,
155                      n.adapt = 2000, n.update = 2000, thin = 1, n.
      iter = 200000)
156 )
157 class(jp2)
158
159 summ_stats2 <- summary(jp2.4)$statistics
160 quants2 <- summary(jp2.4)$quantiles[ , c(1, 5)]
161 rownames(summ_stats2)
162 rho_means2 <- summ_stats2[30:121, 1]
163 cent_mean2 <- data.frame(means = rho_means2)
164
165 summ_stats2.2 <- summary(jp2.2)$statistics
166 plot(jp2.4[, c(2:3)])
167 plot(jp2.4[, c(29:30)])
168
169 p2 <- ggplot(cent_mean2, aes(means))
170 p2 + geom_histogram()
171 p2 + geom_histogram(aes(y=..density..), fill = "blue", col = "
      lightblue") +
172   geom_density(col = "red") + xlab("Centre random effects")
173
174
175 ints2 <- quants2[30:121, ]
176 ci_ind2 <- order(n)
177 ci_lines2 <- ints2[ci_ind2, ]
178
179 fp2_dat <- data.frame(n, y = rho_means2,
```

```
                        lower = ints2[, 1], upper = ints2[ , 2])
fp2_dat <- fp2_dat[ci_ind2, ]
fp2_dat
fp2 <- ggplot(fp2_dat, aes(n, y))
fp2 + geom_point(col = "green", size = 2.5) +
  geom_line(aes(n, lower), col = "blue", se=FALSE, lwd = 0.2) +
  geom_line(aes(n, upper), col = "blue", se=FALSE, lwd = 0.2) +
  theme_bw() + xlab("Patient Volume") + ylab("Complication rate")

fp2 + geom_point(col = "green", size = 2.5) +
  geom_smooth(aes(n, lower), col = "blue", se=FALSE, span= 0.9, lwd
    = 0.5) +
  geom_smooth(aes(n, upper), col = "blue", se=FALSE, span= 0.9, lwd
    = 0.5) +
  theme_bw() + xlab("Patient Volume") + ylab("Complication rate")


# diagnostics
jp2_burn <- window(jp2.2, start = 55000)
summary(jp2_burn)$statistics
plot(window(jp2.2[, 1:4], start = 55000))
autocorr.plot(window(jp2.2[, 25:28], start = 55000))

gelman.diag(jp2_burn, multivariate = FALSE)
geweke.diag(jp2_burn)
raftery.diag(jp2.4)
1 - rejectionRate(jp2_burn)
```