THE UNIVERSITY OF CALGARY

Statistical methods for the meta-analysis of diagnostic tests from studies with varying reference standards

by

Jian Kang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

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

DEPARTMENT OF COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

June, 2009

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Statistical methods for the meta-analysis of diagnostic tests from studies with varying reference standards" submitted by Jian Kang in partial fulfillment of the requirements for the degree of Ph.D.

Dr. Rollin Brant
Department of Community Health Sciences

Dr. William Ghali
Department of Community Health Sciences

Dr. Michael Eliasziw
Department of Community Health Sciences

Dr. Karen Kopciuk
Department of Mathematics and Statistics

Dr. Therese Stukel
Dartmouth Medical School USA and University of Toronto

6/12/09
Date

Abstract

In an ideal study of diagnostic tests, the new diagnostic test and the gold standard should be applied to each patient suspected of the disease. But the reality is that the gold standard is not applicable due to many limitations, such as its invasiveness, high cost, or technical challenge. This project was motivated by the meta-analysis of diagnostic tests for deep vein thrombosis (DVT). In diagnosing DVT, the gold standard is venography, which is invasive and often not applicable to every patient. In fact, it was applied only in a small number of studies. A concurrent reference is ultrasonography, which is not risky to patients and has well known diagnostic characteristics. D-dimer is a new test of interest for DVT. Among the studies of d-dimer, a substantial amount of studies applied ultrasonography as the reference. The aim of this project is to develop statistical methods to estimate the diagnostic performance of d-dimer by synthesizing studies using both references.

By assuming known values of sensitivity and specificity of ultrasonography, maximum likelihood estimation was applied to acquire estimates of coefficients in the log-linear model. When the sensitivity and specificity of ultrasonography were not available, data from a systematic review of ultrasonography [6] were employed. Two approaches to estimating the diagnostic accuracy of d-dimer were compared: admitting the difference between the two references and ignoring the difference. Taking into account heterogeneity across studies, the log-linear model was fitted with random disease prevalence and random association between d-dimer and gold standard. Two algorithms, the Gaussian Hermite integration and the Gibbs sampling, were applied to derive estimates in the random effects model. In the model with two random

effects, a new design matrix of random effects based on -1 1 contrast was applied to improve estimates. This project was a novel application of the Gaussian Hermite integration and the Gibbs sampling in the meta-analysis of incomplete multinomial data.

In summary, this project provided statistical methods for the meta-analysis of diagnostic tests in the absence of complete data. Results between the test of interest and an imperfect reference can be used to estimate the diagnostic performance of the test of interest, provided that appropriate adjustments were performed. This finding has a strong impact in the literature of diagnostic tests: the gold standard may not be necessary in assessing a new test. This is a real milestone in diagnostic tests where the gold standard diagnosis is invasive or harmful to patients.

Acknowledgements

It is my pleasure to thank the many people who made this thesis possible.

It is difficult to overstate my gratitude to my Ph.D. supervisor, Dr. Rollin Brant. With his enthusiasm, his inspiration, and his great efforts to explain things clearly and simply, he helped to make biostatistics fun for me. Throughout my thesis-writing period, he provided encouragement, sound advice, good teaching, and lots of good ideas. I would have been lost without him. Throughout my Ph.D studies, Dr. Brant has become a role model to me in applied statistics. His great ideas in biostatistics have inspired me the most. Besides the great help Dr. Brant has provided on my thesis, he also introduced opportunities for me to involve in various clinical projects. The experience with these projects gave me much more instructions on how to be a successful biostatistician than those from the course work.

I am grateful to the extremely helpful comments from Dr. William Ghali and Dr. Michael Eliasziw with respect to the clinical and statistical aspects of this thesis. Their inputs as clinicians and statisticians provided me more insights to the research in Biostatistics, which is crucial and beneficial to my career. The original data for analysis from Dr. Russell Hull are also acknowledged.

I would like to thank the many people who have taught me mathematics and statistics. My high school math teachers and undergraduate teachers in China have provided me with solid mathematical preparations for graduate studies. My graduate teachers in statistics, professors from Bowling Green State University and the University of Iowa, led me into the field of applied statistics and provided me with a good basis of pursuing Ph.D in Biostatistics. I would like to thank Dr.Michael

Jones from the University of Iowa in particular for the two serial courses: Theory of Biostatistics, which equipped me with solid theoretical tools in constitute of this thesis.

I am indebted to many students and colleagues for providing a stimulating and fun environment in which to learn and grow. I am especially grateful to Dr. Joseph Amuah, who often had great discussions with me in statistics, and Mr. Dave Schulz, who provided a nice and efficient computer platform, the ecogrid server at the University of Calgary, for all the simulations in this thesis.

I am grateful to the faculty and staff members at the Department of Community Health Sciences, for helping the Department to run smoothly and for assisting me in many different ways. Dr. Marilyn Hebert, Crystal Elliott, Florence Yang, and Merl Dalip deserve special mention. In particular, I would like to thank Vicki Stagg. She has provided me with different kinds of support since I arrived in Calgary. She is always caring and resourceful.

I wish to thank my sister and her husband for being always supportive on my Ph.D studies.

Lastly, and most importantly, I wish to thank my parents, Teimei Wen and Weimin Kang. They bore me, raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

Glossary

GLM: Generalized Linear Model

GLMM: Generalized Linear Mixed Model

Silver standard: ultrasonography

Gold standard: venography

DU: d-dimer versus ultrasonography

DV: d-dimer versus venography

UV: ultrasonography versus venography

S_d: sensitivity of d-dimer against venography

C_d: specificity of d-dimer against venography

S_u: sensitivity of ultrasonography against venography

C_u: specificity of ultrasonography against venography

S_{d_1} and C_{d_1}: sensitivity and specificity of d-dimer from the model treating ultrasonography the same as venography

S_{d_2} and C_{d_2}: sensitivity and specificity of d-dimer from the model treating ultrasonography different from venography

MCMC: Markov Chain Monte Carlo

p_{duv}: cell probability in the 2×2×2 contingency table. d=levels of d-dimer, u=levels of ultrasonography, v=levels of venography.

$p_{du.}$: cell probability summing over levels of venography

$p_{d.v}$: cell probability summing over levels of ultrasonography

$p_{.uv}$: cell probability summing over levels of d-dimer

m_{duv}: cell counts in the 2×2×2 contingency table

$m_{du.}$: cell counts summing over levels of venography

$m_{d.v}$: cell counts summing over levels of ultrasonography

$m_{.uv}$: cell counts summing over levels of d-dimer

$table_{du}$: observed table of d-dimer versus ultrasonography

$table_{uv}$: observed table of ultrasonography versus venography

$table_{dv}$: observed table of d-dimer versus venography

$p_{duv|\gamma_1,\gamma_2}$: cell probability conditional on random effects γ_1 and γ_2

Adjusted model: the model adjusting for the difference between the silver standard and the gold standard

Unadjusted model: the model ignoring the difference between the silver standard and the gold standard

1REM: the model with random disease prevalence only

2REM: the model with the random disease prevalence and random association between the test and the gold standard.

Table of Contents

List of Tables

List of Figures

Chapter 1

Background

1.1 Sensitivity and specificity in diagnostic tests

In assessing the clinical performance of diagnostic tests, two widely acknowledged measurements are sensitivity and specificity. The sensitivity is defined as the rate of positive test results in a group of patients who have the disease. The specificity is the proportion of negative test results in healthy patients. An ideal diagnostic test discriminates between diseased and healthy patients without error. Let α and β denote the false negative and false positive rates, then the sensitivity and specificity are 1-α and 1-β, respectively. Two other measurements that are highly related to sensitivity and specificity are the positive and negative predictive values. In reality, the true disease status of a patient is often not known at the time the diagnostic test was applied. Given a positive test result, what is the chance that this patient has the disease? This is referred to as positive predicted value (PPV) of the test. Similarly, the probability that the patient has no disease if the test result is negative is called the negative predicted value (NPV). The relationship between positive and negative predicted value with sensitivity and specificity is addressed by the following expressions.

Notations: s = sensitivity = P(+|disease) c = specificity = P(-|no disease)

P(Disease) = prevalence of disease.

$$PPV = \frac{s \times P(disease)}{s \times P(disease) + (1-c) \times [1 - P(disease)]}$$

$$NPV = \frac{c \times [1 - P(disease)]}{c \times (1 - P(disease)) + (1-s) \times P(disease)}$$

Sensitivity and specificity of a test provide indications of how well the test identifies diseased and healthy patients, respectively. As a function of sensitivity and specificity, the likelihood ratio is an alternative measure of test performance. It is the ratio of the probability of a test result in diseased patients to the probability of the same result in healthy patients. The positive likelihood ratio (PLR) is defined as the ratio of probability of test positive in diseased patients over test positive in non-diseased patients. That is, PLR = sensitivity/(1-specificity). Similarly, the negative likelihood ratio (NLR) = (1-sensitivity)/specificity. Both quantities describe how many times more likely the same results are in diseased patients than in healthy patients. Values above 5 of PLR and below 0.2 of NLR give strong diagnostic evidence [75].

1.2 Contingency tables and log-linear model

Categorical variable is defined as a variable that classifies an object into at least two mutually exclusive categories. Results from a diagnostic test, for example, are classified as positive and negative. For a long time in the literature of categorical data analysis, classification tables for two or more variables have been used, namely "contingency tables". The rows and columns in a contingency table represent levels of the two factors. When a third factor is involved, one 2×2 contingency table at

each level of the third factor is constructed. Similar extensions can be applied to higher dimensions of contingency tables. Each cell count represents the number of outcomes that fall in the category, which is a combination of levels from all factors. The contingency table carries critical information on the association among factors. In the analysis of two-dimensional tables, the Chi-squared test is well known to draw conclusions on the relationship between the two factors. Fisher's exact test is an alternative when cell counts are small.

The structure and methods for the analysis of two-dimensional tables are generally not complicated. The log-linear model has been introduced to analyze high-dimensional contingency tables in extensive literature. When dealing with positive outcomes, the classic linear model is normally considered to be unsatisfactory. The reason is that a certain combination of parameter and covariate values may produce negative values of the outcome. Even if the linear combination may be found to be adequate over the range of the data, extrapolation of the results is often questionable. The logarithm function provides the conversion of a positive number to a number on the real line. This property makes the log-linear model an intuitive candidate for the analysis of contingency tables. In addition, the log-linear model gives rise to a multiplicative association between the response and explanatory variables because of the log function. This association is not seen in the classic linear model.

For low dimensional contingency tables with a small number of parameters, maximum likelihood estimates (MLE) can be obtained algebraically by solving likelihood equations. Difficulties increase as the number of parameters increases. For example, in a three-way contingency table, the conditional independence model involves five parameters. Estimates of these parameters cannot be written out in closed forms.

The iterative proportional fitting procedure has been a traditional method to acquire MLE for log-linear models [12, 20, 89]. Alternatively, the Newton-Raphson algorithm is a general approach to derive MLE. Details of this algorithm are covered in various statistics textbooks.

1.3 Generalized linear model (GLM), generalized linear mixed model (GLMM) and Bayesian analysis

The log-linear model imposes a logarithm relationship between the mean of the outcome and the linear combination of explanatory factors. The model which includes any functional form of the relationship between the outcome and the explanatory factors is defined as the "generalized linear model" [66] (GLM). It takes the form of $g(\mu)=X\beta$, where X is the design matrix and β is the vector of coefficients. The function g is known as the link function. Different link functions give rise to different models. For example, the logistic regression model differs from linear regression in the logit link function on the outcome variable. Explanatory variables in GLM are often regarded as fixed effects.

The inference of the generalized linear model follows a frequentist approach. Under the frequentist approach, the parameter in the likelihood function is an unknown but fixed number. By taking a sample which represents characteristics of the population, the likelihood function can be constructed with respect to unknown parameters. Maximizing the likelihood function produces the best "guess" of the unknown parameter. For a Bayesian approach, however, the foundation of the theory is different. The parameter is regarded as a random variable, which has its own distribution with

mode and curvature parameters. By assuming a probability distribution for the unknown parameter, the knowledge of the parameter can be updated after obtaining a sample from the population. The updated information is referred to as the posterior distribution of the parameter. The updated value is called the posterior estimate of the parameter. The Bayesian posterior estimates and credible sets are analogous to MLE and confidence intervals in the frequentist context.

In a clinical setting, it is sensible to treat some risk factors as fixed but unknown values. For example, the overall effect of a drug in men compared to that in women can be fixed but unknown. In some situations, however, this may not be true, especially at the individual subject level. For example, it makes more sense to treat the effect of a drug on a particular patient as a random variable, which follows a distribution, than as a fixed value. Combining the concepts of fixed and random effects, the GLM can be extended to the generalized linear mixed model (GLMM). In this model, a prior distribution is specified on the random effects. The GLMM encompasses advantages from both GLM and the random effects model. By means of the link function, GLMM allows the analysis of a variety of outcome measurements, discrete or continuous. It enables the accommodation of non-normally distributed responses and specifies a possibly non-linear link between the mean of the response and the predictors. With respect to the distribution of random effects, the normal density is a conventional choice.

The development of computational techniques to solve GLMM, on the other hand, has not been satisfactory. The complexity of GLMM produces difficulties in solving likelihood equations analytically. The expectation-maximization (EM) algorithm was proposed as an advantageous technique for maximum likelihood and restricted

maximum likelihood estimations [2, 61]. The major advance of EM algorithm is that it ensures an elevated likelihood at each iteration, although convergence may be extremely slow. The Gaussian Hermite quadrature (GHQ) is often used for numerical approximations of integrals with Gaussian kernels, i.e., e^{-x^2}. The marginal likelihood of fixed effect parameters in GLMM has to be derived by integrating out the random effects, which often makes the derivation of marginal likelihood analytically intractable. Integrations using Gaussian Hermite quadratures have preferable accuracy in analyzing GLMM when the number of random effects per cluster is small. In recent decades, the Markov Chain Monte Carlo (MCMC), a typical approach for Bayesian analysis, has become valuable in analyzing GLMM. It has become extremely popular for the analysis of complex statistical models. The MCMC procedure reduces the computational complexity of high dimensions to a sequence of much lower ones.

1.4 Application of methods

Thrombosis is an abnormality of an endovascular clot at an inappropriate place and time in the blood [77]. The venous thromboembolism is reflected by alterations in blood flow, in the coagulability of blood, and in the vessel wall [77]. It has clinical indications relating to several diseases, such as myocardial infarction, stroke and cardiovascular disorders, which are leading causes of death in industrialized countries [77]. Deep vein thrombosis (DVT) is a common but often undiagnosed thromboembolic disease. The clinical examination of DVT is not satisfactory. Among suspected patients, venography tests positive for 42% and negative for 58% [77]. Accurate

diagnosis of DVT has significant clinical implications. Patients with false positive results may receive unnecessary or harmful treatments. False negative results of patients may delay their getting necessary and effective treatments. Although the importance of detecting thromboembolic diseases has been widely acknowledged, a perfect diagnostic tool has yet been developed. Diagnostic tests with high accuracies are favorable but frequently they are too expensive or hazardous to be used on high volume population-based screening.

D-dimer, a new diagnostic tool of DVT, has been developed since 1980s, although its error rates have not been confirmed. Error rates can be estimated directly if d-dimer can be applied to the patients whose true disease states are known, but this is usually not feasible. D-dimer is highly sensitive at an elevated level (< 500ug/L) [77]. Studies suggest that pulmonary embolism is unlikely if plasma d-dimers are lower than 500ug/L, which is the most commonly used cutoff [27]. Although there has been extensive research on evaluating d-dimer, the role of d-dimer in diagnosing DVT is not clear because of the presence of multiple assays, laboratory testings and variability of assays [18].

Venography has perfect diagnostic reliabilities in distinguishing disease and healthy patients. It is the *gold standard* in the diagnosis of DVT. It is clinically impractical, however, to apply it to every patient suspected of DVT because it is invasive. Sometimes the test was given to the patients with negative results in a previous test. In this situation, patients with positive results did not go through the reference test at all. On the other hand, ultrasonography, as a non-invasive reference, became an alternative for the diagnosis of DVT in various studies. Its sensitivity and specificity have been well established in the literature. A large number of studies evaluating

the sensitivity and specificity of d-dimer used ultrasound as the reference standard [40, 97]. However, biases on the diagnostic characteristics of d-dimer in these studies were expected if the error rate of ultrasonography was not taken into account. Walter [87] and several researchers have investigated the effects of known error rates of the reference test on the estimates of the new test. Even with small error rates from the reference test, biases in the estimation of the test of interest are substantial [63, 78].

Ideally, if all three tests were applied to each patient, the complete three-dimensional contingency table would be available to estimate characteristics of the test of interest. In practice, however, it is almost impossible to have results from all three tests on each patient, particularly when tests are expensive, time consuming, or invasive. In these circumstances, statistical adjustments have to be applied to obtain corrected estimates when the imperfect reference is used. One of the major purposes of this project is to combine studies using different references to estimate diagnostic characteristics of d-dimer. There has been a variety of clinical areas that experience the same problem as in diagnosing DVT. Two or more references were applied in different studies to diagnose diseases, one is error-free and the others are imperfect but non-invasive. An example is given below to elaborate the importance and potential application of the methods proposed in this project.

Coronary artery disease (CAD) is the leading cause of death across the western world. Prevention and early detection of CAD are critical in clinical practice. X-ray coronary angiography is the current gold standard for identifying clinically significant coronary artery disease [103]. But it is invasive. The commonly used noninvasive reference for CAD is the myocardial perfusion imaging (MPI). In a study that eval-

uated the diagnostic performance of electrocardiograms (ECG), the MPI instead of the gold standard was used as the reference [98]. The authors argued for the use of this reference based on clinical considerations. The stress test results introduced bias on referral to coronary angiography, which was usually the second diagnostic test after ECG. Using MPI as the reference prevented this problem because results from ECG and MPI were acquired at the same test. The advantage of using an accurate and noninvasive diagnosis on CAD has been widely acknowledged. One of the advantageous tools is computed tomography (CT). The clinical usefulness of computed tomography has been well assessed [19] in the diagnosis of CAD. The methods in this project intended to analyze pair-wise marginal tables from three diagnostic tools, which were the gold standard, silver standard, and the test of interest. The gold standard of diagnosing CAD is coronary angiography, and silver standard is MPI. CT is the diagnostic test of interest. In this case, marginal tables of the test with the gold standard and with the silver standard may be collected, respectively. This allows potential application of methods from this project to this setting.

The methods proposed in this project can be applied to the above clinical dilemmas, especially when the gold standard involves invasive or complex procedures, such as biopsy or surgery. Furthermore, a variety of applications may be of interest. First of all, models proposed here may not be limited to reference tests. This corresponds to evaluating the effect of several competing treatments or procedures for a given disease. In the example of coronary artery disease above, competing non-invasive tests are available, such as stress echocardiography [52] and coronary magnetic resonance angiography [103]. Comparisons among these tests are potentially viable using the method suggested in this project. Results of comparisons contribute to construct a

more effective but less invasive diagnostic procedure than using a single diagnostic test. This constitutes the second interesting application of the project. Last but not the least, the models in this project may provide useful information on the evaluations of test performance against the cost of each test. This information may be of interest to health economists to model cost-effectiveness.

Chapter 2

Literature Review

2.1 Meta-analysis and diagnostic tests

2.1.1 Methods in the meta-analysis of diagnostic tests

Diagnostic tests are an active research area in the medical sciences. In particular, the accuracy of diagnostic tests has been the center of this research area for decades. As various studies evaluating the diagnostic accuracy of different tests increases, the meta-analysis becomes important for summarizing the findings. The summary performance measures provided by systematic reviews and meta-analysis of diagnostic tests play an important role in clinical and health policy decision making on the usage of diagnostic tests. As such, methodologies and guidelines for meta-analysis evaluating diagnostic tests are proposed and discussed extensively in the literature [8, 39, 45, 83]. In the following sections, discussions on some of the most commonly applied techniques are presented.

Summary receiver operational characteristics (SROC) curve

In assessing the performance of a single diagnostic test, the receiver operating characteristic (ROC) curve [34] is widely recognized. Graphically, ROC is the plot between sensitivity and 1-specificity on different thresholds of positive diagnosis. The overall measure of accuracy of the test is the area under the ROC curve. The larger the area, the higher the values of sensitivity and specificity are. Poor tests have curves

close to the diagonal line, where the area under the ROC curve is 0.5. In the case of summarizing several diagnostic tests, the summary ROC (SROC) curve has been proposed and widely applied to account for different positive diagnosis thresholds across studies [38, 47, 53, 59, 79]. The SROC curve is constructed from a regression model between the log of diagnostic odds ratio (DOR) and the test positivity criterion (TPC). The DOR and TPC are defined as the following.

$$DOR = log\frac{s}{1-s} - log\frac{1-c}{c}$$
$$TPC = log\frac{s}{1-s} + log\frac{1-c}{c}$$

The regression model proposed by Moses [47] takes the following form.

$$DOR = \alpha + \beta \times TPC$$

Using the sensitivity and specificity estimates from each study in the above model, the estimates of α and β can be obtained by ordinary regression approach. The SROC curve is then constructed by plotting the sensitivity against 1-specificity on the original scale. The transformation of the above model from the logit scale to the original scale was provided by Moses [47] as the following.

$$sensitivity = \frac{1}{1+e^{-\alpha/(1-\beta)} \times (\frac{specificity}{1-specificity})^{(1+\beta)/(1-\beta)}}$$

An alternative approach to constructing an SROC curve is based on true positive and false positive estimates [42]. Rutter and Gatsonis [10] extended this approach to a hierarchical model which accounted for within- and between- study variations. Walter derived properties of the SROC and provided standard errors for the area

under the curve [86]. These approaches, however, did not take into account the errors in the estimates of sensitivity and specificity from each study. In the ordinary regression model, the independent variable was measured without error. The test positive criterion (TPC) is a function of the estimated sensitivity and specificity from each study. Errors in the estimates are expected. Application of the ordinary regression model using the TPC as predictors does not take into account the error in the estimates. This is the major disadvantage of using the summary ROC curve in the meta-analysis of a diagnostic test.

The logit models

In applied statistics, binomial and Poisson responses have been the center of research in the 1990s. Multinomial responses received less development with the majority of the research focused on ordinal data using logit and probit links for cumulative probabilities [13, 16, 32, 50]. Daniels and Gatsonis applied the baseline-category logit in a hierarchical Bayesian model for cluster multinomial data [54]. Hartzel presented a general approach for logit random effects modeling on clustered multinomial responses [31]. These approaches included the Gibbs sampling from the Bayesian standpoint and the maximum likelihood estimation using Gaussian quadratures or the expectation maximization (EM) algorithm in the frequentist context. The normality structure for random effects was assumed. The Poisson log-linear model was advocated by Chen to analyze multinomial data [107].

In the meta-analysis of comparing diagnostic tests, the logit model was proposed by Siadaty [58, 59]. Diagnostic odds ratios can be estimated directly from the logit model, which enables the derivation of the summary ROC curve for each study. Com-

parisons of the area under the curve (AUC) are derived to compare diagnostic tests. Advantages of this model with random effects include the flexibility to allow for missing values and different sample sizes, the ability to adjust for confounding factors and correlations within and between studies, and easy extension to accommodate individual patient data [59]. Similar approaches in the meta-analysis of diagnostic tests are widely available in various statistical software, such as SAS (genmod procedure), R (function geese), and STATA (commands xtgee) [58, 59]. The major concern from Siadaty's approach was that several competing tests reported the tables of the test versus the gold standard. Each test was evaluated against the same gold standard only. The main purpose of this project, however, is to incorporate a large number of studies using the imperfect reference due to the fact that the gold standard was invasive and not applied in many studies. The logit model is not directly applicable to the analysis in this project if one intends to include different references across studies. Alternatively, the meta-analysis may be restricted to studies using the gold standard only. The number of such studies, however, is very small.

2.1.2 Analysis that accounts for different reference standards

In diagnostic tests, differences in criteria to define positive and negative results, subjective assessments of endpoints, and patient conditions, are major reasons for different test results. Many researchers have addressed the clinical importance of including heterogeneity among studies in meta-analysis [17, 45, 92]. However, attentions to the difference in reference standards applied in each study have not been received in the meta-analysis of diagnostic tests. Most meta-analysis of diagnostic tests emphasize the method of combining summary statistics, sensitivity and specificity, or functions

of these two components. Accounting for different references should be recognized as the major concern of pooling results from different studies. Ignoring this information produces biased conclusions about the accuracy of the diagnostic test [63, 78].

Among the reviews of diagnostic test in cancer, 53% of them included studies from multiple reference tests, only 14% of them included studies from a single reference test, and 33% of the reviews did not report a reference test [83]. Although heterogeneity in studies due to different reference tests was addressed in the literature [38], statistical solutions to this problem were not concrete. Walter and Irwig [88] conducted a comprehensive review on estimations in misclassified categorical data. They elaborated on estimation in different scenarios which corresponded to different combinations of the number of tests per individual and the number of populations. The minimum number of observers (diagnostic tests) for the identifiability of all parameters is 3 per individual for any number of populations. In other words, patients have to go through at least 3 diagnostic tests in order to identify all parameters. A latent class analysis using EM algorithm [2] was proposed to improve estimates of the SROC curve when the reference was subject to error [87]. Improvements using a Bayesian approach were also discussed by several authors [46, 80]. But all of these approaches aimed at corrections either when only one type of table was used, i.e., the test of interest versus "silver standard", or in the situation where each individual received all tests. De Bock [24] and colleagues suggested estimation via the EM algorithm when there were at least two types of marginal tables, i.e., not all the tests were applied to each subject. This approach, however, assumed the sensitivity and specificity of each test were the same across studies. This was not true if studies used different thresholds of positivity. In our study, variations in the characteristics

of the test among different studies were taken into account in the model via random effects.

This project drew its inspiration from the meta-analysis of diagnostic tests for deep vein thrombosis. The problem arising from different reference standards not only affected clinical estimations of d-dimer but also had significant impacts on the meta-analysis. It was difficult to conclude anything about the sensitivity and specificity of d-dimer in the presence of different references. This gave rise to difficulties in the meta-analysis of the diagnostic performance of d-dimer. Conventional meta-analytic methods cannot be applied directly to combine data from the *gold* and *silver* standard by ignoring the difference between these two references. Tables from pair-wise combinations of the three tests, i.e., marginal tables, were incorporated to acquire accurate estimates and standard errors on the test of interest. Making use of available data from all three marginal tables was statistically preferred over using data from the gold standard alone. When characteristics of the silver standard become well established, sensitivity and specificity of the test of interest can be estimated using tables between the test and the silver standard. This indicates that tables with silver standard carry useful information on the test of interest. According to the scientific philosophy, all relevant and available evidence should be included in the meta-analysis.

2.2 Misclassification in epidemiology

From an epidemiological perspective, the problem discussed above can be regarded as exposure misclassification. Diagnostic results from the gold standard are analo-

gous to true classifications, whereas results from the silver standard, ultrasound, can be viewed as misclassifications. In a generalized definition, Walter and Irwig [88] named it an "observation", which may refer to a diagnostic test, a different data source, or a different occasion to apply the same method. Tables from studies using ultrasonography as the reference were misclassified. Consequently, estimates of the test of interest from these studies were biased if adjustments were not made. It was shown that bias was substantial even with small diagnostic imperfection from the silver standard [63, 78]. The magnitude of bias depended on the diagnostic characteristics of ultrasonography. Different approaches to adjust bias from exposure misclassification have been proposed in the literature and summarized below.

2.2.1 Conventional approaches

It has been long recognized that measurement errors were among the major weakness of epidemiological studies. Initially, understanding the effects of measurement errors on exposure-disease relationship was the focus of methodological research. The two patterns of errors have been well known as differential and non-differential misclassification of exposure. If the rates of misclassification in the two exposure groups are the same, it is non-differential. Otherwise, if the rate of misclassification depends on the disease status, it is differential. It is well known that non-differential misclassification in exposure generally produces bias in the odds ratio toward the null value [44]. The uncertainty about the direction of exposure-disease association, however, can increase in non-differential misclassification [64]. Differential misclassification, on the other hand, can affect the odds ratio in either protective or harmful direction. A large collection of literature is devoted to the corrections of bias due to misclassifi-

cation. The maximum likelihood estimation (MLE) was a conventional approach to solve the likelihood equation in the presence of misclassification. The matrix method and inverse matrix method were proposed by several authors for a more straightforward approach than MLE to correct bias from misclassification [78, 55]. The matrix method has been recommended by textbooks [44] and the variance estimation was given by Greenland [82]. In the matrix method, the data were partitioned into two samples. One was regarded as the validation study and the other was regarded as the main study. Expected cell counts in the observed table were a function of expected cell counts in the unobserved table and misclassification parameters (sensitivities and specificities). By replacing the expected cell counts with corresponding observed counts, and using the validation study to estimate misclassification parameters, the cell counts in the unobserved table can be estimated. The corrected log odds ratio was then calculated by the estimated cell counts.

In 1990, Marshall [78] introduced the inverse matrix method for corrections based on predictive values. Similar to the matrix method, information in the unobserved table, represented by predicted probabilities instead of cell counts, was a function of information in the observed table and the predictive values. The predictive values were estimated from the validation data. The corrected log odds ratio was calculated based on estimated probabilities in the unobserved table. Marshall also improved estimations in the matrix method by replacing cell counts with probabilities but keeping sensitivity and specificity as misclassification parameters [78].

Morrissey and Spiegelman [55] compared efficiencies of the three methods and found that: the inverse matrix was more efficient than the matrix method for differential misclassification and the MLE was more efficient than the matrix method for

non-differential misclassification. Several authors have proposed methods to correct for non-differential misclassification of exposure [85, 14, 104, 28]. Flanders et al [101] considered adjustments for differential misclassification of exposure with respect to disease status. They estimated the exposure and disease relationship by a pooled stratum specific odds ratio. Kosinski and Flanders [3] proposed a logistic regression approach via EM algorithm to correct estimates in the case of exposure misclassification. This approach required two imperfect tests but not the gold standard. Note that the above methods focused on adjustments in a single study with imperfect reference rather than combining the results from different studies. In other words, the adjustments proposed by above methods were study-specific and not directly applicable to meta-analysis.

2.2.2 Bayesian approaches

Corrections of misclassification can also be implemented via the traditional Bayesian approach by assuming prior information on parameters. Joseph [46] estimated all parameters by means of the beta prior on diagnostic parameters and uniform prior on the disease prevalence. Gustafson [63] extended this approach to investigate the exposure-disease association (odds ratios) with rough but not exact information on misclassification probabilities. The rough information was constructed by the validation study, in which patients went through all the tests. With respect to algorithms, the Gibbs sampler was the conventional device to obtain Bayesian posterior marginal distributions of parameters. In the context of meta-analysis, the application of Bayesian hierarchical models is well established [5, 11, 35].

Although Bayesian approaches produced stronger conclusions on parameters than

frequentist approaches, the conclusions were derived at the cost of stronger assumptions, i.e., the prior information. Choices of prior distributions have been discussed by several authors. Proposed methods included direct matching of percentiles, means and standard deviations to a distribution, [70, 93]. Matching functions of 95% probability ranges of sensitivity and specificity to parameters in the distribution family, the prior distribution can be constructed [46]. For example, the center and a quarter of the range can be matched with the mean and standard deviation of the beta distribution, respectively [46]. Parameters of the beta distribution can then be solved as functions of the range of sensitivity and specificity. Furthermore, vigorous assessments of convergence of the posterior distribution are still under discussion. It is often difficult to decide when it is safe to terminate the sampler and conclude convergence. An example by Cowles [56] showed that the convergence diagnostics did not always agree with one another and not any one was superior over the other. One common conclusion from above methods was that more iterations were required in the presence of high correlations among the parameters [56].

In this project, tables of d-dimer and ultrasound were misclassified. Corrections on estimates were implemented via the maximum likelihood estimation by solving the likelihood equations with constraints. The random effects model allowed the odds ratio between d-dimer and venography to vary from study to study. The Gaussian Hermite integration was applied as a frequentist approach to estimate parameters in the random effects model. In the context of meta-analysis of diagnostic test, this project was a novel application of the Gaussian Hermite integration to the meta-analysis of diagnostic test. It was a simple approach to analyze random effects models with high accuracy. The Gibbs sampling for the generalized linear mixed model, as

a Bayesian approach, was employed to obtain posterior samples of parameters. The Gibbs sampling approach became popular in analyzing complex statistical models in the past decade. It was often applied in meta-analysis to derive the summary measure of test accuracy [10, 99]. Both the Gaussian Hermite integration and Gibbs sampling were applied in this project. They were theoretically and computationally straightforward, compared to other approaches, such as the EM algorithm.

Chapter 3

Methods

3.1 Outline of procedures

3.1.1 Description of data

As discussed in previous chapters, the diagnostic results from d-dimer studies can be summarized into two types of tables: d-dimer versus ultrasonography and d-dimer versus venography. In the context of a contingency table with three factors, the diagnostic data from d-dimer studies were not the complete three-dimensional contingency tables. None of the studies had patients go through all three tests due to various limitations, such as the availability of the test, invasiveness of tests, health status of the patient. Some studies used venography (V) as the reference for d-dimer (D) and others used ultrasonography (U). In the context of contingency table, data from these studies were regarded as marginal tables. As discussed in Chapter 2, information on diagnostic characteristics of ultrasonography was required in order to combine these tables correctly. This information can be presented in two forms: 1. known values of sensitivity and specificity of ultrasonography; 2. observed tables between ultrasonography and venography. In the latter case, the data for analysis were three types of tables: d-dimer versus ultrasonography, d-dimer versus venography, and ultrasonography versus venography. In the meta-analysis of d-dimer, the fixed effects model and the random effects model were considered. In particular, two forms of data were considered in the fixed effects model.

1. Two types of marginal tables were available: D-V, D-U, with known values of $SENS_u$ and $SPEC_u$, derived from external sources.

2. Three types of marginal tables were available: D-V, D-U, and U-V.

In the random effects model, only the second form of data was considered in the analysis. In other words, the tables between ultrasonography and venography were collected for analysis, which was a common case in clinical practice.

3.1.2 The log-linear model and assumptions

Ultrasonography as the gold standard

In the diagnostic test of DVT, ultrasonography was often applied as the reference test for d-dimer in many studies. If ultrasonography was treated the same as the gold standard in the meta-analysis, the DU and DV tables were regarded as diagnostic results from the same reference. In this situation, the meta-analysis was to combine several 2×2 tables and the log-linear model took the following form.

$$\log(m_{dv}) = \beta_0 + \beta_1 \mathrm{D} + \beta_2 \mathrm{V} + \beta_3 \mathrm{DV}$$

In this model, m_{dv} represented cell counts in the 2×2 table with i=0,1 and j=0,1, which were two levels of each test. Using the cell probabilities and the observed tables between the two tests, the likelihood can be constructed and maximization can be performed to obtain estimates of the cell probabilities. In this case, the information between ultrasonography and venography was not used in the analysis because ultrasonography was assumed the same as venography.

Ultrasonography as the silver standard

If ultrasonography was treated as a reference different from venography, the meta-analysis of d-dimer involved three tests, d-dimer, ultrasonography, and venography. The log-linear model for 2×2×2 contingency table was expressed as the following.

$$\log(m_{duv}) = \beta_0 + \beta_1 \mathrm{D} + \beta_2 \mathrm{U} + \beta_3 \mathrm{V} + \beta_4 \mathrm{DV} + \beta_5 \mathrm{UV} + \beta_6 \mathrm{DU} + \beta_7 \mathrm{DUV}$$

In this model, m_{duv} represented cell counts in the three-dimensional table with d,u,v representing the levels of d-dimer, ultrasonography, and venography, respectively. For instance, m_{000} was the number of negative results in all three tests and m_{111} was the number of all positive results.

In this model, DV, UV, and DU were the two-way interactions and DUV was the three-way interaction of all three tests. Note that β_0 was not an independent parameter because the cell probabilities had to sum to 1 in the multinomial distribution. It was a function of the rest parameters in the model, which is derived in the next section. Note that the estimation of coefficients in the above model depended on the available data. For example, in order to estimate the three-way interaction, the complete three-dimensional table was required. In the d-dimer studies, however, such a table with complete dimension was not observed. Therefore, not all the parameters in the log-linear model were estimable. Besides, assumptions were set up for different clinical settings on the three tests. A general discussion on different assumptions was presented by several authors [90, 106], which is summarized below.

- Complete independence: $\beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$
- Conditional independence between d-dimer and ultrasonography given the value of venography: $\beta_6 = \beta_7 = 0$

- Association between any two tests was not affected by the third test: $\beta_7 = 0$

In diagnostic tests, the most widely acceptable assumption was the independence between competing tests conditional on the gold standard, i.e., the second assumption above. In the diagnosis of DVT, venography was regarded as the gold standard. The conditional independence assumption indicated that if the true diagnosis from venography was known, the result from d-dimer diagnosis was not affected by the result from ultrasonography, and vice versa. In the epidemiologic context, this assumption implied that the odds ratio of d-dimer and ultrasonography was 1 in either the diseased patients or the healthy patients. By making this assumption, two components in the log-linear model were set to zero. These two components were the interaction between d-dimer and ultrasound and the three-way interaction, i.e., β_6=β_7=0. In other words, under the conditional independence assumption between d-dimer and ultrasonography, the log-linear model can be written as the following.

$$\log(m_{duv}) = \beta_0 + \beta_1 \mathrm{D} + \beta_2 \mathrm{U} + \beta_3 \mathrm{V} + \beta_4 \mathrm{DV} + \beta_5 \mathrm{UV}$$

The conditional independence was assumed throughout the analysis in this project. In the matrix format, the above model can be re-written as follows.

$$\log(m_{duv}) = \mathrm{X}\beta$$

In this expression, X was the design matrix for the fixed effects (the diagnostic tests) and was expressed as the following.

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Based on this matrix, the order of the cell counts were (m_{000} m_{100} m_{010} m_{110} m_{001} m_{101} m_{011} m_{111}). By subtracting the log of the table total from the above model, the cell counts can be translated to corresponding cell probabilities.

In addition, venography was considered an error-free reference. This indicated that the probability of test positive in venography corresponded to the prevalence of deep vein thrombosis. In the log-linear model, the test positive in venography was represented by the coefficient of venography. The interaction between d-dimer and venography corresponded to the log of odds ratio between these two factors. Similarly, the UV interaction corresponded to the log of odds ratio between ultrasonography and venography. Elaborations on these indications would be presented in section 3.2.

3.1.3 Random effects model

The fixed effects log-linear model treated the model coefficients as fixed but unknown constants across studies. It assumed that all the test results were from the same diagnostic environment. In other words, studies differed only by random errors. However,

these assumptions were not necessarily valid in most clinical environments. Studies were typically conducted at different locations or under different conditions. Each set of data was collected independently under different environments. In the context of diagnostic tests, studies differed in many clinical and epidemiological conditions, such as the disease prevalence, laboratory procedures, patient characteristics, availability of tests, positivity criteria, and so forth. For example, tests of d-dimer as a diagnostic tool often differed in the difference in assays, the availability of central laboratory, and point-of-care of testing [68]. Diagnostic accuracy of d-dimer often varied across studies. The sources of variation in studies of diagnostic accuracy were investigated and summarized in the literature [67].

Among the several heterogeneities, disease prevalence was one of the major variations across studies [67]. As discussed in the previous section, the prevalence of disease was represented by the probability of test positive in venography. This was considered the first random effect in the log-linear mixed model. The following example gave a brief illustration. Two studies had noticeable difference in the prevalence of the disease, 0.1 and 0.8, respectively. Tables from the two studies were shown in Table 3.1 and Table 3.2. In Study 1, the sensitivity of the test was $7/10 = 0.7$

Tests	T+	T-	Total
Disease	7	3	10
no disease	18	72	90

Table 3.1: Disease versus diagnostic test from study 1

and specificity was $72/90 = 0.8$. In Study 2, however, the sensitivity of the test was $24/80 = 0.3$ and specificity $10/20 = 0.5$. Although both studies had the same table

Tests	T+	T-	Total
Disease	24	56	80
no disease	10	10	20

Table 3.2: Disease versus diagnostic test from study 2

total of 100, the numbers of diseased patients in the two tables are quite different due to extreme values of the prevalence. If the difference in prevalence was not taken into account in the analysis, one would collapse the two tables into Table 3.3. The

Tests	T+	T-	Total
Disease	31	59	90
no disease	28	82	110

Table 3.3: Disease versus diagnostic test combining study 1 and study 2

resulting sensitivity of the test was $31/90 = 0.3$ and specificity was 82/110=0.74. Apparently, the pooled sensitivity was attenuated by data from the high prevalence study, whereas the specificity was highly weighted by the low prevalence study. With the same table total, high prevalence of disease resulted in large number of diseased patients. Therefore, the number of disease and positive test patients was larger than that in the low prevalence tables, when calculating the pooled sensitivity. Similarly, when calculating the pooled specificity, weights given to the tables with low disease prevalence were higher than those given to the tables with high disease prevalence because the number of healthy patients was larger in the study with low disease prevalence than in the study with high disease prevalence. In fact, this can be regarded as the interaction between the study and the disease prevalence, i.e., effect

modification by study. The importance of accounting for different prevalence of disease among studies was revealed.

As a relatively new diagnostic test, the clinical performance of d-dimer had not been well established. Studies may result in different estimations of the association between d-dimer and venography, which was represented by the odds ratio between the two tests. In the log-linear model, the odds ratio between d-dimer and venography was represented by the DV interaction. Elaboration on this association was presented in the next section. In order to account for the difference in odds ratios across studies, the random interaction between d-dimer and venography was added to the log-linear model. This random effect allowed variations in the test performance of d-dimer across studies. The normal distribution was assumed on the random effects.

3.2 Fixed effects model

3.2.1 Log-linear model ignoring the imperfection of the silver standard

If ultrasonography was regarded the same as venography, tables between d-dimer and ultrasonography can be regarded as d-dimer and venography. The estimation of diagnostic characteristics of d-dimer can be derived by incorporating the two types of tables without adjusting for the different reference tests. The log-linear model in this situation was expressed as the following.

$$\log(m_{dv}) = \beta_0 + \beta_1 \mathrm{D} + \beta_2 \mathrm{V} + \beta_3 \mathrm{DV}$$

Specifically, the log of four cell probabilities can be expressed as the following.

$\log m_{11} = \beta_0 + \beta_1 + \beta_2 + \beta_3$

$\log m_{01} = \beta_0 + \beta_2$

$\log m_{10} = \beta_0 + \beta_1$

$\log m_{00} = \beta_0$

β_0 was a function of β_1, β_2, and β_3 because the four cell counts had to sum to the table total. By simple algebra, the corresponding four cell probabilities were expressed below.

$p_{11} = \frac{e^{\beta_1+\beta_2+\beta_3}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2+\beta_3}}$

$p_{01} = \frac{e^{\beta_2}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2+\beta_3}}$

$p_{10} = \frac{e^{\beta_1}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2+\beta_3}}$

$p_{00} = \frac{1}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2+\beta_3}}$

Using these expressions and the observed tables of d-dimer and venography, the likelihood function for the i^{th} table was constructed as the following.

$\log L_i = x^i_{11} \log p_{11} + x^i_{01} \log p_{01} + x^i_{10} \log p_{10} + x^i_{00} \log p_{00}$

The sum of log likelihood from each table can be maximized with respect to the model coefficients β_1, β_2, and β_3, using conventional algorithms. Using the functional relationship between the cell probabilities and model coefficients, the maximum likelihood estimates of cell probabilities can be derived. Applying the delta method, the variance covariance matrix of the cell probabilities can be derived.

3.2.2 Log-linear model adjusting for the imperfection of the silver standard

By assuming conditional independence between d-dimer and ultrasonography given the status of venography, the log-linear model was expressed in the following format.

$$\log(\mathbf{m}) = \beta_0 + \beta_1 \mathrm{D} + \beta_2 \mathrm{U} + \beta_3 \mathrm{V} + \beta_4 \mathrm{DV} + \beta_5 \mathrm{UV}$$

In this model, $\mathbf{m}$ was the vector of cell counts m_{duv} in the 2×2×2 contingency table. The indices of m_{duv} represented the levels of each diagnostic test: d=0, 1 for d-dimer negative and positive; u=0, 1 for ultrasonography negative and positive; v=0, 1 for venography negative and positive. Let p_{duv} denote the cell probabilities in the 2×2×2 contingency table. Expressions of these probabilities in terms of model coefficients, β, were listed below.

$$\log(p_{111}) = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5$$

$$\log(p_{011}) = \beta_0 + \beta_2 + \beta_3 + \beta_5$$

$$\log(p_{101}) = \beta_0 + \beta_1 + \beta_3 + \beta_4$$

$$\log(p_{001}) = \beta_0 + \beta_3$$

$$\log(p_{110}) = \beta_0 + \beta_1 + \beta_2$$

$$\log(p_{010}) = \beta_0 + \beta_2$$

$$\log(p_{100}) = \beta_0 + \beta_1$$

$$\log(p_{000}) = \beta_0$$

Recall that β_0 was not an independent parameter. It was a function of the rest βs because all the cell probabilities had to sum to 1 in the multinomial distribution. The following expressions provided an elaboration on this issue.

$p_{duv} = e^{\beta_0} \times e^{\beta_1 D + \beta_2 U + \beta_3 V + \beta_4 DV + \beta_5 UV}$

$\sum_{d=0}^{1} \sum_{u=0}^{1} \sum_{v=0}^{1} p_{duv} = e^{\beta_0} \times \sum_{duv} e^{\beta_1 D + \beta_2 U + \beta_3 V + \beta_4 DV + \beta_5 UV} = 1$

Hence, β_0 = -log(1 + e^{β_1} + e^{β_2} + $e^{\beta_1+\beta_2}$ + e^{β_3} + $e^{\beta_1+\beta_3+\beta_4}$ + $e^{\beta_2+\beta_3+\beta_5}$ + $e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}$). The cell probabilities can be re-written as the following.

$$p_{111} = \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{011} = \frac{e^{\beta_2+\beta_3+\beta_5}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{101} = \frac{e^{\beta_1+\beta_3+\beta_4}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{001} = \frac{e^{\beta_3}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{110} = \frac{e^{\beta_1+\beta_2}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{010} = \frac{e^{\beta_2}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{100} = \frac{e^{\beta_1}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{000} = \frac{1}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

From above expressions, the model coefficients for interactions had clinical meanings. For example, $\beta_4 = \log(p_{111})$ - $\log(p_{011})$ - ($\log(p_{110})$ - $\log(p_{010})$) = $\log\frac{p_{111}\times p_{010}}{p_{011}\times p_{110}}$. The last expression represented the log of odds ratio between d-dimer and venography when ultrasonography=1, i.e., in patients with positive test results on ultrasonography. Similarly, $\beta_5 = \log(p_{111})$ - $\log(p_{101})$ - ($\log(p_{110})$ - $\log(p_{100})$) = $\log\frac{p_{111}\times p_{100}}{p_{101}\times p_{110}}$, which represented the log of odds ratio between ultrasonography and venography in patients with positive test results on d-dimer.

If we let $\mathbf{p} = (p_{000}\ p_{100}\ p_{010}\ p_{110}\ p_{001}\ p_{101}\ p_{011}\ p_{111})$, then the likelihood of the marginal table can be expressed in terms of $\mathbf{p}$ and the observed cell counts. For instance, the log likelihood of a DU table was written as the following.

$$\log(L_{du}) = \sum_{du} x_{du.} \log(p_{du.})$$

In this expression, $x_{du.}$ was the vector of cell counts from the table between d-dimer and ultrasonography. The order of cell counts in this table was $x_{00.}\ x_{10.}\ x_{01.}\ x_{11.}$, which was consistent with the order of $p_{du.}$. The $p_{du.}$ represented the summation of probabilities over the levels of venography for any d^{th} level of d-dimer and u^{th} level of ultrasonography, i.e., $p_{du.} = p_{du1} + p_{du0}$. Similarly, $p_{d.v} = p_{d1v} + p_{d0v}$ and $p_{.uv} = p_{1uv} + p_{0uv}$.

Following a similar procedure, the log likelihood of a DV table and a UV table can be derived. In the fixed effects model, the joint likelihood function was the product of the likelihood from each table because the tables were collected independently from different studies. Based on the joint likelihood, maximum likelihood estimates (MLE) of βs can be obtained. MLE of sensitivity and specificity of d-dimer can be derived accordingly.

3.2.3 Algorithm of analysis with known sensitivity and specificity of the silver standard

Description of the problem

When two types of tables were available, i.e., DU and DV tables, the diagnostic information from ultrasonography was required in order to distinguish ultrasonography from venography. In this case, the sensitivity and specificity of ultrasonography, denoted $SENS_u$ and $SPEC_u$, were assumed known and regarded as two constraints on the likelihood function. Expressions of $SENS_u$ and $SPEC_u$ in terms of cell probabilities were listed below.

$$SENS_u = \frac{p_{111} + p_{011}}{p_{111} + p_{011} + p_{101} + p_{001}}$$

$$SPEC_u = \frac{p_{000} + p_{100}}{p_{000} + p_{100} + p_{010} + p_{110}}$$

Using the expressions of p_{duv} in the previous section, the expressions of $SENS_u$ and $SPEC_u$ can be simplified as the following.

$$SENS_u = \frac{e^{\beta_2+\beta_5}}{1 + e^{\beta_2+\beta_5}}$$

$$SPEC_u = \frac{1}{1 + e^{\beta_2}}$$

By simple algebra, β_2 and β_5 can be written as functions of $SENS_u$ and $SPEC_u$ as $\beta_2 = log(\frac{1-SPEC_u}{SPEC_u})$ and $\beta_5 = log(\frac{SENS_u}{1-SENS_u}) - log(\frac{1-SPEC_u}{SPEC_u})$. With known values of $SENS_u$ and $SPEC_u$, the values of β_2 and β_5 were determined. In other words, knowing the values of $SENS_u$ and $SPEC_u$ was the same as knowing the values of β_2 and β_5 provided that $SENS_u$ and $SPEC_u$ did not attain the boundaries of 0 or 1. The parameters for estimation became β_1, β_3, and β_4. The vector of coefficients can be written as the following.

$$\beta = \begin{pmatrix} \beta_1 \\ log(\frac{1-SPEC_u}{SPEC_u}) \\ \beta_3 \\ \beta_4 \\ log(\frac{SENS_u}{1-SENS_u}) - log(\frac{1-SPEC_u}{SPEC_u}) \end{pmatrix}$$

By incorporating this vector into the likelihood function, maximization procedure can be implemented via conventional approaches. Each table used the same vector of β to construct the likelihood. The joint likelihood was the product of the likelihood from each table. Maximization was then performed on the joint likelihood to obtain estimates of β_1, β_3 and β_4.

Similar to the procedure described in the previous section, the log likelihood for each table took one of the following forms.

$$\log(L_{du}) = \sum_{du} x_{du.} \log(p_{du.})$$
$$\log(L_{uv}) = \sum_{uv} x_{.uv} \log(p_{.uv})$$
$$\log(L_{du}) = \sum_{dv} x_{d.v} \log(p_{d.v})$$

The log of joint likelihood function was the sum of the log likelihood from each table. By means of the Newton-Raphson algorithm, the log(L) can be maximized via an iterative process. At the end of the algorithm, the estimates and corresponding hessian matrix can be derived with respect to the model coefficients.

Applying the Newton-Raphson algorithm

In brief, the Newton-Raphson algorithm can be described as follows: To find a root of f(θ)=0 given an initial value θ_0, using the iteration $\theta_{k+1} = \theta_k - \frac{f(\theta_k)}{f'(\theta_k)}$ for k = 0,1,2,...,m until convergence. Convergence was assessed by $|\theta_{k+1} - \theta_k| < \xi$, which was an arbitrary small positive number. Applying the multivariate version of this theorem to solve $f(\beta) = \mathbf{0}$ produced the maximum likelihood estimates (MLE) of coefficients in the log-linear model. Note that the convergence criterion was changed to $max(|\theta_{k+1} - \theta_k|) < \xi$ because θ_k was a vector in the multivariate case. This criterion ensured the convergence of all parameters because the maximum of differences was bounded by ξ. The parameters of interest in this problem were the sensitivity and specificity of d-dimer using venography as the reference. After the MLE of coefficients β_1, β_3 and β_4 in the log-linear model were estimated, MLE of cell probabilities in the complete three-dimensional table can be derived. Because the sensitivity and specificity of d-dimer were functions of cell probabilities, the MLE of sensitivity and specificity of d-dimer would be derived by the MLE of cell probabilities in the following expressions.

$$\hat{SENS}_d = \frac{\hat{p}_{111} + \hat{p}_{101}}{\hat{p}_{111} + \hat{p}_{101} + \hat{p}_{011} + \hat{p}_{001}}$$

$$\hat{SPEC}_d = \frac{\hat{p}_{000} + \hat{p}_{010}}{\hat{p}_{000} + \hat{p}_{010} + \hat{p}_{100} + \hat{p}_{110}}$$

If $\hat{p}_{duv}$ were represented by model coefficients $\hat{\beta}$, the expressions of $\hat{SENS}_d$ and $\hat{SPEC}_d$ can be written as:

$$\hat{SENS}_d = \frac{e^{\hat{\beta}_1+\hat{\beta}_4}}{1+e^{\hat{\beta}_1+\hat{\beta}_4}}$$

$$\hat{SPEC}_d = \frac{1}{1+e^{\hat{\beta}_1}}$$

Substituting the MLE of β_1 and β_4 in above expressions produced the MLE of sensitivity and specificity of d-dimer, namely, $SENS_d$ and $SPEC_d$.

In the estimation of β_1, β_3 and β_4 via Newton-Raphson algorithm, the score vector and hessian matrix were required. The score vector was the vector of first derivatives of the joint likelihood function with respect to each parameter. The hessian matrix was the matrix of second derivatives of the joint likelihood with respect to each pair of parameters. In order to obtain derivatives of the joint likelihood of each parameter, the derivatives of cell probabilities with respect to each parameter should be derived, which was listed below.

Derivatives of p_{duv} with respect to β

Let $\sum e^{X\beta}$ denote the denominator of p_{duv}, which was $(1+e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2} + e^{\beta_3} + e^{\beta_1+\beta_3+\beta_4} + e^{\beta_2+\beta_3+\beta_5} + e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5})$. Derivatives of $\sum e^{X\beta}$ with respect to β_1, β_3 and β_4 were calculated as below.

$$\frac{\partial \sum e^{X\beta}}{\partial \beta_1} = e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5} + e^{\beta_1+\beta_3+\beta_4} + e^{\beta_1+\beta_2} + e^{\beta_1}$$

$$\frac{\partial \sum e^{X\beta}}{\partial \beta_3} = e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5} + e^{\beta_2+\beta_3+\beta_5} + e^{\beta_1+\beta_3+\beta_4} + e^{\beta_3}$$

$$\frac{\partial \sum e^{X\beta}}{\partial \beta_4} = e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5} + e^{\beta_1+\beta_3+\beta_4}$$

Using these expressions, the first derivative of p_{duv} with respect to each of β_1, β_3 and β_4 were calculated as follows.

$$\frac{\partial p_{111}}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}{\sum e^{X\beta}}\right) = \frac{\sum e^{X\beta} e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5} - e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}\frac{\partial \sum e^{X\beta}}{\partial \beta_1}}{(\sum e^{X\beta})^2}$$

$$= \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}{\sum e^{X\beta}} - \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}{\sum e^{X\beta}} \times \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5} + e^{\beta_1+\beta_3+\beta_4} + e^{\beta_1+\beta_2} + e^{\beta_1}}{\sum e^{X\beta}}$$

$$= p_{111} - p_{111}(p_{111} + p_{101} + p_{110} + p_{100})$$

$$\frac{\partial p_{011}}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{e^{\beta_2+\beta_3+\beta_5}}{\sum e^{X\beta}}\right) = -\frac{e^{\beta_2+\beta_3+\beta_5}}{(\sum e^{X\beta})^2} \times \frac{\partial \sum e^{X\beta}}{\partial \beta_1}$$

$$= 0 - p_{011}(p_{111} + p_{101} + p_{110} + p_{100})$$

$$\frac{\partial p_{101}}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{e^{\beta_1+\beta_3+\beta_4}}{\sum e^{X\beta}}\right) = \frac{\sum e^{X\beta} e^{\beta_1+\beta_3+\beta_4} - e^{\beta_1+\beta_3+\beta_4}\frac{\partial \sum e^{X\beta}}{\partial \beta_1}}{(\sum e^{X\beta})^2}$$

$$= p_{101} - p_{101}(p_{111} + p_{101} + p_{110} + p_{100})$$

$$\frac{\partial p_{001}}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{e^{\beta_3}}{\sum e^{X\beta}}\right) = -\frac{e^{\beta_3}}{(\sum e^{X\beta})^2}\frac{\partial \sum e^{X\beta}}{\partial \beta_1}$$

$$= 0 - p_{001}(p_{111} + p_{101} + p_{110} + p_{100})$$

$$\frac{\partial p_{110}}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{e^{\beta_1+\beta_2}}{\sum e^{X\beta}}\right) = \frac{\sum e^{X\beta} e^{\beta_1+\beta_2} - e^{\beta_1+\beta_2}\frac{\partial \sum e^{X\beta}}{\partial \beta_1}}{(\sum e^{X\beta})^2}$$

$$= p_{110} - p_{110}(p_{111} + p_{101} + p_{110} + p_{100})$$

$$\frac{\partial p_{010}}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{e^{\beta_2}}{\sum e^{X\beta}}\right) = -\frac{e^{\beta_2}}{(\sum e^{X\beta})^2}\frac{\partial \sum e^{X\beta}}{\partial \beta_1}$$

$$= 0 - p_{010}(p_{111} + p_{101} + p_{110} + p_{100})$$

$$\frac{\partial p_{100}}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{e^{\beta_1}}{\sum e^{X\beta}}\right) = \frac{\sum e^{X\beta} e^{\beta_1} - e^{\beta_1} \frac{\partial \sum e^{X\beta}}{\partial \beta_1}}{(\sum e^{X\beta})^2}$$

$$= p_{100} - p_{100}(p_{111} + p_{101} + p_{110} + p_{100})$$

$$\frac{\partial p_{100}}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{e^{\beta_1}}{\sum e^{X\beta}}\right) = - \frac{1}{(\sum e^{X\beta})^2} \frac{\partial \sum e^{X\beta}}{\partial \beta_1}$$

$$= 0 - p_{000}(p_{111} + p_{101} + p_{110} + p_{100})$$

In the matrix form, derivatives of p_{duv} with respect to β_1 can be summarized as below.

$$\frac{\partial p_{duv}}{\partial \beta_1} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} - \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} \times (p_{111} + p_{101} + p_{110} + p_{100})$$

Note that $\sum \frac{\partial p_{duv}}{\partial \beta_1} = p_{1..} - p_{1..} \times \sum p_{duv} = 0$, where $p_{1..}$ denoted the summation of cell probabilities over levels of each of ultrasonography and venography given a positive result in d-dimer.

Following similar procedures, the first derivatives of p_{duv} with respect to β_3 and β_4 can be obtained as the following.

$$\frac{\partial p_{duuv}}{\partial \beta_3} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} - \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} \times (p_{111} + p_{101} + p_{011} + p_{001})$$

$$\frac{\partial p_{duuv}}{\partial \beta_4} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} - \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} \times (p_{111} + p_{101})$$

Again, $\sum \frac{\partial p_{duuv}}{\partial \beta_3} = \sum \frac{\partial p_{duuv}}{\partial \beta_4} = 0$.

Derivatives of log likelihood with respect to β

The log of joint likelihood was the summation of the log likelihood from each marginal table. Let $x_{du.}$ and $x_{d.v}$ represent the observed cell counts in DU and DV tables, respectively. The log of joint likelihood of 1 DU table and 1 DV table was expressed below.

$$\text{logL} = \log L_{du} + \log L_{dv} = x_{11.} \log (p_{111} + p_{110}) + x_{01.} \log (p_{011} + p_{010}) + x_{10.} \log (p_{101} + p_{100}) + x_{00.} \log (p_{001} + p_{000}) + x_{1.1} \log (p_{111} + p_{101}) + x_{0.1} \log (p_{011} + p_{001}) + x_{1.0} \log (p_{110} + p_{100}) + x_{0.0} \log (p_{010} + p_{000})$$

The p_{duv} was expressed in terms of β after taking into account the constraints that all probabilities sum to 1 and that β_2 and β_5 were determined by $SENS_u$ and $SPEC_u$. The first-order partial derivative of the log likelihood with respect to β_1 was obtained by the following calculation.

$$\frac{\partial \log L}{\partial \beta_1} = \frac{x_{11.}}{p_{111}+p_{110}}\left(\frac{\partial p_{111}}{\partial \beta_1} + \frac{\partial p_{110}}{\partial \beta_1}\right) + \frac{x_{01.}}{p_{011}+p_{010}}\left(\frac{\partial p_{011}}{\partial \beta_1} + \frac{\partial p_{010}}{\partial \beta_1}\right) + \frac{x_{10.}}{p_{101}+p_{100}}\left(\frac{\partial p_{101}}{\partial \beta_1} + \frac{\partial p_{100}}{\partial \beta_1}\right) + \frac{x_{00.}}{p_{001}+p_{000}}\left(\frac{\partial p_{001}}{\partial \beta_1} + \frac{\partial p_{000}}{\partial \beta_1}\right) + \frac{x_{1.1}}{p_{111}+p_{101}}\left(\frac{\partial p_{111}}{\partial \beta_1} + \frac{\partial p_{101}}{\partial \beta_1}\right) + \frac{x_{0.1}}{p_{011}+p_{001}}\left(\frac{\partial p_{011}}{\partial \beta_1} + \frac{\partial p_{001}}{\partial \beta_1}\right) + \frac{x_{1.0}}{p_{110}+p_{100}}\left(\frac{\partial p_{110}}{\partial \beta_1} + \frac{\partial p_{100}}{\partial \beta_1}\right) + \frac{x_{0.0}}{p_{010}+p_{000}}\left(\frac{\partial p_{010}}{\partial \beta_1} + \frac{\partial p_{000}}{\partial \beta_1}\right)$$

Substituting expressions of p_{duv} in terms of β into the derivatives above gave rise to the first derivative of log of joint likelihood with respect to β_1. By some algebraic work, the first derivative of log likelihood with respect to β_1 was simplified as the following expression.

$$\frac{\partial \log L}{\partial \beta_1} = x_{11.} + x_{10.} - p_{1..} \sum x_{ij.} + x_{1.1} + x_{1.0} - p_{1..} \sum x_{i.k}$$

Similarly, first derivatives of the log likelihood with respect to β_3 and β_4 were expressed as follows.

$$\frac{\partial \log L}{\partial \beta_3} = \frac{x_{11.}p_{111}}{p_{111}+p_{110}} + \frac{x_{01.}p_{011}}{p_{011}+p_{010}} + \frac{x_{10.}p_{101}}{p_{101}+p_{100}} + \frac{x_{00.}p_{001}}{p_{001}+p_{000}} - p_{..1} \sum x_{ij.} + x_{1.1} + x_{0.1} - p_{..1} \sum x_{i.k}$$

$$\frac{\partial \log L}{\partial \beta_4} = \frac{x_{11.}p_{111}}{p_{111}+p_{110}} + \frac{x_{10.}p_{101}}{p_{101}+p_{100}} - p_{1.1} \sum x_{ij.} + x_{1.1} - p_{1.1} \sum x_{i.k}$$

The score vector was constructed using the expressions above.

$$\text{score} = \left(\frac{\partial \log L}{\partial \beta_1} \quad \frac{\partial \log L}{\partial \beta_3} \quad \frac{\partial \log L}{\partial \beta_4} \right)$$

The second order derivatives of the log likelihood with respect to β were derived from above expressions.

$$\frac{\partial^2 \log L}{\partial \beta_1^2} = - \sum x_{ij.} p_{1..}(1 - p_{1..}) - \sum x_{i.k} p_{1..}(1 - p_{1..})$$

$$\frac{\partial^2 \log L}{\partial \beta_3^2} = x_{11.} \frac{p_{111}}{p_{11.}}(1 - \frac{p_{111}}{p_{11.}}) + x_{01.} \frac{p_{011}}{p_{01.}}(1 - \frac{p_{011}}{p_{01.}}) + x_{10.} \frac{p_{101}}{p_{10.}}(1 - \frac{p_{101}}{p_{10.}}) + x_{00.} \frac{p_{001}}{p_{00.}}(1 - \frac{p_{001}}{p_{00.}}) - \sum x_{ij.} p_{..1}(1 - p_{..1}) - \sum x_{i.k} p_{..1}(1 - p_{..1})$$

$$\frac{\partial^2 \log L}{\partial \beta_4^2} = x_{11.} \frac{p_{111}}{p_{11.}}(1 - \frac{p_{111}}{p_{11.}}) + x_{10.} \frac{p_{101}}{p_{10.}}(1 - \frac{p_{101}}{p_{10.}}) - \sum x_{ij.} p_{1.1}(1 - p_{1.1}) - \sum x_{i.k} p_{1.1}(1 - p_{1.1})$$

$$\frac{\partial^2 \log L}{\partial \beta_1 \beta_3} = - \sum x_{ij.} p_{1.1} + \sum x_{ij.} p_{..1} p_{1..} - \sum x_{i.k} p_{1.1} + \sum x_{i.k} p_{..1} p_{1..}$$

$$\frac{\partial^2 \log L}{\partial \beta_1 \beta_4} = - \sum x_{ij.} p_{1.1}(1 - p_{1..}) - \sum x_{i.k} p_{1.1}(1 - p_{1..})$$

$$\frac{\partial^2 \log L}{\partial \beta_3 \beta_4} = \frac{\partial^2 \log L}{\partial \beta_4 \beta_3} = x_{11.} \frac{p_{111}}{p_{11.}}(1 - \frac{p_{111}}{p_{11.}}) + x_{10.} \frac{p_{101}}{p_{10.}}(1 - \frac{p_{101}}{p_{10.}}) - \sum x_{ij.} p_{1.1}(1 - p_{..1}) - \sum x_{i.k} p_{1.1}(1 - p_{..1})$$

The hessian matrix was the matrix of second derivatives of the log likelihood and symmetric about the diagonal. The diagonal and below-diagonal elements of the hessian matrix can be constructed using above expressions in the following format.

$$\text{hessian} = \begin{pmatrix} \frac{\partial^2 \log L}{\partial \beta_1^2} & & \\ \frac{\partial^2 \log L}{\partial \beta_1 \beta_3} & \frac{\partial^2 \log L}{\partial \beta_3^2} & \\ \frac{\partial^2 \log L}{\partial \beta_1 \beta_4} & \frac{\partial^2 \log L}{\partial \beta_3 \beta_4} & \frac{\partial^2 \log L}{\partial \beta_4^2} \end{pmatrix}$$

The Newton-Raphson procedure incorporated the score vector and the hessian matrix iteratively to acquire the maximum likelihood estimates (MLE) of parameters. At convergence of the algorithm, the variance covariance matrix of β can be

approximated by the inverse of the observed information matrix evaluated at the MLE of β. The observed information matrix was calculated as the negative of hessian matrix. The observed information matrix was used as an asymptotic equivalence of the Fisher's information matrix. Let $\hat{V}_{\hat{\beta}}$ denote the estimated variance covariance matrix of $\hat{\beta}$. The relationship between hessian matrix and the estimated variance covariance matrix of β was expressed as follows.

$$\hat{V}_{\hat{\beta}} = (-hessian_{\hat{\beta}})^{-1}$$

The nlm() function in R used the score vector and hessian matrix calculated above to derive MLE of parameters in the log likelihood function.

Note that the parameters of interest were the sensitivity and specificity of d-dimer, which were functions of model coefficients, β. Because the variance covariance matrix of β can be derived from the hessian matrix evaluated at the MLE of β. The variance covariance matrix of sensitivity and specificity of d-dimer can be calculated via the multivariate delta method.

In order to apply the delta method, first derivatives of sensitivity and specificity with respect to β are required. The functional forms of sensitivity and specificity with respect to β are displayed below.

$$SENS_d = \frac{p_{111}+p_{101}}{p_{111}+p_{101}+p_{011}+p_{001}} = \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_3}} = \frac{e^{\beta_1+\beta_4}}{1+e^{\beta_1+\beta_4}}$$

$$SPEC_d = \frac{p_{000}+p_{010}}{p_{000}+p_{010}+p_{100}+p_{110}} = \frac{1+e^{\beta_2}}{1+e^{\beta_2}+e^{\beta_1}+e^{\beta_1+\beta_2}} = \frac{1}{1+e^{\beta_1}}$$

The first order derivative matrix of $SENS_d$ and $SPEC_d$ was constructed as the following:

$$D_{sc\beta} = \begin{pmatrix} \frac{\partial SENS_d}{\partial \beta_1} & \frac{\partial SENS_d}{\partial \beta_3} & \frac{\partial SENS_d}{\partial \beta_4} \\ \frac{\partial SPEC_d}{\partial \beta_1} & \frac{\partial SPEC_d}{\partial \beta_3} & \frac{\partial SPEC_d}{\partial \beta_4} \end{pmatrix} = \begin{pmatrix} \frac{e^{\beta_1+\beta_4}}{(1+e^{\beta_1+\beta_4})^2} & 0 & \frac{e^{\beta_1+\beta_4}}{(1+e^{\beta_1+\beta_4})^2} \\ -\frac{e^{\beta_1}}{(1+e^{\beta_1})^2} & 0 & 0 \end{pmatrix}.$$

Applying this derivative matrix to the delta method with estimated values of β, the variance-covariance matrix for sensitivity and specificity were calculated as $D_{sc\hat{\beta}}\hat{V}_{\hat{\beta}}D'_{sc\hat{\beta}}$.

3.2.4 Algorithm of analysis of cross-tables between the silver standard and the gold standard

Description of the problem

In the literature of diagnostic tests, tables between the silver standard and the gold standard may be available instead of the true sensitivity and specificity of the silver standard. In this situation, the tables for analysis were DU, DV, and UV tables. Given these tables, the statistical solution to find the maximum likelihood estimates (MLE) was straightforward by writing out the joint likelihood from the three types of tables and applying the Newton-Raphson algorithm.

Estimation

As noted in section 3.2.2, the log likelihood of each marginal table can be expressed in the function of observed cell counts and the cell probabilities. The log likelihoods of DU, UV, and DV tables had the following expressions.

$$\log(L_{du}) = \sum_{du} x_{du.} \log(p_{du.})$$
$$\log(L_{uv}) = \sum_{uv} x_{.uv} \log(p_{.uv})$$
$$\log(L_{du}) = \sum_{dv} x_{d.v} \log(p_{d.v})$$

The log joint likelihood was then given by: $\log(\mathrm{L}) = \log(L_{du}) + \log(L_{uv}) + \log(L_{dv})$. In these log likelihoods, $x_{du.}$, $x_{.uv}$ and $x_{d.v}$ were vectors of observed cell counts in the DU, UV, and DV tables, respectively. In order to acquire the maximum likelihood

estimates (MLE) of model coefficients, the nlm() function in R was employed. The nlm() function used the Newton-type algorithm to locate the minimum of a function and produced the asymptotic hessian matrix. It is a convenient device to carry out minimization of an unconstrained function. The negative log likelihood function and the model parameters were specified in the nlm() function to obtain the MLEs of the parameters. At convergence, the estimates, score vector, and negative hessian matrix were produced by the nlm() function. The inverse of the negative hessian matrix provided the estimated variance covariance matrix of the model coefficients. In order to obtain the variance covariance matrix of the sensitivity and specificity of d-dimer, the first order derivatives of sensitivity and sepecificity of d-dimer with respect to β_1 and β_4 were the same as those provided in section 3.2.3. The complete derivative matrix can be constructed below.

$$D_{sc\beta} = \begin{pmatrix} \frac{\partial SENS_d}{\partial \beta_1} & \frac{\partial SENS_d}{\partial \beta_2} & \frac{\partial SENS_d}{\partial \beta_3} & \frac{\partial SENS_d}{\partial \beta_4} & \frac{\partial SENS_d}{\partial \beta_5} \\ \frac{\partial SPEC_d}{\partial \beta_1} & \frac{\partial SENS_d}{\partial \beta_2} & \frac{\partial SPEC_d}{\partial \beta_3} & \frac{\partial SPEC_d}{\partial \beta_4} & \frac{\partial SENS_d}{\partial \beta_5} \end{pmatrix} =$$

$$\begin{pmatrix} \frac{e^{\beta_1+\beta_4}}{(1+e^{\beta_1+\beta_4})^2} & 0 & 0 & \frac{e^{\beta_1+\beta_4}}{(1+e^{\beta_1+\beta_4})^2} & 0 \\ -\frac{e^{\beta_1}}{(1+e^{\beta_1})^2} & 0 & 0 & 0 & 0 \end{pmatrix}$$

Using the matrix of derivatives, the estimated variance covariance matrix of sensitivity and specificity can be obtained by the delta method using the variance covariance matrix of β.

3.3 Random effects model

As discussed in 3.1.3, there was heterogeneity among studies on the diagnostic accuracy of d-dimer. Studies differed not only in the reference standard, but also in

a set of random factors, such as the prevalence of disease, laboratory procedures, patient characteristics, and so forth. By adding random effects in the model, systematic variations from study to study were taken into account. The difference in the disease prevalence and the diagnostic characteristics of d-dimer were considered major contributors to the heterogeneity among studies.

3.3.1 Model accounting for the heterogeneity in disease prevalence

In this section, the model that took into account heterogeneity in disease prevalence among studies was considered. As mentioned in previous sections, the coefficient of venography in the log-linear model represented the prevalence of deep vein thrombosis (DVT) if venography was the perfect reference. Taking into account the heterogeneity in this coefficient, the random effect of venography was added to the model. Unlike conventional random effects models, however, the random intercept is not considered. The reason is that in the log-linear model for multinomial distribution, the intercept β_0 is not an independent parameter. It is, instead, a function of the rest of the β because the cell probabilities should sum to 1. With this unusual intercept, adding a random intercept is not meaningful. The log-linear model with the random coefficient of venography is given below.

$$\log(\mathbf{m}) = \mathrm{X}\beta + \mathrm{Z}\gamma = \beta_0 + \beta_1\mathrm{D} + \beta_2\mathrm{U} + \beta_3\mathrm{V} + \beta_4\mathrm{DV} + \beta_5\mathrm{UV} + \gamma\mathrm{V}$$

As in conventional random effects models, the random effect of disease prevalence in the log-linear model was assumed to have a normal distribution with mean 0 and variance, σ^2. In this case, the random effect from each cluster was assumed to come from the same normal distribution. In other words, the σ^2 was assumed to be the

same across studies and it was one of the parameters to be estimated.

3.3.2 Model accounting for heterogeneities in disease prevalence and the association between the test and the gold standard

The second random effect under consideration was the interaction between d-dimer and venography. In the log-linear model from section 3.2.2, the interaction between d-dimer and venography corresponded to the log odds ratio between d-dimer and venography, i.e., the test performance of d-dimer. This may be a systematic variation among studies due to difference in positivity thresholds. Taking into account this variation, the second random effects log-linear model included both the random venography coefficient and the random interaction between d-dimer and venography. The log-linear model with two random effects can be expressed as the following.

$$\log(\mathbf{m}) = \mathrm{X}\beta + \mathrm{Z}\gamma = \beta_0 + \beta_1\mathrm{D} + \beta_2\mathrm{U} + \beta_3\mathrm{V} + \beta_4\mathrm{DV} + \beta_5\mathrm{UV} + \gamma_1\mathrm{V} + \gamma_2\mathrm{DV}$$

In this model, X took the same form as in the fixed effects model and Z was a subset of the X matrix because the random effect was a subset of fixed effects. This may not necessarily be true in other random effects model. The matrix Z was constructed by extracting columns of V and DV from the X matrix.

$$Z = \begin{pmatrix} V & DV \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Notice that Z employed the 0-1 contrast as in the conventional random effects model. This contrast, however, may have a significant impact on the estimation because the column of random interaction between d-dimer and venography only affected two cells probabilities in the contingency table. The rest of the six cells, all zeros in the Z matrix were not affected by the random DV interaction. Although the random intercept β_0 had an effect on all cells, it may not overcome the impact from the zero values in the Z matrix.

A different design matrix of random effects using the -1 1 contrast, namely Z^*, was considered. This matrix had a greater impact on the cell counts than the original 0-1 Z matrix. In order to derive this new matrix, consider the following -1 1 contrasts for the fixed effects d-dimer(D), ultrasonography(U) and venography(V).

$$X^* = \begin{pmatrix} \beta_0 & D & U & V \\ 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Multiplying the columns of D and V in X^* produced the column of DV in the new Z^* matrix. Combining this new DV column and the V column in X^*, the new Z^* matrix had the following format.

$$\mathrm{Z} = \begin{pmatrix} V & DV \\ -1 & 1 \\ -1 & -1 \\ -1 & 1 \\ -1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$$

The -1 1 contrast attenuated the "0" effect when multiplying the columns of d-dimer and venography. Compared to the original Z matrix, the new Z^* affected all cells in the 2×2×2 contingency table. Both the Z and Z^* matrix were applied in the log-linear model with two random effects.

Distribution of random effects

The distribution of random effects in the conventional linear mixed models was specified as the normal distribution with mean 0 and a variance-covariance matrix. In most cases, the random effects were assumed independent. Similar structures and assumptions on random effects were applied in generalized linear mixed models (GLMM), such as models with logit, log, and probit link functions [23, 29, 76]. In other words, the random effects $\gamma=(\gamma_1\ \gamma_2)$ were jointly normally distributed with mean $\mathbf{0}$ and variance of each random effect forming the diagonal variance matrix Σ. All the analysis of random effects model in this project was based on this assumption. In the matrix format, the log linear model was summarized as follows.

$$\log\begin{pmatrix} m_c^{000} \\ m_c^{100} \\ m_c^{010} \\ m_c^{110} \\ m_c^{001} \\ m_c^{101} \\ m_c^{011} \\ m_c^{111} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ -1 & -1 \\ -1 & 1 \\ -1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \gamma_{1c} \\ \gamma_{2c} \end{bmatrix},$$

$$\begin{pmatrix} \gamma_{1c} \\ \gamma_{2c} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}\right)$$

The m_c^{duv} represented cell counts in the c^{th} table and γ_{1c} and γ_{2c} were the two random effects in the c^{th} table. The random effects were assumed independent within each table and between tables.

3.3.3 Estimation using Gaussian Hermite integration

The joint distribution in the random effects model was calculated as the product of the likelihood and the distribution of random effects. In this project, the likelihood of the data had a multinomial distribution. The random effects followed the bivariate normal distribution with mean 0 and variance matrix Σ. In order to estimate fixed effect coefficients in the random effects model, maximization was performed on the marginal likelihood of fixed effects. According to probability theory, the marginal likelihood is obtained by integrating the joint distribution with respect to the random effects. The normal density of the random effects is, therefore, a component of the joint likelihood. Integration over the normal distribution, however, did not yield a closed form. This has been the difficulty in analyzing random effects models in the literature. Different techniques were applied to approximate the integrals. In this project, the Gaussian Hermite approximation was employed because of its high accuracy and easy implementation.

Gaussian Hermite integration

In the analysis of random effects model, integrating out random effects from the joint distribution function has been a challenge for frequentists. As the number of random effects increased, the difficulties of integration increased. The Gaussian Hermite integration is an advantageous numerical integration approach with high accuracy. It originated from the Gaussian formula for integrations over infinite interval. Given sets of a_k and w_k, the integration of function $f(x)$ can be approximated by summations. This approximation is expressed below.

$$\int_{-\infty}^{+\infty} w(x)f(x)dx \approx \sum_{k=1}^{n} w_k f(a_k)$$

The $w(x)$ is called the weight function. When $w(x) = e^{-x^2}$, the approximation was called the Hermite formula and took the following form.

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)dx \approx \sum_{k=1}^{n} w_k f(a_k)$$

This is the basic idea behind "Gaussian Hermite integration" (GHI). The weights w_k are functions of the abscissas a_k, and n denotes the number of pairs of a_k and w_k. Detailed expressions for these two sets of quantities are given by Davis [69]. Abramowitz and Stegun [51] calculated values of a_k and w_k for different values of n. The accuracy of the integration increases as the number of pairs of abscissas and weights increases, although improvement is achieved at the cost of computational time.

In the model with one random effect, i.e., random disease prevalence only, the joint distribution was the product of the likelihood from the observed table and the normal density of random effects. In order to acquire estimates for the model coefficients, maximization should be performed on the marginal likelihood which resulted from integrating out random effects. The joint distribution was expressed as the following.

$$f(data|\beta,\gamma)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\gamma^2}{2\sigma^2}}$$

The function f(data|β, γ) denoted the likelihood of observed tables. Integration of this function was performed with respect to γ to obtain the marginal likelihood for maximization.

$$\int_{-\infty}^{\infty} f(data|\beta,\gamma)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{\gamma^2}{2\sigma^2}}d\gamma$$

Notice that this integral does not fit the GHI approximation directly because the function of γ is not in the standard form as in the Hermite formula. Transformations are required. Let $x = \frac{\gamma}{\sqrt{2}\sigma}$. The integral is equivalent to the following expression.

$$\int_{-\infty}^{\infty} f(data|\beta, \gamma = \sqrt{2}x\sigma)\frac{1}{\sqrt{2\pi}\sigma}e^{-x^2}\sqrt{2}\sigma dx = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} f(data|\beta, \gamma = \sqrt{2}x\sigma)e^{-x^2}dx$$

This expression now has the same functional form as the Hermite formula. The approximation from abscissas and weights can be applied. The integral is then approximated by the following summation.

$$\frac{1}{\sqrt{\pi}}\sum_{k=1}^{n} w_k f(data|\beta, \gamma = \sqrt{2}\sigma a_k)$$

This procedure can be extended to higher dimensions of integrals. When there are two variables to be integrated, $X = (x_1\ x_2)$ for instance, the Hermite formula is written as the following.

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-X'X} f(X)dX \approx \sum_{j=1}^{n}\sum_{k=1}^{n} w_j w_k f(a_j, a_k)$$

This extension was applied in the model with two random effects. The maximization was then performed over the marginal likelihood, i.e., the likelihood integrating out all random effects. In order to acquire the marginal likelihood, the two random effects were integrated out from the joint distribution. Specifically, for the i^{th} study, if $\begin{pmatrix} \gamma_{i1} \\ \gamma_{i2} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}\right)$, the integral to derive the marginal likelihood is displayed below.

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(data|\beta, \gamma_{i1}, \gamma_{i2})\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{\gamma_{i1}^2}{2\sigma_1^2}}\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{\gamma_{i1}^2}{2\sigma_2^2}}d\gamma_{i1}d\gamma_{i2}$$

Again, this did not fit into the Hermite formula directly. In order to match the standard expression, transformation was required. Let $\frac{\gamma_{i1}}{\sqrt{2}\sigma_1} = a_{i1}$ and $\frac{\gamma_{i2}}{\sqrt{2}\sigma_2} = a_{i2}$, then $d\gamma_{i1} = \sqrt{2}\sigma_1 da_{i1}$ and $d\gamma_{i2} = \sqrt{2}\sigma_2 da_{i2}$. With these transformations, the integral above can be re-written as the following.

$$\frac{1}{\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(data|\beta, \gamma_{i1} = \sqrt{2}\sigma_1 a_{i1}, \gamma_{i2} = \sqrt{2}\sigma_2 a_{i2}) e^{-a_{i1}^2} e^{-a_{i2}^2} da_{i1} da_{i2}$$

The transformed integral matches the Hermite formula and is approximated by the following expression.

$$\frac{1}{\pi}\sum_{l=1}^{n}\sum_{q=1}^{n} w_l w_q f(data|\beta, \gamma_{i1} = \sqrt{2}\sigma_1 a_{li1}, \gamma_{i2} = \sqrt{2}\sigma_2 a_{qi2})$$

In this expression, $f(data|\beta, \gamma_{i1} = \sqrt{2}\sigma_1 a_{li1}, \gamma_{i2} = \sqrt{2}\sigma_2 a_{qi2})$ is the likelihood, where positions of γ_{i1} and γ_{i2} are replaced by the functions of abscissas. The functional form of the likelihood was derived from the log-linear model, which involved two random effects as presented in previous sections. The log likelihood of the i^{th} DU table, for example, was calculated as the following.

$$\log\left(L_{du}^{i}\right) = x_{du}^{i} \log\left(p_{du}\right)$$

The x_{du}^{i} was the vector of observed cell counts in the i^{th} DU table. The cell probabilities were expressions in terms of fixed effect coefficients and random effects. Similarly, the log likelihoods of the j^{th} UV table and the k^{th} DV table were given by the following expressions.

$$\log\left(L_{uv}^{j}\right) = x_{uv}^{j} \log(p_{uv})$$
$$\log\left(L_{dv}^{k}\right) = x_{dv}^{k} \log(p_{dv})$$

Treating L^i_{du}, for instance, as $f(data|\beta, \gamma_{i1}, \gamma_{i2})$ and replacing γ_{i1} and γ_{i2} with $\sqrt{2}\sigma_1 a_{li1}$ and $\sqrt{2}\sigma_1 a_{qi2}$, the marginal likelihood of the i^{th} DU table was approximated by the following expression.

$$L^i_{du} =$$

$$\frac{1}{\pi}\sum_{l=1}^{n}\sum_{q=1}^{n} w_l w_q e^{x^i_{du} log(duMat \times e^{X\beta + Z(\sqrt{2}\sigma_1 a_{li1}, \sqrt{2}\sigma_2 a_{qi2})}) - N^i_{du} log(\sum e^{X\beta + Z(\sqrt{2}\sigma_1 a_{li1}, \sqrt{2}\sigma_2 a_{qi2})})}$$

In this expression, duMat was the design matrix to obtain the marginal cell probabilities of the DU table.

Similarly, the marginal likelihoods of the j^{th} UV and k^{th} DV tables can be derived. The product of these marginal likelihoods composed the marginal likelihood of observed tables and fixed effect coefficients. This was the function to be maximized with respect to β, σ_1 and σ_2. The Gaussian Hermite integration procedure for this project can be summarized in the steps below.

- Write out the likelihood for each table.
- Use the Gaussian Hermite abscissas and weights to approximate the integral of joint likelihood and acquire the marginal likelihood (the function with no random effects).
- Take the product of the marginal likelihoods and maximize this product using conventional Newton-Raphson type algorithm.

Note that the variances of random effects σ_1^2 and σ_2^2 were parameters in the marginal likelihood functions. Maximizations were performed to acquire estimates for these two quantities along with the fixed effect coefficients. A trick to avoid negative variances, however, was suggested. In the function for maximization, σ_1^2 and

σ_2^2 were re-written as $\sigma_1^2 = e^{\log \sigma_1^2}$ and $\sigma_2^2 = e^{\log \sigma_2^2}$. The $\log \sigma_1^2$ and $\log \sigma_2^2$ became the parameters to be estimated. At the convergence of the Newton-Raphson algorithm, estimates of $\log \sigma_1^2$, $\log \sigma_2^2$ and β were acquired. Exponentiations of $\log \sigma_1^2$ and $\log \sigma_2^2$ provided estimates of σ_1^2 and σ_2^2.

In addition, the choice of number of abscissas and weights was considered. Although "20" was the conventional choice in most generalized linear mixed models (GLMM), estimates using the 20-point abscissas were not stable for the likelihood in this problem. The number was increased to "25" to improve accuracy at the cost of computational intensity. Stability was achieved when the number of abscissas and weights was changed to 25. Abscissas and weights for 25 points were derived from the ghq() function in the glmmML package in R. Values of 20 points abscissas and weights from this source were verified with Tables 25.9 and 25.10 in the book by Abramowitz and Stegun [51]. This evaluation was to ensure correct values of 25-point abscissas and weights produced by the ghq() function in R.

Unconditional cell probabilities

Recall that the log-linear model with random effects took the following form.

$$\log\left(m_{duv|\beta,\gamma}\right) = X\beta + Z\gamma$$

The corresponding cell probabilities were conditional on the random effects in this model. The parameters of interest in the context of diagnosing DVT, however, were the sensitivity and specificity of d-dimer, which were based upon unconditional cell probabilities. In other words, the estimates of sensitivity and specificity of d-dimer

from the meta-analysis should represent the overall performance attributes of d-dimer rather than study-specific characteristics. In order to obtain these estimates, an additional step was required to derive unconditional sensitivity and specificity of d-dimer. The functions being integrated were the cell probabilities.

$$p_{duv|\beta,\gamma} = \frac{e^{X\beta+Z\gamma}}{\sum e^{X\beta+Z\gamma}}$$

Integrations of these probabilities were performed over the random effects in order to derive unconditional cell probabilities. Using the Gaussian Hermite abscissas and weights again, the integrands of cell probabilities can be approximated. Note that each row of above function represented a specific cell probability and hence the function was integrated row-by-row with respect to the random effects. For example, the r^{th} row of the X and Z matrix corresponded to the r^{th} cell probability, where r=1,2,...8. Approximation of this probability by the abscissas and weights was expressed below.

$$\frac{1}{\pi}\sum_{l=1}^{n}\sum_{q=1}^{n} w_l w_q \frac{e^{X[r,]\beta+Z[r,]\gamma}}{\sum e^{X\beta+Z\gamma}}$$

Again in this expression, γ was replaced by the vector ($\sqrt{2}\sigma_1 a_{l1}$, $\sqrt{2}\sigma_2 a_{q2}$). The estimated sensitivity and specificity can be obtained using the unconditional cell probabilities.

In the estimation of model coefficients, the observed information matrix can be derived. It is the negative of second derivatives of the log likelihood with respect

to each parameter. The inverse of the observed information matrix produced the asymptotic variance covariance matrix of the model coefficients β. Based on this matrix, the estimated variance covariance matrix of cell probabilities can be derived, namely $\hat{V}_{p_{duv}}$, via the delta method. Let $\hat{V}_\beta$ denote the estimated variance covariance matrix of $(\beta_1\ \beta_2\ \beta_3\ \beta_4\ \beta_5)$. The cell probabilities p_{duv} were functions of these five βs. Derivatives of p_{duv} with respect to βs were required in order to use the delta method. Let $D_{p\beta}$ denote the derivative matrix of cell probabilities with respect to β. It follows from the delta method that $\hat{V}_{p_{duv}} = D_{p\beta}\hat{V}_\beta D'_{p\beta}$. The derivatives of cell probabilities with respect to β were calculated below.

Derivatives of p_{duv} with respect to β

The expressions of p_{duv} in terms of β and γ were listed below.

$p_{111} = e^{\beta_0+\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2}$

$p_{011} = e^{\beta_0+\beta_2+\beta_3+\beta_5+\gamma_1-\gamma_2}$

$p_{101} = e^{\beta_0+\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2}$

$p_{001} = e^{\beta_0+\beta_3+\gamma_1-\gamma_2}$

$p_{110} = e^{\beta_0+\beta_1+\beta_2-\gamma_1-\gamma_2}$

$p_{010} = e^{\beta_0+\beta_2-\gamma_1+\gamma_2}$

$p_{100} = e^{\beta_0+\beta_1-\gamma_1-\gamma_2}$

$p_{000} = e^{\beta_0-\gamma_1+\gamma_2}$

Again, note that β_0 was not an independent parameter. It can be calculated as the following. Because $\sum p_{duv} = e^{\beta_0}(e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2}+e^{\beta_2+\beta_3+\beta_5+\gamma_1-\gamma_2}+e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2}+e^{\beta_3+\gamma_1-\gamma_2}+e^{\beta_1+\beta_2-\gamma_1-\gamma_2}+e^{\beta_2-\gamma_1+\gamma_2}+e^{\beta_1-\gamma_1-\gamma_2}+e^{-\gamma_1+\gamma_2}) = 1$. It follows that $\beta_0 =$ -log $(\sum e^{X\beta+Z\gamma})$.

Let $\sum e^{X\beta+Z\gamma}$ denote $e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_2+\beta_3+\beta_5+\gamma_1-\gamma_2} + e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2} + e^{\beta_3+\gamma_1-\gamma_2} + e^{\beta_1+\beta_2-\gamma_1-\gamma_2} + e^{\beta_2-\gamma_1+\gamma_2} + e^{\beta_1-\gamma_1-\gamma_2} + e^{-\gamma_1-\gamma_2}$.

$$p_{111} = \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}}$$
$$p_{011} = \frac{e^{\beta_2+\beta_3+\beta_5+\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}}$$
$$p_{101} = \frac{e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}}.$$
$$p_{001} = \frac{e^{\beta_3+\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}}$$
$$p_{110} = \frac{e^{\beta_1+\beta_2-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}}$$
$$p_{010} = \frac{e^{\beta_2-\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}}$$
$$p_{100} = \frac{e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}}$$
$$p_{000} = \frac{e^{-\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}}$$

Derivatives of p_{duv} with respect to β were based on the re-written expressions of p_{duv}. Several steps are displayed below to obtain the derivatives. First of all, derivatives of the denominator of cell probabilities with respect to each β were calculated.

$$\frac{\partial \sum e^{X\beta+Z\gamma}}{\partial \beta_1} = e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2} + e^{\beta_1+\beta_2-\gamma_1-\gamma_2} + e^{\beta_1-\gamma_1-\gamma_2}$$
$$\frac{\partial \sum e^{X\beta+Z\gamma}}{\partial \beta_2} = e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_2+\beta_3+\beta_5+\gamma_1-\gamma_2} + e^{\beta_1+\beta_2-\gamma_1-\gamma_2} + e^{\beta_2-\gamma_1+\gamma_2}$$
$$\frac{\partial \sum e^{X\beta+Z\gamma}}{\partial \beta_3} = e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_2+\beta_3+\beta_5+\gamma_1-\gamma_2} + e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2} + e^{\beta_3+\gamma_1-\gamma_2}$$
$$\frac{\partial \sum e^{X\beta+Z\gamma}}{\partial \beta_4} = e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2}$$
$$\frac{\partial \sum e^{X\beta+Z\gamma}}{\partial \beta_5} = e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_2+\beta_3+\beta_5+\gamma_1-\gamma_2}$$

Using the derivatives of the denominator with respect to each coefficient, the derivatives of cell probabilities can be derived. For example, the derivative of p_{111} with respect to β_1 can be calculated as the following.

$$\frac{\partial p_{111}}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \int \int_{-\infty}^{+\infty} \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$
$$= \int \int_{-\infty}^{+\infty} \frac{\sum e^{X\beta+Z\gamma} \times e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} - e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} \times \frac{\partial \sum e^{X\beta+Z\gamma}}{\partial \beta_1}}{(\sum e^{X\beta+Z\gamma})^2} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

$$= \int\int_{-\infty}^{+\infty} \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2 - \int\int_{-\infty}^{+\infty} \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}} \times$$
$$\frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2}+e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2}+e^{\beta_1+\beta_2-\gamma_1-\gamma_2}+e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2$$

Note that the first integral in this expression was actually the unconditional probability p_{111} after integrating out the two random effects. Let p_{duv} denote the probabilities integrating out random effects and $p_{duv|\gamma_1,\gamma_2}$ denote the conditional probabilities on random effects, the above derivative was equivalent to the following expression.

$$\frac{\partial p_{111}}{\partial \beta_1} = p_{111} - \int\int_{-\infty}^{+\infty} p_{111|\gamma_1,\gamma_2} \times (p_{111|\gamma_1,\gamma_2} + p_{101|\gamma_1,\gamma_2} + p_{110|\gamma_1,\gamma_2} + p_{100|\gamma_1,\gamma_2}) f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2$$

Following similar procedures, derivatives of other cell probabilities are:

$$\frac{p_{011}}{\partial \beta_1} = \int\int_{-\infty}^{+\infty} \frac{\partial}{\partial \beta_1} \frac{e^{\beta_2+\beta_3+\beta_5+\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2$$
$$= 0 - \int\int_{-\infty}^{+\infty} \frac{e^{\beta_2+\beta_3+\beta_5+\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} \times$$
$$\frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2}+e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2}+e^{\beta_1+\beta_2-\gamma_1-\gamma_2}+e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2$$
$$= 0 - \int\int_{-\infty}^{+\infty} p_{011|\gamma_1,\gamma_2} \times (p_{111|\gamma_1,\gamma_2}+p_{101|\gamma_1,\gamma_2}+p_{110|\gamma_1,\gamma_2}+p_{100|\gamma_1,\gamma_2}) f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2$$

$$\frac{\partial p_{101}}{\partial \beta_1} = \int\int_{-\infty}^{+\infty} \frac{\sum e^{X\beta+Z\gamma} \times e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2} - e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2} \times \frac{\partial \sum e^{X\beta+Z\gamma}}{\partial \beta_1}}{(\sum e^{X\beta+Z\gamma})^2} f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2$$
$$= \int\int_{-\infty}^{+\infty} \frac{e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2 - \int\int_{-\infty}^{+\infty} \frac{e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}} \times$$
$$\frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2}+e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2}+e^{\beta_1+\beta_2-\gamma_1-\gamma_2}+e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2$$
$$= p_{101} - \int\int_{-\infty}^{+\infty} p_{101|\gamma_1,\gamma_2} \times (p_{111|\gamma_1,\gamma_2}+p_{101|\gamma_1,\gamma_2}+p_{110|\gamma_1,\gamma_2}+p_{100|\gamma_1,\gamma_2}) f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2$$

$$\frac{\partial p_{001}}{\partial \beta_1} = 0 - \int\int_{-\infty}^{+\infty} \frac{e^{\beta_3+\gamma_1-\gamma_2} \times \frac{\partial \sum e^{X\beta+Z\gamma}}{\partial \beta_1}}{(\sum X\beta+Z\gamma)^2} f(\gamma_1,\gamma_2)d\gamma_1 d\gamma_2$$

$$= 0 - \int\int_{-\infty}^{+\infty} p_{001|\gamma_1,\gamma_2} \times (p_{111|\gamma_1,\gamma_2} + p_{101|\gamma_1,\gamma_2} + p_{110|\gamma_1,\gamma_2} + p_{100|\gamma_1,\gamma_2}) f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

$$\frac{\partial p_{110}}{\partial \beta_1} = \int\int_{-\infty}^{+\infty} \frac{e^{\beta_1+\beta_2-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2 - \int\int_{-\infty}^{+\infty} \frac{e^{\beta_1+\beta_2-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} \times$$
$$\frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2} + e^{\beta_1+\beta_2-\gamma_1-\gamma_2} + e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$
$$= p_{110} - \int\int_{-\infty}^{+\infty} p_{110|\gamma_1,\gamma_2} \times (p_{111|\gamma_1,\gamma_2} + p_{101|\gamma_1,\gamma_2} + p_{110|\gamma_1,\gamma_2} + p_{100|\gamma_1,\gamma_2}) f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

$$\frac{\partial p_{010}}{\partial \beta_1} = 0 - \int\int_{-\infty}^{+\infty} \frac{e^{\beta_2-\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}} \times$$
$$\frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2} + e^{\beta_1+\beta_2-\gamma_1-\gamma_2} + e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$
$$= 0 - \int\int_{-\infty}^{+\infty} p_{010|\gamma_1,\gamma_2} \times (p_{111|\gamma_1,\gamma_2} + p_{101|\gamma_1,\gamma_2} + p_{110|\gamma_1,\gamma_2} + p_{100|\gamma_1,\gamma_2}) f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

$$\frac{\partial p_{100}}{\partial \beta_1} = \int\int_{-\infty}^{+\infty} \frac{e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2 - \int\int_{-\infty}^{+\infty} \frac{e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} \times$$
$$\frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2} + e^{\beta_1+\beta_2-\gamma_1-\gamma_2} + e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$
$$= p_{100} - \int\int_{-\infty}^{+\infty} p_{100|\gamma_1,\gamma_2} \times (p_{111|\gamma_1,\gamma_2} + p_{101|\gamma_1,\gamma_2} + p_{110|\gamma_1,\gamma_2} + p_{100|\gamma_1,\gamma_2}) f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

$$\frac{\partial p_{000}}{\partial \beta_1} = 0 - \int\int_{-\infty}^{+\infty} \frac{e^{-\gamma_1+\gamma_2}}{\sum e^{X\beta+Z\gamma}}$$
$$\times \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5+\gamma_1+\gamma_2} + e^{\beta_1+\beta_3+\beta_4+\gamma_1+\gamma_2} + e^{\beta_1+\beta_2-\gamma_1-\gamma_2} + e^{\beta_1-\gamma_1-\gamma_2}}{\sum e^{X\beta+Z\gamma}} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$
$$= 0 - \int\int_{-\infty}^{+\infty} p_{000|\gamma_1,\gamma_2} \times (p_{111|\gamma_1,\gamma_2} + p_{101|\gamma_1,\gamma_2} + p_{110|\gamma_1,\gamma_2} + p_{100|\gamma_1,\gamma_2}) f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

In matrix format, the above derivatives can be summarized as the following expressions, where $p_{1..|\gamma_1,\gamma_2} = p_{111|\gamma_1,\gamma_2} + p_{101|\gamma_1,\gamma_2} + p_{110|\gamma_1,\gamma_2} + p_{100|\gamma_1,\gamma_2}$.

$$\frac{\partial p_{duv}}{\partial \beta_1} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} - \int \int_{-\infty}^{+\infty} \begin{pmatrix} p_{111|\gamma_1,\gamma_2} \\ p_{011|\gamma_1,\gamma_2} \\ p_{101\gamma_1,\gamma_2} \\ p_{001|\gamma_1,\gamma_2} \\ p_{110|\gamma_1,\gamma_2} \\ p_{010|\gamma_1,\gamma_2} \\ p_{100|\gamma_1,\gamma_2} \\ p_{000|\gamma_1,\gamma_2} \end{pmatrix} \times p_{1..|\gamma_1,\gamma_2} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

By similar procedures, derivatives of p_{duv} with respect to β_2, β_3, β_4, and β_5 were displayed below in matrix formats.

$$\frac{\partial p_{duv}}{\partial \beta_2} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} - \int \int_{-\infty}^{+\infty} \begin{pmatrix} p_{111|\gamma_1,\gamma_2} \\ p_{011|\gamma_1,\gamma_2} \\ p_{101\gamma_1,\gamma_2} \\ p_{001|\gamma_1,\gamma_2} \\ p_{110|\gamma_1,\gamma_2} \\ p_{010|\gamma_1,\gamma_2} \\ p_{100|\gamma_1,\gamma_2} \\ p_{000|\gamma_1,\gamma_2} \end{pmatrix} \times p_{.1.|\gamma_1,\gamma_2} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

$$\frac{\partial p_{duv}}{\partial \beta_3} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} - \int \int_{-\infty}^{+\infty} \begin{pmatrix} p_{111|\gamma_1,\gamma_2} \\ p_{011|\gamma_1,\gamma_2} \\ p_{101\gamma_1,\gamma_2} \\ p_{001|\gamma_1,\gamma_2} \\ p_{110|\gamma_1,\gamma_2} \\ p_{010|\gamma_1,\gamma_2} \\ p_{100|\gamma_1,\gamma_2} \\ p_{000|\gamma_1,\gamma_2} \end{pmatrix} \times p_{..1|\gamma_1,\gamma_2} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

$$\frac{\partial p_{duv}}{\partial \beta_4} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} - \int \int_{-\infty}^{+\infty} \begin{pmatrix} p_{111|\gamma_1,\gamma_2} \\ p_{011|\gamma_1,\gamma_2} \\ p_{101\gamma_1,\gamma_2} \\ p_{001|\gamma_1,\gamma_2} \\ p_{110|\gamma_1,\gamma_2} \\ p_{010|\gamma_1,\gamma_2} \\ p_{100|\gamma_1,\gamma_2} \\ p_{000|\gamma_1,\gamma_2} \end{pmatrix} \times p_{1.1|\gamma_1,\gamma_2} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

$$\frac{\partial p_{duv}}{\partial \beta_5} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{111} \\ p_{011} \\ p_{101} \\ p_{001} \\ p_{110} \\ p_{010} \\ p_{100} \\ p_{000} \end{pmatrix} - \int\int_{-\infty}^{+\infty} \begin{pmatrix} p_{111|\gamma_1,\gamma_2} \\ p_{011|\gamma_1,\gamma_2} \\ p_{101\gamma_1,\gamma_2} \\ p_{001|\gamma_1,\gamma_2} \\ p_{110|\gamma_1,\gamma_2} \\ p_{010|\gamma_1,\gamma_2} \\ p_{100|\gamma_1,\gamma_2} \\ p_{000|\gamma_1,\gamma_2} \end{pmatrix} \times p_{.11|\gamma_1,\gamma_2} f(\gamma_1,\gamma_2) d\gamma_1 d\gamma_2$$

The summary matrix of partial derivatives can be constructed below.

$$D_{p\beta} = \frac{\partial p_{jik}}{\partial \beta} = \begin{pmatrix} \frac{\partial p_{000}}{\partial \beta_1} & \frac{\partial p_{000}}{\partial \beta_2} & \frac{\partial p_{000}}{\partial \beta_3} & \frac{\partial p_{000}}{\partial \beta_4} & \frac{\partial p_{000}}{\partial \beta_5} \\ \frac{\partial p_{100}}{\partial \beta_1} & \frac{\partial p_{100}}{\partial \beta_2} & \frac{\partial p_{100}}{\partial \beta_3} & \frac{\partial p_{100}}{\partial \beta_4} & \frac{\partial p_{100}}{\partial \beta_5} \\ \frac{\partial p_{010}}{\partial \beta_1} & \frac{\partial p_{010}}{\partial \beta_2} & \frac{\partial p_{010}}{\partial \beta_3} & \frac{\partial p_{010}}{\partial \beta_4} & \frac{\partial p_{010}}{\partial \beta_5} \\ \frac{\partial p_{110}}{\partial \beta_1} & \frac{\partial p_{110}}{\partial \beta_2} & \frac{\partial p_{110}}{\partial \beta_3} & \frac{\partial p_{110}}{\partial \beta_4} & \frac{\partial p_{110}}{\partial \beta_5} \\ \frac{\partial p_{001}}{\partial \beta_1} & \frac{\partial p_{001}}{\partial \beta_2} & \frac{\partial p_{001}}{\partial \beta_3} & \frac{\partial p_{001}}{\partial \beta_4} & \frac{\partial p_{001}}{\partial \beta_5} \\ \frac{\partial p_{101}}{\partial \beta_1} & \frac{\partial p_{101}}{\partial \beta_2} & \frac{\partial p_{101}}{\partial \beta_3} & \frac{\partial p_{101}}{\partial \beta_4} & \frac{\partial p_{101}}{\partial \beta_5} \\ \frac{\partial p_{011}}{\partial \beta_1} & \frac{\partial p_{011}}{\partial \beta_2} & \frac{\partial p_{011}}{\partial \beta_3} & \frac{\partial p_{011}}{\partial \beta_4} & \frac{\partial p_{011}}{\partial \beta_5} \\ \frac{\partial p_{111}}{\partial \beta_1} & \frac{\partial p_{111}}{\partial \beta_2} & \frac{\partial p_{111}}{\partial \beta_3} & \frac{\partial p_{111}}{\partial \beta_4} & \frac{\partial p_{111}}{\partial \beta_5} \end{pmatrix}$$

Based on these derivatives and the variance matrix of β, the estimated variance matrix of p_{duv} can be derived using the delta method.

$$\hat{V}_{p_{duv}} = D_{p\beta} \hat{V}_{\beta} D'_{p\beta}$$

Variance estimation of $SENS_d$ and $SPEC_d$

In order to acquire the estimated variance covariance matrix of $\hat{SENS}_d$ and $\hat{SPEC}_d$, the delta method was applied again by using derivatives of $SENS_d$ and $SPEC_d$ with respect to p_{duv}.

$$D_{scp} = \begin{pmatrix} 0 & 0 & 0 & 0 & -\frac{p_{101}+p_{111}}{p^2_{..1}} & \frac{p_{001}+p_{011}}{p^2_{..1}} & -\frac{p_{101}+p_{111}}{p^2_{..1}} & \frac{p_{001}+p_{011}}{p^2_{..1}} \\ \frac{p_{100}+p_{110}}{p^2_{..0}} & -\frac{p_{010}+p_{000}}{p^2_{..0}} & \frac{p_{100}+p_{110}}{p^2_{..0}} & -\frac{p_{010}+p_{000}}{p^2_{..0}} & 0 & 0 & 0 & 0 \end{pmatrix}$$

The estimated variance covariance matrix of $\hat{SENS}_d$ and $\hat{SPEC}_d$ was derived by:

$$\hat{V}_{sc} = D_{scp}\hat{V}_{p_{duv}}D'_{scp}.$$

3.3.4 Estimations using the Gibbs sampling

The Gibbs sampling

As the number of random effects in the generalized linear mixed model (GLMM) increased, the computational burden from the Gaussian Hermite integration increased. Bayesian techniques served as alternatives and have become more and more popular for solving complex statistical models. In particular, the Gibbs sampling is widely applied in Bayesian models, especially those with high-dimensional hierarchical structure [46, 62, 80, 84, 102]. The theory behind Gibbs sampling can be summarized as the following. Suppose that the full conditional distributions for three variables, X, Y, and Z, were available. In other words, f(X|Y, Z), f(Y|X,Z) and f(Z|X,Y) were known density functions. Given a set of starting values of X, Y, and Z, draws from the full conditional distributions were obtained for each of X, Y, and Z. For a large number of consecutive draws, the joint distribution of (X, Y, Z) can be approxi-

mated by the sample distribution of (X, Y, Z) at convergence. The algorithm can be summarized in the following steps.

1. Give a set of starting values (X^0, Y^0, and Z^0).
2. Sample X^{i+1} from $f_x(X|Y^i, Z^i)$.
3. Sample Y^{i+1} from $f_y(Y|X^{i+1}, Z^i)$.
4. Sample Z^{i+1} from $f_z(Z|X^{i+1}, Y^{i+1})$.
5. Repeat steps 2 - 4 until convergence.

Geman and Geman [81] showed that the sample distribution of (X, Y, Z) converges exponentially to the joint distribution $f_{xyz}(X, Y, Z)$ as the number of iterations approaches infinity. Adapting the idea from the Gibbs sampling, Zeger and Karim proposed an algorithm to overcome computational difficulties in analyzing the generalized linear mixed model [94]. This project applied the algorithm from Zeger and Karim with modifications to perform the meta-analysis of d-dimer.

Before elaborating details of the algorithm, assumptions of the procedure were considered. As proposed by Zeger and Karim, the distribution of fixed effect coefficients β conditional on random effects was independent of the variances of random effects. Similarly, the distribution of the variances of random effects was independent of the fixed effects. These assumptions can be summarized as the following.

- $f(\beta|\gamma, data, \sigma_1, \sigma_2) = f(\beta|\gamma, data)$
- $f(\sigma_1, \sigma_2|\beta, \gamma, data) = f(\sigma_1, \sigma_2|\gamma)$

The design matrices in the model, X and Z, were the same as before. The parameters to be estimated were β, σ_1 and σ_2.

Full conditional distributions of the parameters

The full conditional distributions of these parameters were derived as the following. As stated in Zeger and Karim's paper, the conditional distribution of β is approximated by a multivariate normal distribution with mean $\hat{\beta}$ and variance $\hat{V}_\beta$. The estimated parameters $\hat{\beta}$ and $\hat{V}_\beta$ were obtained by solving the log-linear model with random effects at the values from the previous step. $\hat{\beta}$ is the vector of estimated coefficients from the log-linear model and $\hat{V}_\beta$ is the inverse of the Fisher information matrix. In our problem, the Fisher information matrix was replaced by the observed information matrix. An updated sample of the fixed effect coefficients β^* was acquired by taking a sampled value from the multivariate Gaussian distribution, $\mathrm{N}(\hat{\beta}, \hat{V}_\beta)$.

Updates of variances of random effects from the conditional distribution of $\Sigma=(\sigma_1^2, \sigma_2^2)$ given the random effects γ was produced by the following steps.

1. Calculate $S = \sum_{i=1}^{I} \gamma_i \times \gamma_i'$, where I is the number of independent studies.
2. Calculate the Choleski decomposition of S^{-1}, denoted by H, i.e., $S^{-1} = H'H$.
3. Generate W* from Wishart distribution with $I - q + 1$ degrees of freedom and parameter S, where q is the number of random effects in the model.
4. The variance matrix of random effects is updated by $\Sigma = (H'W^*H)^{-1}$.

In our problem, the number of clusters was the number of independent studies, i.e., tables of diagnostic test results from different studies. The number of random

effects in the model was 2, i.e., $q = 2$. In addition, the random effects were assumed independent. So the off-diagonal elements of Σ were zero.

Zeger and Karim claimed that the full conditional distribution of random effects γ given the data and current values of fixed effects coefficients β did not have a closed form and must be derived by numerical techniques. The idea was to find the mode and curvature of the joint distribution and to apply the rejection sampling to acquire a sampling point of the random effect. This sampling point can be considered as an updated value from the conditional distribution of random effects. For conventional GLMM with complete data, the mode and curvature can be derived by the iterative weighted least squares as mentioned by Zeger and Karim. In our problem, however, the form of the likelihood was different from classic models where complete data were available. The joint likelihood was the product of likelihoods from all marginal tables. The solutions of $\hat{\gamma}_i$ and estimated variances proposed by Zeger and Karim were not applicable to our problem. Instead, the likelihood was treated as a function of the unknown random effect parameters. Maximization of the joint distribution, which was the product of the likelihood and distribution of random effects, can be achieved through the Newton-type algorithm. Given values of β and variances of random effects from the previous step, the joint likelihood is a function of the random effects only. In this circumstance, the mode and curvature of the joint distribution can be derived by obtaining the maximum likelihood estimates (MLE) of random effects. Because β and Σ are independent, the joint distribution is given by the following expression.

$$f(data_i|\gamma_i, \beta)f(\gamma_i|\Sigma)f(\beta, \Sigma) = f(data_i|\gamma_i, \beta)f(\gamma_i|\Sigma)f(\Sigma)$$

In this expression, $f(data_i|\gamma,\beta)$ is the likelihood from the i^{th} table, namely, L^i_{du}. Because $f(\Sigma)$ is the distribution of the variance evaluated at the updated value of variance from previous step, it is considered a "constant" with respect to the random effects γ_i. It is a proportional factor in the maximization of the joint distribution and can be removed. The joint distribution is then expressed as the product of the likelihood and the distribution of random effects.

$$L^i_{du} \times f(\gamma_i|\Sigma),$$

where $f(\gamma_i|\Sigma)$ is the bivariate normal density. Again, omitting the "constant" in the density, $f(\gamma_i|\Sigma)$ is proportional to:

$$f(\gamma_i|\Sigma) \propto e^{-\frac{1}{2}(\gamma_i-0)'\Sigma^{-1}(\gamma_i-0)}.$$

Note that the vector of random effects γ_i for the i^{th} cluster is independent of random effects for other clusters. Information in the i^{th} likelihood, L^i, is related to γ_i only. Given the values of β, maximization of L^i with respect to γ_i is not affected by likelihoods obtained from other tables. This is different from the estimation of fixed effects β, in which all the tables consist of information about the common β parameters. The random effects are cluster-specific. Given current values of β, maximization of the joint likelihood with respect to γ_i for each cluster is the same as maximization of likelihood from each cluster.

Let $\hat{\gamma}_i$ and $\hat{v}_i$ denote the estimated mode and curvature of the joint distribution, $p(\gamma_i)$. It is expressed as $p(\gamma_i) = f(data_i|\beta,\gamma_i)f(\gamma_i|\Sigma)$. Denote another Gaussian density $g(\gamma_i)$ with estimated mean $\hat{\gamma}_i$ and variance $c_2\hat{v}_i$, i.e., N($\hat{\gamma}_i$, $c_2\hat{v}_i$). The rejection sampling algorithm was then applied to update the values of random effects.

1. generate γ_i^* from $g(\gamma_i) = \mathrm{N}(\hat{\gamma}_i,\, c_2\hat{v}_i)$
2. calculate $c_{1i} = \frac{p(\hat{\gamma}_i)}{g(\hat{\gamma}_i)}$
3. generate a uniform (0, 1) random value u and let $\gamma_i^{(k+1)} = \gamma_i^*$ if $\frac{p(\gamma_i^*)}{c_{1i}g(\gamma_i^*)} < u$, otherwise return to step 1.

Zeger and Karim suggested $c_2 = 2$. By the end of these three steps, the vector of random effects was updated.

The Gibbs sampling procedure

The full conditional distribution of each parameter given the rest was constructed above. The Gibbs sampling algorithm can be summarized in the following steps.

1. Specify initial values of $\gamma_i^{(0)}$.
2. Estimate fixed effects $\hat{\beta}$ and variance $\hat{V}_\beta$ based on the likelihood of all the data with current values of $\gamma_i^{(k)}$.
3. Update values of $\beta^{(k+1)}$ by sampling from $\mathrm{N}(\hat{\beta},\, \hat{V}_\beta)$.
4. Update values of variance of random effects $\Sigma^{(k)}$ based on $\gamma^{(k)}$.
5. Estimate random effects $\hat{\gamma}$ based on the likelihood with $\beta^{(k+1)}$ and the pre-defined distribution of random effects with variance $\Sigma^{(k)}$.
6. Apply the rejection algorithm to update the random effects $\gamma^{(k+1)}$.
7. Repeat steps 2 - 6 until convergence.

Before implementing the above algorithm, the choice of the number of MCMC iterations was considered. This issue is related to the convergence assessment in MCMC. Although Cowles had reviewed a broad range of methods of convergence assessment, there was not a concrete conclusion on which method was superior over the others. Zeger and Karim pointed out that the variance of random effects was the slowest to converge. In particular, if the variance was small, the sequence would have extreme long-term dependence. In this problem, 5000, as in most MCMC sampler, was chosen to be the number of iterations. With 2000 burn-in, a total of 7000 MCMC iterations was determined. Convergence of the Gibbs samples was assessed by histograms of posterior sample distributions of the parameters. A normal density curve centered at the sample mean and with sample variance as curvature parameter was super-imposed on the histogram of each parameter.

Theoretically speaking, the choice of initial values did not affect the estimation results. The choice of initial values, however, *did* have strong impacts on the rate of convergence of MCMC, especially in the case of slow convergence. The initial values of random effects in this project were sampled from standard normal distribution. Recall that the distribution of random effects was assumed as the normal density with mean zero and unknown variance. The sampling distribution of initial values was similar to the pre-defined distribution but with variance set at one.

Chapter 4

Application

4.1 Description of d-dimer data

Chapter 3 provided expositions of different algorithms to analyze the d-dimer data, where the complete three-dimensional contingency tables were not available. The meta-analysis by Stein and colleagues [68] synthesized diagnostic data from studies using different cutoffs and different assays. As with other diagnostic tests, the choice of cutoff for positivity affected the sensitivity and specificity of d-dimer. The assays differed in sensitivity, specificity and variability among patients with suspected deep vein thrombosis (DVT). For this project, data extracted from the d-dimer paper were confined to a particular cutoff value and assay. The cutoff chosen was "500" and the assay was "SL". The combination of 500 cutoff and "SL" yielded the largest number of studies among other choices. Within this set of studies, tables from studies using either ultrasonography or venography, but not the both, were selected.

Test property	1	2	3	4
true positive	16	32	21	31
false negative	5	5	15	7
false positive	4	11	6	24
true negative	28	21	54	33
sensitivity	0.76	0.865	0.59	0.816
specificity	0.87	0.656	0.90	0.571

Table 4.1: DV tables from d-dimer study [68]

The DV and DU tables from d-dimer study are summarized in Tables 4.1 and 4.2,

Test property	1	2	3
true positive	44	55	25
false negative	1	20	4
false positive	43	36	33
true negative	12	60	46
sensitivity	0.98	0.73	0.862
specificity	0.22	0.625	0.582

Table 4.2: DU tables from d-dimer study [68]

respectively. The third type of marginal tables was ultrasonography versus venography, which was not specified in the paper of d-dimer [68]. A review article of the sensitivity and specificity of ultrasonography was selected [6]. In this review, 8 studies were assessed. The five studies with complete data on sensitivity and specificity of ultrasonography were chosen. Data from tables between ultrasonography

Test property	1	2	3	4	5
true positive	7	14	26	19	17
false negative	1	11	1	0	4
false positive	0	0	2	1	6
true negative	6	18	29	14	20
sensitivity	0.88	0.56	0.96	1	0.81
specificity	1	1	0.94	0.93	0.77

Table 4.3: UV tables from the literature [6]

and venography are summarized in Table 4.3 and served as the UV tables in this project.

4.2 Outline of applications

The analysis in this chapter was based on the two types of marginal tables from the d-dimer paper [68] and the tables of ultrasonography and venography from the

review of ultrasonography [6]. The conditional independence between d-dimer and ultrasonography given venography was assumed in all models and simulations. The sensitivity and specificity of d-dimer, namely S_d and C_d, were the parameters of interest in all models and simulations. The analysis included two modeling strategies: the fixed effects model and the random effects model.

In the fixed effects model, two scenarios were considered: known sensitivity and specificity of ultrasonography versus the observed UV tables. In the first scenario, the sensitivity and specificity of ultrasonography were assumed to be 0.95. Two models were applied in the analysis: treating ultrasonography as silver standard (adjusted model) and treating ultrasonography as a perfect reference (unadjusted model). In the latter model, sensitivity and specificity of ultrasonography were not applicable, nor were the tables between ultrasonography and venography. Estimates of S_d and C_d from the two models were compared. In the simulations, the true sensitivity and specificity were chosen as 0.81 and 0.77 respectively from the review of ultrasonography. The estimated coefficients of the model in the adjusted model were used as true parameter values in the simulation. Bias and mean squared errors of estimates from simulations of the two procedures were compared.

In the second scenario, observed tables between ultrasonography and venography were incorporated. No parameter values were assumed. Two models were considered: treating ultrasonography as the silver standard (adjusted model) and treating ultrasonography as the gold standard (unadjusted model). Comparisons on the estimates and standard errors from the two models were conducted. Simulations were performed to examine and compare the properties of estimates. The estimated coefficients from the adjusted model were used as pre-defined parameter values in the

simulation. Three types of marginal tables were generated in the simulation. 5000 simulations were conducted. Bias and mean squared errors of estimates from simulations were compared.

With respect to the random effects model, two random effects were considered: random venography and random interaction between d-dimer and venography. The data for analysis were the same tables as in the fixed effects model. Two algorithms were considered: Gaussian Hermite integration and the Gibbs sampling suggested by Zeger and Karim [94]. Results from these procedures were compared in terms of estimated model coefficients and the variance of random effects. Simulations were conducted using the Gaussian Hermite integration procedure. Estimated coefficients and variances of random effects were applied as true parameter values in the simulation. In all simulations, bias, mean squared error (MSE), and coverage of 95% confidence intervals were calculated.

4.3 Fixed effects model

Under the assumption of conditional independence between d-dimer and ultrasonography, the log-linear model for the three tests was expressed as the following.

$$\log(m_{duv}) = X\beta = \beta_0 + \beta_1 \mathrm{D} + \beta_2 \mathrm{U} + \beta_3 \mathrm{V} + \beta_4 \mathrm{DV} + \beta_5 \mathrm{UV}$$

In this model, m_{duv} represented individual cell counts; D, U, and V denote d-dimer, ultrasonography, and venography, respectively. With the assumption of conditional independence between D and U, five independent coefficients, (β_1 β_2 β_3 β_4 β_5), were the parameters to be estimated. The design matrix X took the same form as that presented in Chapter 3.

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Again, β_0 is a function of the rest βs in the form of $\beta_0 = -log(\sum e^{X_{2-6}\beta_{1-5}})$, where X_{2-6} is the 2^{th} to 6^{th} columns of the X matrix and β_{1-5} is the vector of β_1 to β_5. The likelihood of the marginal table can be constructed from the observed data table, X, and β_{1-5}.

4.3.1 Analysis of two marginal tables with known sensitivity and specificity of the silver standard

The first scenario under consideration is a simple structure of available data. In this situation, only one DU table and one DV table were available. The UV table was not collected but sensitivity and specificity of ultrasonography were known as 0.95. As discussed in Chapter 3, the likelihood in this situation was subject to constraints. The information from the sensitivity and specificity of ultrasonography can be transformed into values of model coefficients. Following this principle, the likelihood can be written as a function of the remaining coefficients and maximized with regards to these coefficients. When there were more than one DU and DV tables, the logarithm of the joint likelihood can be expressed as the sum of the

logarithm of the likelihood from each table. Maximization can be performed on the joint likelihood.

Most similar DU and DV tables

Among the DU and DV tables in the previous section, the tables with similar sensitivity and specificity were selected, which occurred in the last column of both tables. The sensitivity and specificity in the DU table were 0.862 and 0.582, respectively. The sensitivity and specificity in the DV table were 0.816 and 0.571.

DV table:

tests	V+	V-
D+	31	24
D-	7	33

DU table:

tests	U+	U-
D+	25	33
D-	4	46

With the DU and DV tables specified above, cell probabilities were estimated using the Newton-type algorithm.

$\hat{p}_{000}$	$\hat{p}_{100}$	$\hat{p}_{010}$	$\hat{p}_{110}$	$\hat{p}_{001}$	$\hat{p}_{101}$	$\hat{p}_{011}$	$\hat{p}_{111}$
0.374	0.265	0.0197	0.0139	0.0025	0.0139	0.0474	0.2639

The estimated sensitivity and specificity were calculated from the estimated cell probabilities.

$$\hat{S}_d = 0.848,\ \hat{C}_d = 0.585$$

By the delta method, the estimate variances for S_d and C_d were obtained.

$\hat{V}_{S_d}$	$\hat{V}_{C_d}$
0.00217	0.00185

The corresponding standard errors were 0.047 and 0.043, respectively.

If ultrasonography was treated the same as venography, the estimated sensitivity and specificity of d-dimer were:

$$\hat{S}_d = 0.836,\ \hat{C}_d = 0.581.$$

The standard errors of $\hat{S}_d$ and $\hat{C}_d$ were 0.045 and 0.042, respectively.

Most distinct DU and DV tables

The analysis above was performed on the most similar DU and DV tables. In this section, the most different DU and DV tables, based on sensitivity and specificity, were examined. The DU table had sensitivity 0.98 and specificity 0.22. The DV table had sensitivity 0.59 and specificity 0.90.

DV table:

tests	U+	U-
D+	21	6
D-	15	54

DU table:

tests	U+	U-
D+	44	43
D-	1	12

The sensitivity and specificity of ultrasonography were assumed 0.95. The estimated cell probabilities were presented below.

$\hat{p}_{000}$	$\hat{p}_{100}$	$\hat{p}_{010}$	$\hat{p}_{110}$	$\hat{p}_{001}$	$\hat{p}_{101}$	$\hat{p}_{011}$	$\hat{p}_{111}$
0.32	0.231	0.0168	0.0122	0.00407	0.0169	0.0774	0.3215

Estimated S_d and C_d were $\hat{S}_d = 0.806$ and $\hat{C}_d = 0.581$ with standard errors 0.044 and 0.048, respectively.

If the ultrasonography was treated the same as venography, the estimated sensitivity and specificity were $\hat{S}_d = 0.802$ and $\hat{C}_d = 0.574$ with standard errors 0.044 and 0.046.

Analysis of all DU and DV tables

In the situations discussed above, only one DU and one DV tables were used in the analysis. In this section, all the available tables from the d-dimer paper [68] were ana-

lyzed. Again, the known parameters of sensitivity and specificity of ultrasonography were set at 0.95.

Using the maximum likelihood algorithm proposed in chapter 3, the cell probabilities were estimated below.

$\hat{p}_{000}$	$\hat{p}_{100}$	$\hat{p}_{010}$	$\hat{p}_{110}$	$\hat{p}_{001}$	$\hat{p}_{101}$	$\hat{p}_{011}$	$\hat{p}_{111}$
0.313	0.073	0.0165	0.00384	0.011	0.0186	0.210	0.354

Applying the estimated cell probabilities, S_d and C_d were estimated as $\hat{S}_d = 0.81$ and $\hat{C}_d = 0.63$ with standard errors 0.0245 and 0.0246, respectively.

If ultrasonography was treated as the gold standard, the estimated $\hat{S}_d$ and $\hat{C}_d$ using all the DU and DV tables were 0.797 and 0.618, respectively. The standard errors were 0.024 and 0.024, respectively.

4.3.2 Simulations of the model with known sensitivity and specificity of the silver standard

In this section, simulations were conducted to examine the performance of the estimates. Two scenarios were considered. In the first scenario, only one DU table and one DV table were generated in each iteration. Estimated sensitivity and specificity of d-dimer were obtained at the end of each simulation. In the second scenario, 3 DU tables and 4 DV tables were generated and estimates were obtained at the end of each simulation. The table total was set at 1000 for all simulations. 5000 simulations were performed for each model in each scenario.

Based on the observed tables of ultrasonography, the known sensitivity and specificity of ultrasonography were chosen at 0.81 and 0.77, respectively. Besides, 4 sets of

sensitivity and specificity of d-dimer were used in the simulation as true parameter values, which were extracted from Table 4.1.

S_d	C_d
0.59	0.90
0.816	0.571
0.76	0.87
0.865	0.656

Table 4.4: Parameter values of sensitivity and specificity of d-dimer for simulations in the fixed effects model

Simulations using one DU table and one DV table

In the first scenario, only one DU table and one DV table were generated at each simulation. Two models were fitted: admitting the imperfection of ultrasonography and ignoring the difference between ultrasonography and venography. The data were generated using estimated probabilities from the model admitting the imperfection of ultrasonography. Estimates were obtained from each model at the end of each simulation. Bias with standard errors, mean squared error, and average coverage probability of 95% confidence intervals were obtained for each model.

Before providing details of the simulation procedure, elaboration of generating cell probabilities is presented below. First of all, model coefficients should be expressed as functions of the known parameters S_d, C_d, S_u, C_u, and prevalence of disease. In Chapter 3, the associations between these quantities were provided and summarized below.

$$S_d = \frac{e^{\beta_1+\beta_4}}{1+e^{\beta_1+\beta_4}}$$

$$C_d = \frac{1}{1+e^{\beta_1}}$$

$$S_u = \frac{e^{\beta_2+\beta_5}}{1+e^{\beta_2+\beta_5}}$$

$$C_u = \frac{1}{1+e^{\beta_2}}$$

By simple algebra, these equations can be solved for the model coefficients β_1, β_2, β_4, and β_5, as shown below.

$$\beta_1 = log\frac{1-C_d}{C_d}$$

$$\beta_2 = log\frac{1-C_u}{C_u}$$

$$\beta_4 = log\frac{S_d}{1-S_d} - log\frac{1-C_d}{C_d}$$

$$\beta_5 = log\frac{S_u}{1-S_u} - log\frac{1-C_u}{C_u}$$

The expression for β_3 was derived by the following steps. First of all, the prevalence was expressed as the sum of cell probabilities at venography=1.

$prevalence = P(V^+) = p_{..1} = p_{111} + p_{011} + p_{101} + p_{001}$

Each of the above 4 components on the right-hand side can be expressed as the function of model coefficients.

$$p_{111} = \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{011} = \frac{e^{\beta_2+\beta_3+\beta_5}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{101} = \frac{e^{\beta_1+\beta_3+\beta_4}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

$$p_{001} = \frac{e^{\beta_3}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

The prevalence can then be expressed as the sum of these four fractions.

$$P(V^+) = \frac{e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_3}}{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}+e^{\beta_3}+e^{\beta_1+\beta_3+\beta_4}+e^{\beta_2+\beta_3+\beta_5}+e^{\beta_1+\beta_2+\beta_3+\beta_4+\beta_5}}$$

By some algebra, β_3 can be calculated by the following function.

$$\beta_3 = log\frac{P(V^+)}{1-P(V^+)} - log\frac{1+e^{\beta_1}+e^{\beta_2}+e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_4}+e^{\beta_2+\beta_5}+e^{\beta_1+\beta_2+\beta_4+\beta_5}}$$

Substituting representations of β_1, β_2, β_4, β_5, and the prevalence into the function above gave rise to the expression of β_3.

Using the expressions of all these coefficients, the true cell probabilities can be derived. The tables can then be generated by the cell probabilities and table total. The simulation procedure is summarized below.

1. Calculate parameter values of cell probabilities based on known sensitivity and specificity of ultrasonography and d-dimer as well as disease prevalence set at 0.7.

2. Generate a three-way contingency table based on cell probabilities in step 1 and table total 1000.

3. Obtain the marginal DU table by summing cell counts from step 2 over the index of venography.

4. Generate a new three-way contingency table using cell probabilities in step 1.

5. Obtain the marginal DV table by summing cell counts from step 4 over the index of ultrasonography.

6. Fit the model admitting that ultrasonography is an imperfect reference to the tables generated in step 3 and 5. Estimate sensitivity and specificity of d-dimer and corresponding 95% confidence intervals were obtained.

7. Fit the model ignoring the difference between ultrasonography and venography to the tables generated in step 3 and 5. Estimates of sensitivity and specificity of d-dimer and corresponding 95% confidence intervals were obtained.

8. Repeat steps 2 - 7 5000 times. For each model, the bias and mean squared error of estimated sensitivity and specificity of d-dimer were calculated. 95% coverages of the 95% confidence interval of sensitivity and specificity of d-dimer were calculated for each model.

Settings	$S_d(0.59)$ $C_d(0.90)$ $S_u(0.81)$ $C_u(0.77)$	$S_d(0.816)$ $C_d(0.571)$ $S_u(0.81)$ $C_u(0.77)$
Bias S_{d1} (s.e.)	-0.0253 (0.0136)	-0.0199 (0.0109)
Bias C_{d1} (s.e.)	-0.0979 (0.0155)	-0.0775 (0.0193)
MSE S_{d1}	0.000827	0.000514
MSE C_{d1}	0.00983	0.00638
95% coverage S_{d1}	0.535	0.568
95% coverage C_{d1}	0	0.0228
Bias S_{d2} (s.e.)	-7.57e-05 (0.0143)	0.000172 (0.0115)
Bias C_{d2} (s.e.)	0.000223 (0.0159)	0.000181 (0.0237)
MSE S_{d2}	0.000204	0.000132
MSE C_{d2}	0.000254	0.000561
95% coverage S_{d2}	0.949	0.952
95% coverage C_{d2}	0.945	0.953

Table 4.5: S_{d1} and C_{d1}: sensitivity and specificity of d-dimer from the unadjusted fixed effects model using one table of each type; S_{d2} and C_{d2}: sensitivity and specificity of d-dimer from the adjusted fixed effects model using one table of each type.

Table 4.5 summarizes the results from simulations using the first two sets of sensitivity and specificity of d-dimer in Table 4.4. Table 4.6 displayed the results from simulations when the true sensitivity and specificity of d-dimer took the last two rows of values in Table 4.4.

In both tables, biases of S_d and C_d from the model ignoring the imperfection of ultrasonography were much larger than those from the model adjusting for the difference between ultrasonography and venography. Estimates of S_d and C_d from the adjusted model were almost unbiased. The corresponding coverages of S_d and C_d were very close to 95%. The coverages of S_d and C_d in the unadjusted model,

Settings	$S_d(0.76)$ $C_d(0.87)$ $S_u(0.81)$ $C_u(0.77)$	$S_d(0.865)$ $C_d(0.656)$ $S_u(0.81)$ $C_u(0.77)$
Bias S_{d1} (s.e.)	-0.0321 (0.0121)	-0.0270 (0.00990)
Bias C_{d1} (s.e.)	-0.126 (0.0168)	-0.104 (0.0192)
MSE S_{d1}	0.00118	0.000824
MSE S_{d1}	0.0161	0.0112
95% coverage S_{d1}	0.242	0.229
95% coverage C_{d1}	0	2e-04
Bias S_{d2} (s.e.)	0.000399 (0.0127)	-8.74e-06 (0.0104)
Bias C_{d2} (s.e.)	0.000412 (0.0179)	0.000441 (0.0233)
MSE S_{d1}	0.000162	0.000108
MSE S_{d1}	0.000322	0.000542
95% coverage S_{d2}	0.949	0.955
95% coverage C_{d2}	0.944	0.948

Table 4.6: S_{d1} and C_{d1}: sensitivity and specificity of d-dimer from the unadjusted fixed effects model using one table of each type; S_{d2} and C_{d2}: sensitivity and specificity of d-dimer from the adjusted fixed effects model using one table of each type.

however, were extremely low, especially in specificity. This was due to the large bias and the relatively small standard error in estimating specificity in the unadjusted model.

Figures 4.1, 4.2 and 4.3 displays the sampling distributions from the unadjusted and adjusted models when the true sensitivity and specificity of d-dimer were 0.76 and 0.87, respectively. The vertical lines in all the histograms denoted the true parameter values. The true sensitivity fell at the right tail end of the sampling distribution of estimates from the unadjusted model. The distance between the center of the sampling distribution and the true sensitivity was the absolute bias 0.032, which was larger than twice the standard error 0.024. Twice the standard error was approximately half the width of the 95% confidence intervals. In other words, the absolute bias was larger than half the width of the 95% confidence intervals. Consequently, only a small portion of the confidence intervals on the right-hand side

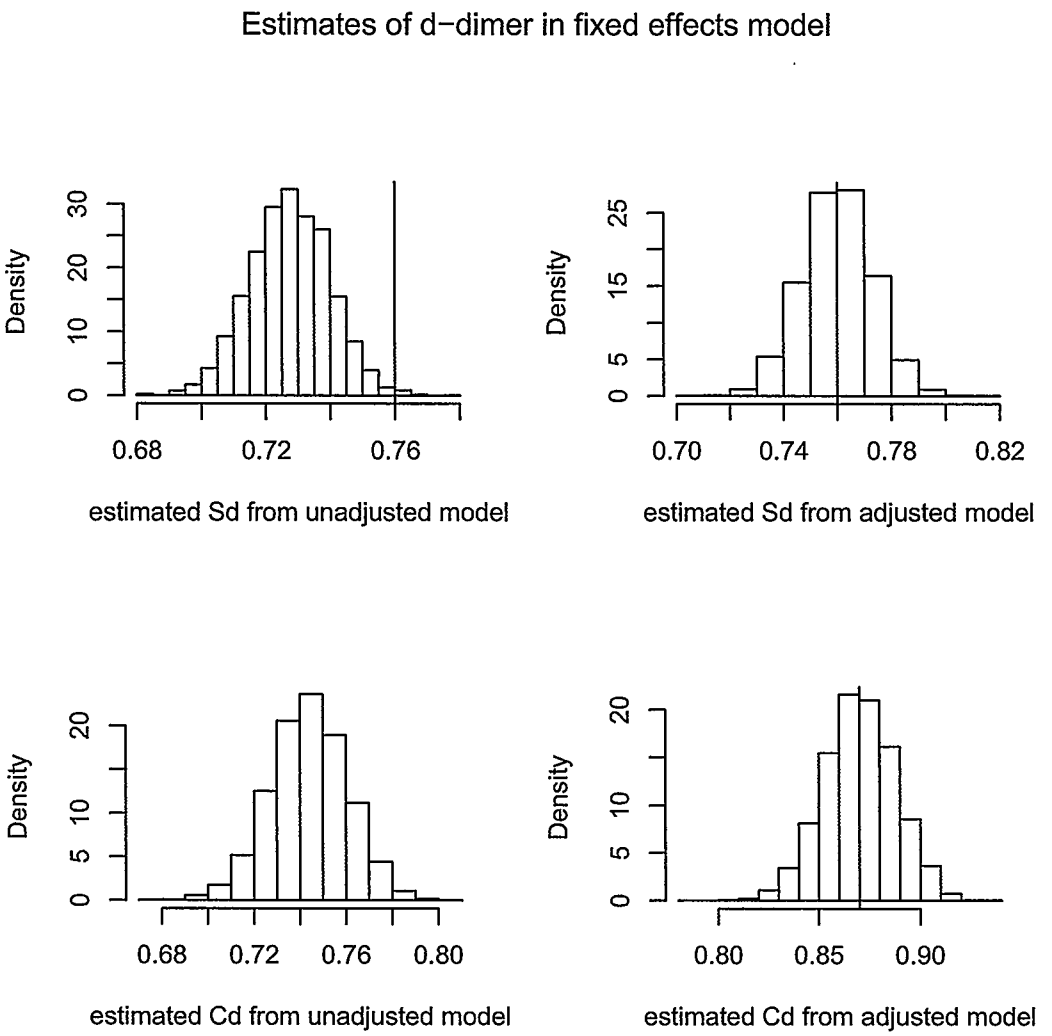

Figure 4.1: Histograms of estimated sensitivity and specificity from unadjusted and adjusted fixed effects model using one table of each type

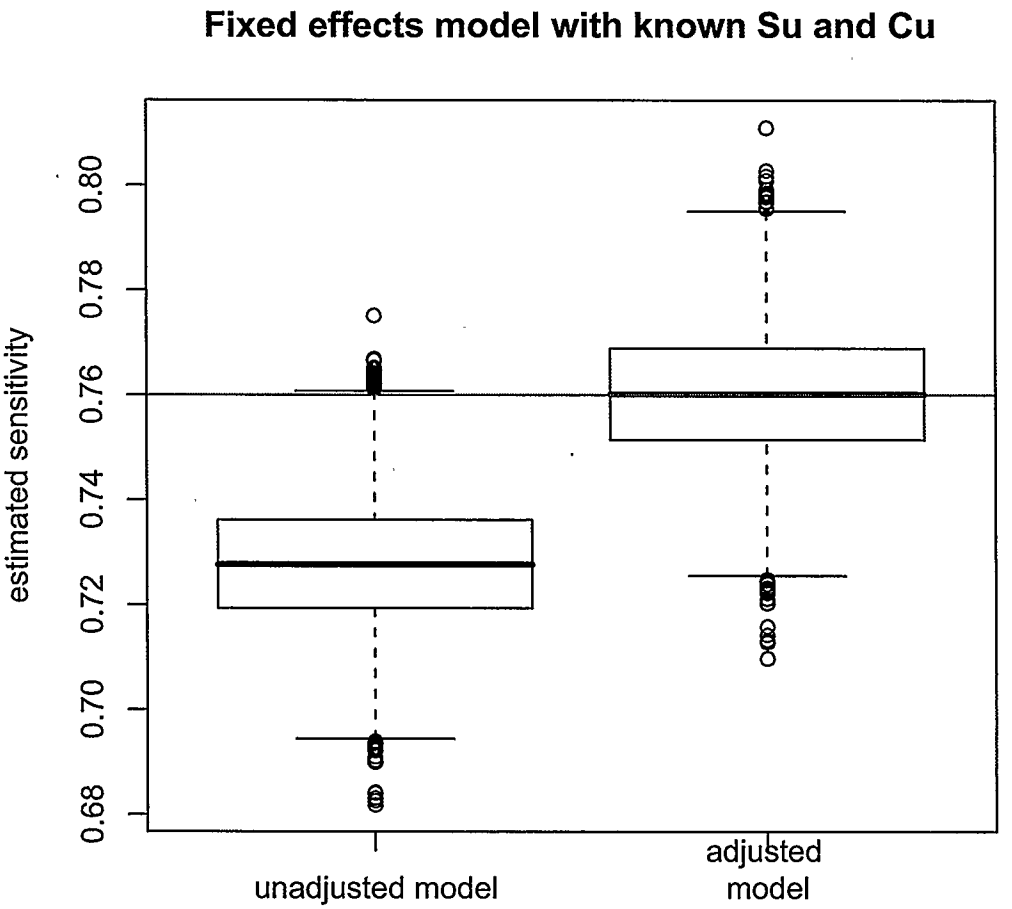

Figure 4.2: Boxplots of estimated sensitivity from unadjusted and adjusted fixed effects model using one table of each type. True value of sensitivity is 0.76.

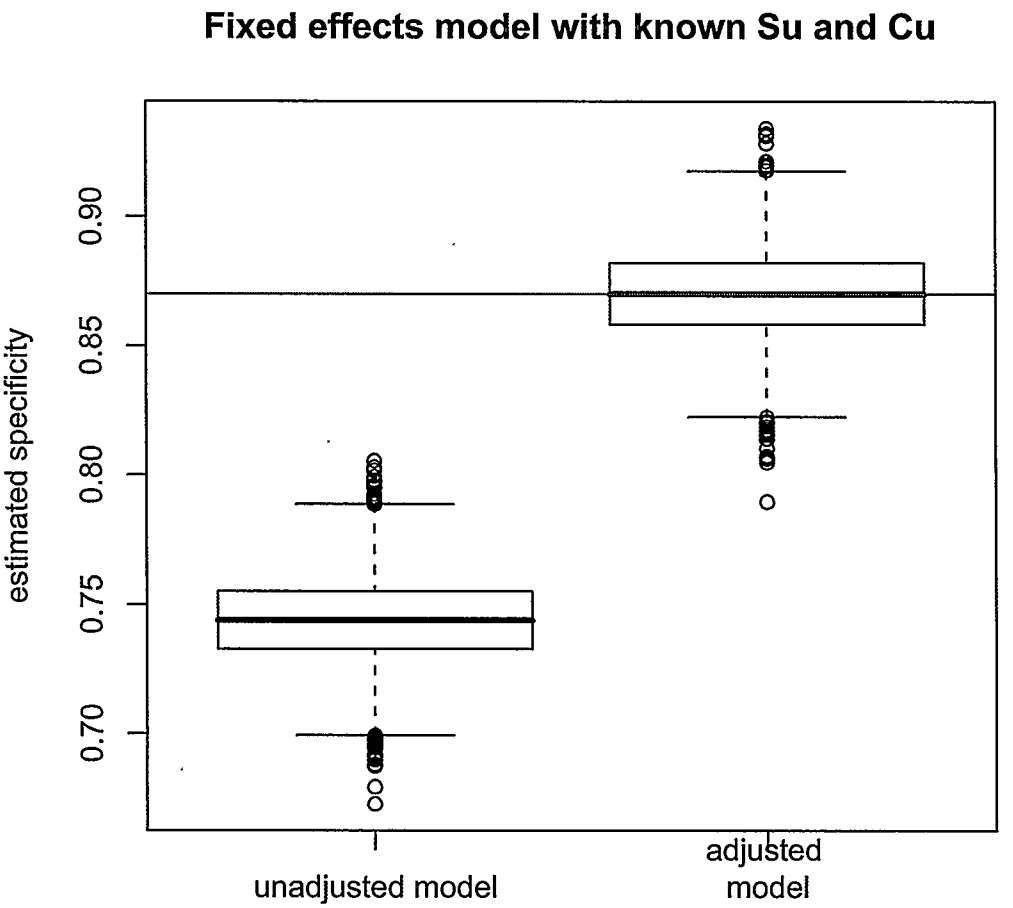

Figure 4.3: Boxplots of estimated specificity from unadjusted and adjusted fixed effects model using one table of each type. True value of specificity is 0.87.

of the sampling distribution covered the true parameter values. Specifically, the confidence intervals of estimated sensitivities larger than 0.76-0.024=0.736 covered the true sensitivity (0.76), where 0.736 was approximately the 75 percentile of the sampling distribution. Therefore, the coverage of sensitivity was approximately 0.25. The magnitude of bias was substantially large in specificity from the unadjusted model, almost 10 times the standard error. Therefore, the coverage of true specificity was zero. In contrast, the true sensitivity and specificity were located at the center of the sampling distribution of estimates from the adjusted model. The coverage of true parameter values was very close to 95% in this adjusted model setting.

In the analysis from the previous section, sensitivity and specificity of ultrasonography were very high, 0.95. The resulting estimated sensitivity and specificity values of d-dimer from the two models were very close to each other. In the simulation in this section, however, the sensitivity and specificity of ultrasonography were not as high as 0.95. Results from simulations indicated that bias would be substantial with reference test that was even moderately different from the gold standard.

Simulations using multiple DU and DV tables

In this scenario, multiple tables of each type of DU and DV were generated at each iteration. Parameter values of model coefficients were calculated following the same procedure as that in the previous simulations. The procedure of simulation was summarized below.

1. Calculate parameter values of cell probabilities based on known sensitivity and specificity of ultrasonography and d-dimer as well as fixed disease prevalence of 0.7.

2. Generate a three-way contingency table using cell probabilities in step 1.

3. Calculate a marginal DU table by summing cell counts from step 2 over the index of venography.

4. Repeat steps 2 and 3 three times to generate three independent DU tables.

5. Generate a three-way contingency table using cell probabilities in step 1.

6. Calculate a marginal DV table by summing cell counts from step 2 over the index of ultrasonography.

7. Repeat steps 5 and 6 four times to generate four independent DV tables.

8. Fit the model acknowledging the imperfection of ultrasonography to the tables in step 4 and 7. Estimate sensitivity and specificity of d-dimer and corresponding 95% confidence intervals.

9. Fit the model ignoring the imperfection of ultrasonography to the tables in step 4 and 7. Estimate sensitivity and specificity of d-dimer and corresponding 95% confidence intervals.

10. Repeat steps 2 - 9 5000 times. For each model, calculate the bias and mean squared error of estimated sensitivity and specificity of d-dimer, 95% coverage of the 95% confidence intervals.

Results from simulations are summarized in Table 4.7 and Table 4.8. Estimates from the adjusted approach were almost unbiased with 95% coverage of sensitivity and specificity of d-dimer. In the analysis of multiple DU and DV tables, biases of S_d

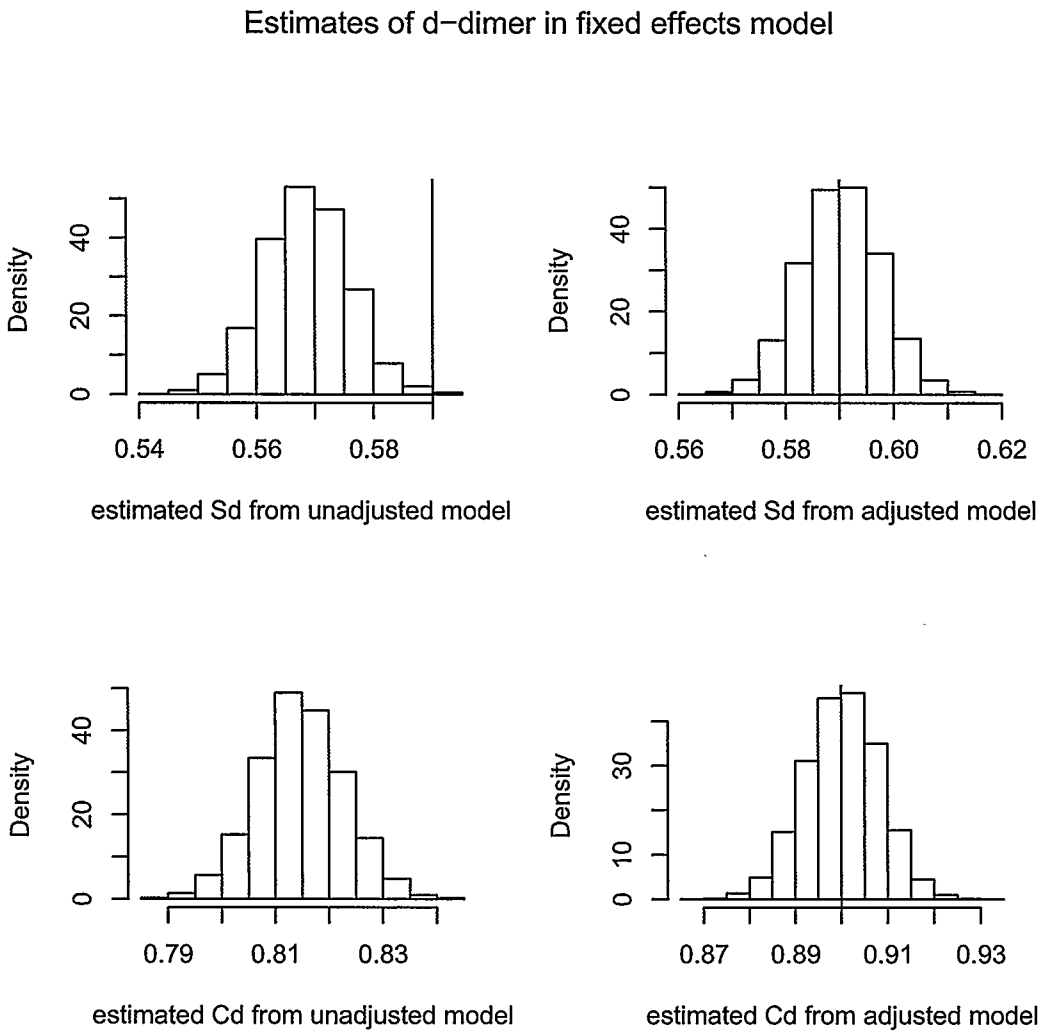

Figure 4.4: Histograms of estimated sensitivity and specificity from unadjusted and adjusted fixed effects model using multiple tables with known sensitivity and specificity of the silver standard

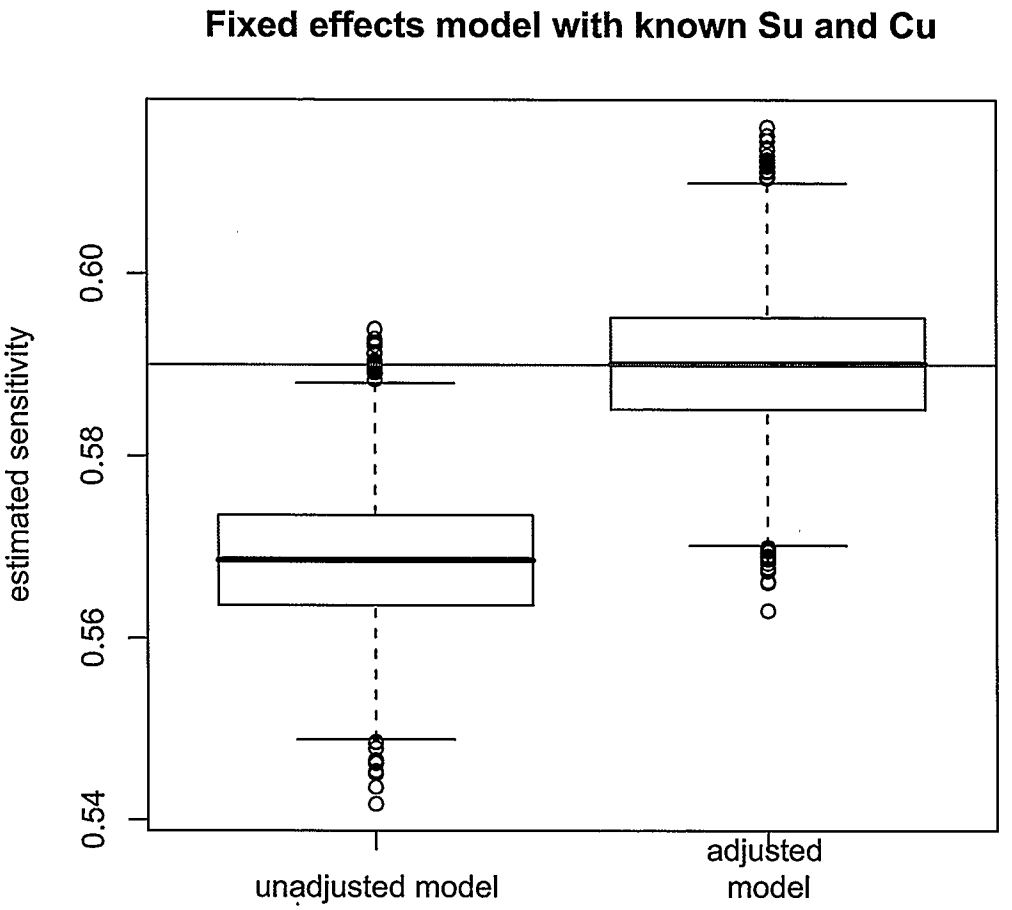

Figure 4.5: Boxplots of estimated specificity from unadjusted and adjusted fixed effects model using multiple tables with known sensitivity and specificity of the silver standard

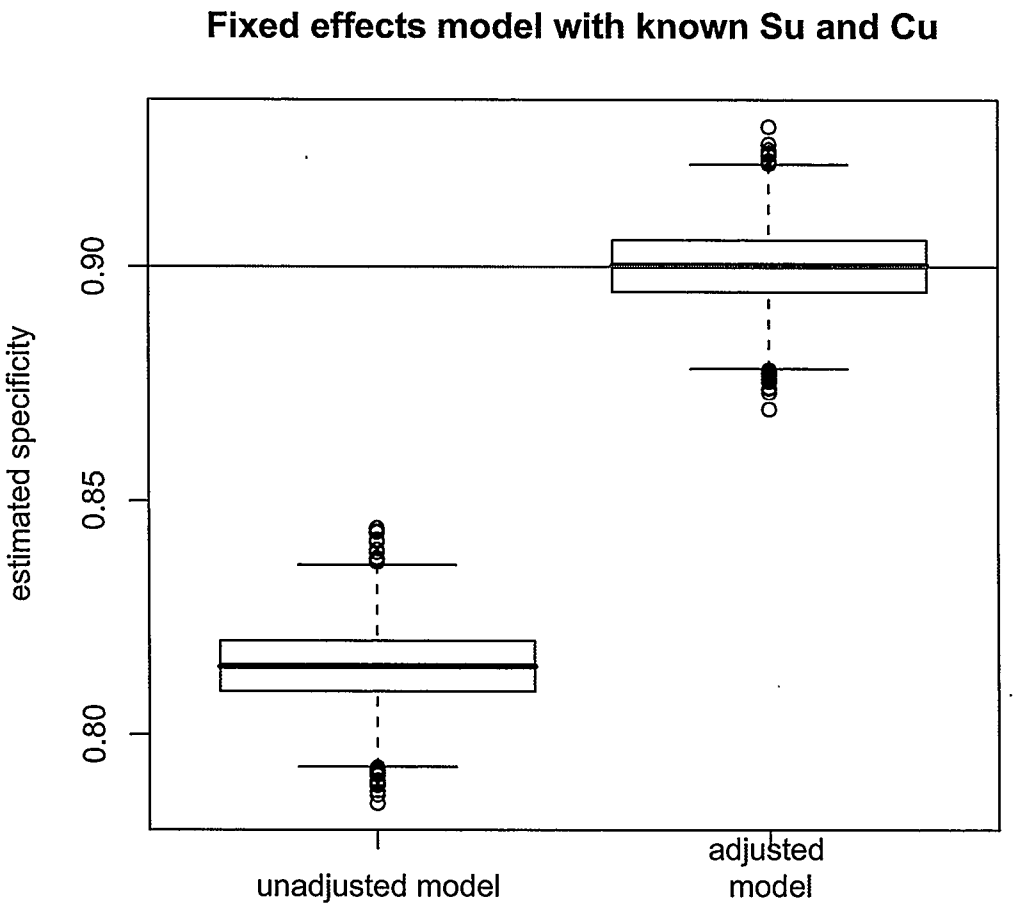

Figure 4.6: Boxplots of estimated specificity from unadjusted and adjusted fixed effects model using multiple tables with known sensitivity and specificity of the silver standard

Settings	$S_d(0.59)$ $C_d(0.90)$ $S_u(0.81)$ $C_u(0.77)$	$S_d(0.816)$ $C_d(0.571)$ $S_u(0.81)$ $C_u(0.77)$
Bias S_{d1} (s.e.)	-0.0215 (0.00720)	-0.0171 (0.00574)
Bias C_{d1} (s.e.)	-0.0854 (0.00807)	-0.0675 (0.0102)
MSE S_{d1}	0.000513	0.000324
MSE C_{d1}	0.00735	0.00466
95% coverage S_{d1}	0.152	0.156
95% coverage C_{d1}	0	0
Bias S_{d2} (s.e.)	0.000102 (0.00741)	-4.24e-05 (0.00600)
Bias C_{d2} (s.e.)	9.58e-05 (0.00801)	-4.59e-05 (0.0120)
MSE	5.50e-05	3.60e-05
MSE	6.42e-05	0.000144
95% coverage S_{d2}	0.953	0.949
95% coverage C_{d2}	0.954	0.954

Table 4.7: S_{d1} and C_{d1}: sensitivity and specificity of d-dimer from the unadjusted fixed effects model using multiple tables; S_{d2} and C_{d2}: sensitivity and specificity of d-dimer from the adjusted fixed effects model using multiple tables

and C_d estimates from the unadjusted model were slightly reduced compared to those when only one table from each type was analyzed. But the standard errors of the bias were substantially reduced when multiple tables were analyzed. The reduction in standard error outweighed that in the bias. Consequently, the coverage of sensitivity in the simulation with multiple tables was lower than that in the simulation with only one DU table using the unadjusted model.

The sampling distributions of the unadjusted and adjusted models were presented in Figures 4.4, 4.5, and 4.6. The estimates from the unadjusted model were centered away from the true parameter values, whereas those from the adjusted model were centered on the true parameter values. Similar to the discussion in the situation with one table, the large biases relative to the standard errors provided a plausible explanation of why the 95% coverage probability from the unadjusted model was very small.

Settings	$S_d(0.76)$ $C_d(0.87)$ $S_u(0.81)$ $C_u(0.77)$	$S_d(0.865)$ $C_d(0.656)$ $S_u(0.81)$ $C_u(0.77)$
Bias S_{d1} (s.e.)	-0.0278 (0.00640)	-0.0229 (0.00527)
Bias C_{d1} (s.e.)	-0.110 (0.00880)	-0.0907 (0.0103)
MSE S_{d1}	0.000812	0.000550
MSE C_{d1}	0.0121	0.00833
95% coverage S_{d1}	0.0072	0.007
95% coverage C_{d1}	0	0
Bias S_{d2} (s.e.)	-8.06e-05 (0.00664)	6.64e-05 (0.00545)
Bias C_{d2} (s.e.)	-3.39e-05 (0.00888)	-6.06e-06 (0.0119)
MSE S_{d1}	4.42e-05	2.97e-05
MSE C_{d1}	7.89e-05	0.000142
95% coverage S_{d2}	0.950	0.951
95% coverage C_{d2}	0.954	0.952

Table 4.8: S_{d1} and C_{d1}: sensitivity and specificity of d-dimer from the unadjusted fixed effects model using multiple DU and DV tables; S_{d2} and C_{d2}: sensitivity and specificity of d-dimer from the adjusted fixed effects model using multiple DU and DV tables

4.3.3 The effect of disease prevalence on the bias in the unadjusted model

In the results above, the biases in estimating sensitivity and specificity of d-dimer from the unadjusted model were much larger than those from the model accounting for the difference in reference tests. To further investigate this bias, simulations were performed using different values of disease prevalence and diagnostic characteristics of ultrasonography. The prevalences of disease were chosen from 0.1 to 0.9 with the spacing of 0.2. Three sets of sensitivity and specificity of ultrasonography were applied. Two forms of data were considered: 1 table for each marginal type and multiple tables for each marginal type. For each form of data, 1000 simulations were performed with table total 1000.

At the end of the simulation, the magnitudes of biases in estimated sensitivity and specificity were plotted against the values of disease prevalence. Results from

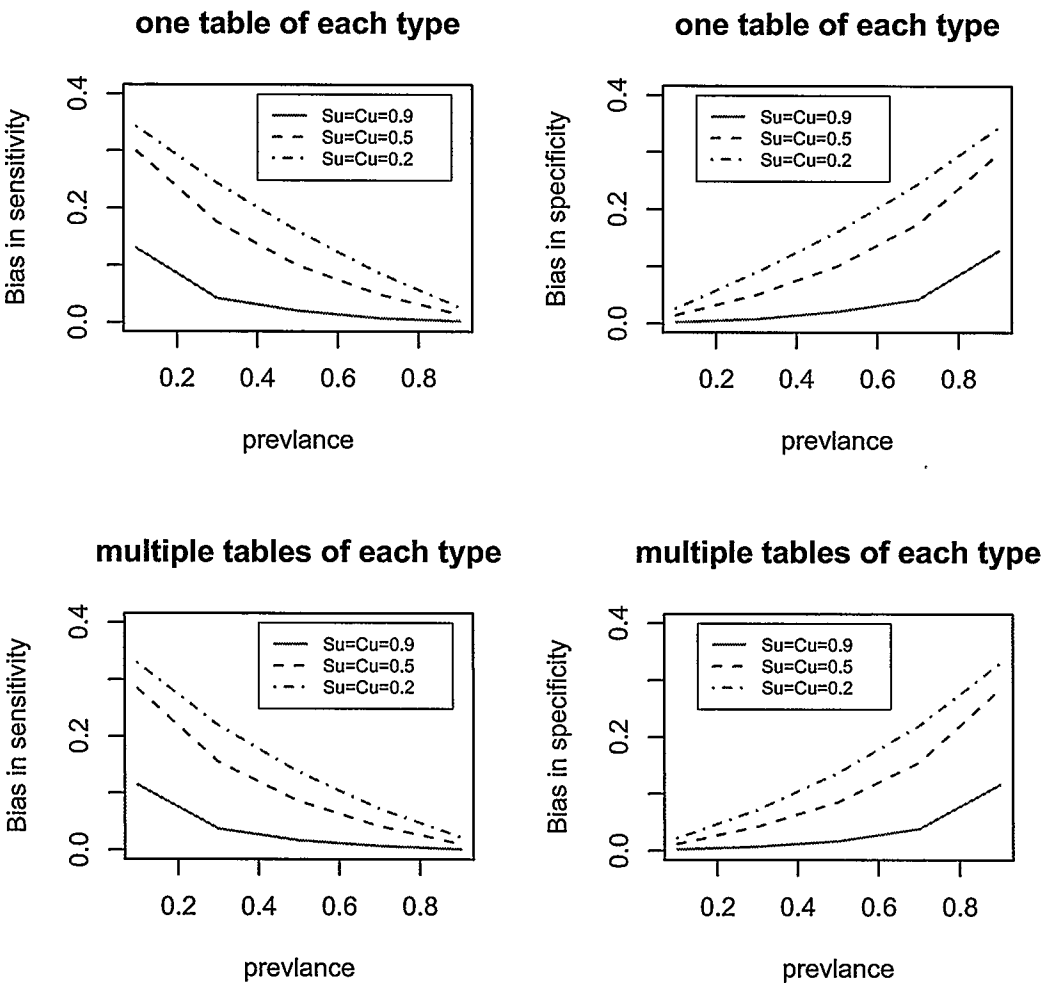

Figure 4.7: Effect of disease prevalence on the magnitude of bias in sensitivity and specificity using the unadjusted model when sensitivity and specificity of the silver standard are known

simulations were summarized and displayed in Figure 4.7. When there were 1 DU and 1 DV tables for analysis, the absolute bias in estimating sensitivity decreased as the prevalence of disease increased, regardless the choice of sensitivity and specificity of the silver standard. The absolute bias in estimating specificity, on the other hand, increased as the prevalence of disease increased, regardless the values of diagnostic characteristics of the silver standard. For the same value of disease prevalence, biases were smaller when the silver standard had high sensitivity and specificity than those when a poor reference standard was applied. The same phenomena were observed in the simulations with multiple tables.

4.3.4 Comparison between the analysis using all tables and the analysis using tables from the test and gold standard only

The above simulations provided evidence that the model adjusting for the imperfection of ultrasonography produced estimates with very small biases and had much higher efficiency than the model ignoring the difference between the two references. In this section, the adjusted model was compared to the model using the DV tables only. Both approaches were expected to produce estimates with similar biases. The number of tables was the same as that in the previous simulation. The same sets of parameter values of S_d, C_d, S_u, C_u and disease prevalence of 0.7 were applied.

Results in Table 4.9 and Table 4.10 indicate that coverages of 95% confidence interval were very close to 0.95 in both analyses. The standard errors and mean squared errors using all the tables were smaller than those from the analysis using DV tables only. The ratio of the mean squared errors from the analysis using both DU and DV tables over that from the analysis using DV tables only was smaller than

Settings	$S_d(0.59)$ $C_d(0.90)$ $S_u(0.81)$ $C_u(0.77)$	$S_d(0.816)$ $C_d(0.571)$ $S_u(0.81)$ $C_u(0.77)$
Bias S_{d1} (s.e.)	0.000101 (0.00932)	-0.000103 (0.00744)
Bias C_{d1} (s.e.)	-0.000112 (0.0085)	-0.000160 (0.0143)
MSE S_{d1}	8.68e-05	5.53e-05
MSE C_{d1}	7.22e-05	0.000205
95% coverage S_{d1}	0.952	0.941
95% coverage C_{d1}	0.952	0.948
Bias S_{d2} (s.e.)	-4.32e-05 (0.00744)	-0.000119 (0.00611)
Bias C_{d2} (s.e.)	-0.000121 (0.00797)	-0.000170 (0.0124)
MSE S_{d2}	5.53e-05	3.74e-05
MSE C_{d2}	6.35e-05	0.000153
95% coverage S_{d2}	0.951	0.946
95% coverage C_{d2}	0.952	0.946

Table 4.9: S_{d1} and C_{d1}: sensitivity and specificity of d-dimer from the fixed effects analysis using DV tables only; S_{d2} and C_{d2}: sensitivity and specificity of d-dimer from the fixed effects analysis using all DU and DV tables

1. This indicated that the model using both DU and DV tables was more efficient than the model using DV tables only. Using studies from the gold standard alone and excluding studies using the imperfect reference in the meta-analysis resulted in loss of efficiency.

In summary, the estimators from the model adjusting for the difference between the two references and using all tables had very small biases. Incorporating the data from d-dimer and the imperfect reference provided more information on the diagnostic characteristics of d-dimer and, hence, smaller mean squared errors than using tables from the gold standard alone.

4.3.5 Analysis of three types of marginal tables

The model above dealt with situations where DU and DV tables were available but not the UV table. The clinical performance of the silver standard, instead, was

Settings	$S_d(0.76)$ $C_d(0.87)$ $S_u(0.81)$ $C_u(0.77)$	$S_d(0.865)$ $C_d(0.656)$ $S_u(0.81)$ $C_u(0.77)$
Bias S_{d1} (s.e.)	-0.000198 (0.0080)	0.000120 (0.00639)
Bias C_{d1} (s.e.)	-8.36e-05 (0.0097)	-0.000214 (0.0136)
MSE S_{d1}	6.46e-05	4.08e-05
MSE C_{d1}	9.45e-05	0.000186
95% coverage S_{d1}	0.950	0.951
95% coverage C_{d1}	0.951	0.952
Bias S_{d2} (s.e.)	-0.000149 (0.00663)	0.000124 (0.00547)
Bias C_{d2} (c.e.)	-7.06e-05 (0.0091)	-9.71e-05 (0.0120)
MSE S_{d2}	4.39e-05	3.00e-05
MSE C_{d2}	8.29e-05	0.000143
95% coverage S_{d2}	0.950	0.948
95% coverage C_{d2}	0.951	0.951

Table 4.10: S_{d1} and C_{d1}: sensitivity and specificity of d-dimer from the fixed effects analysis using DV tables only; S_{d2} and C_{d2}: sensitivity and specificity of d-dimer from the fixed effects analysis using all DU and DV tables

known. In a general situation of meta-analysis of diagnostic tests, the UV tables were available and all three types of marginal tables were collected. The log likelihood of the i^{th} DU table, for example, can be written as the following.

$$\log(L^i_{du}) = table^i_{du} \times \log(p_{du.})$$

In this expression, $table^i_{du}$ was the i^{th} observed table of d-dimer and ultrasonography. The marginal probabilities $p_{du.}$ were sum of p_{duv} over levels of venography, taking into account the constraint of all probabilities summing to 1. Expressions of p_{duv} in terms of model coefficients were given in Chapter 3. Similar functional forms for the log likelihoods of the j^{th} UV table and the k^{th} DV table can be derived. The log of joint likelihood function to be maximized was then calculated as the following.

$$\text{logL} = \sum_i \log(L^i_{du}) + \sum_j \log(L^j_{uv}) + \sum_k \log(L^k_{dv})$$

Maximization of the joint likelihood was implemented via the nlm() function in R. The nlm() function applies the Newton-type procedure to locate the minimum of a function. The negative log likelihood was specified in the nlm() and a random sample of six uniformly distributed values was given as the starting values of the estimation. The nlm() function produced the second derivatives of the likelihood with respect to the parameters β. The inverse of the negative of second derivatives provided an asymptotic estimated variance covariance matrix of $\hat{\beta}$.

In this analysis, all the tables from Table 4.1, 4.2, and 4.3 were used. The estimated cell probabilities were listed below.

$\hat{p}_{000}$	$\hat{p}_{100}$	$\hat{p}_{010}$	$\hat{p}_{110}$	$\hat{p}_{001}$	$\hat{p}_{101}$	$\hat{p}_{011}$	$\hat{p}_{111}$
0.334	0.163	0.0258	0.0126	0.018	0.0825	0.0652	0.299

The estimated sensitivity and specificity of d-dimer were given as the following.

$$\hat{S}_d = 0.821, \hat{C}_d = 0.672$$

By the delta method, the estimate variances for S_d and C_d were obtained.

$\hat{V}_{S_d}$	$\hat{V}_{C_d}$
0.000600	0.000818

The standard errors of $\hat{S}_d$ and $\hat{C}_d$ were 0.0245 and 0.0286, respectively.

4.3.6 Simulations on the model using three types of marginal tables

Simulations were conducted using estimated coefficients in the above analysis as parameter values. One marginal table of each type was obtained by generating a multinomial three-dimensional contingency table and summing over levels of the third factor. The estimated cell probabilities were listed below.

$\hat{p_{000}}$	$\hat{p_{100}}$	$\hat{p_{010}}$	$\hat{p_{110}}$	$\hat{p_{001}}$	$\hat{p_{101}}$	$\hat{p_{011}}$	$\hat{p_{111}}$
0.334	0.163	0.0258	0.0126	0.018	0.0825	0.0652	0.299

Data from the d-dimer paper [68] consisted of 3 DU tables and 4 DV tables. In the review of ultrasonography [6], 5 UV tables were available. In the simulations, the same number of marginal tables of each type was generated. In other words, 3 DU tables, 4 DV tables, and 5 UV tables were generated at each iteration. The simulation procedure can be summarized as below.

1. Generate one multinomial table from Multinom(1000, p_{duv}).
2. Obtain a marginal table from step 1.
3. Repeat steps 1 and 2 to generate marginal tables for each type: 3 DU, 4 DV, and 5 UV tables.
4. Fit the fixed effects model into the (3+4+5)=12 tables and acquire estimates of β, $\hat{\beta}$, and its variance-covariance matrix, $\hat{V}_\beta$.
5. Estimate sensitivity and specificity of d-dimer, $\hat{S}_d$ and $\hat{C}_d$, based on $\hat{\beta}$.
6. Estimate variances of sensitivity and specificity of d-dimer using the delta method.
7. Construct 95% confidence interval of $\hat{S}_d$ and $\hat{C}_d$ using the point estimates and corresponding estimated variances using these expressions:

$$\hat{S}_d \pm 1.96\sqrt{\hat{V}_{\hat{S}_d}} \text{ and } \hat{C}_d \pm 1.96\sqrt{\hat{V}_{\hat{C}_d}}$$

8. If the 95% confidence interval included the true sensitivity (specificity), count as 1 in the coverage; if not, count as 0.

9. Repeat steps 1-8 5000 times and calculate bias, mean square error (MSE), and the average coverage rate for sensitivity and specificity.

Parameters	True values	Mean of estimates	Bias (s.e.)	MSE	95% coverage
S_d	0.821	0.821	5.38e-05 (0.00756)	5.72e-05	0.945
C_d	0.672	0.672	1.37e-05 (0.00834)	6.95e-05	0.952

Table 4.11: Simulation on the fixed effects model using three types of marginal tables

Results in Table 4.11 indicated that the algorithm produced estimates of S_d and C_d with very small biases and small mean squared errors. The coverage rates of the estimates from simulations were very close to 95%.

Treating ultrasonography as the gold standard

If ultrasonography was treated the same as venography, the marginal tables reduced to 1 type, i.e., DV tables. Consequently, the DV tables for analysis combined DU and DV tables, i.e., 7 tables in total. The log-linear model can be written in the following form.

$$\log(m_{dv}) = \beta_0^* + \beta_1^* D + \beta_2^* V + \beta_3^* DV$$

The analysis produced point estimates of β^*s as the following.

β_1^*	β_2^*	β_3^*
-0.481	-1.494	1.85

The corresponding estimate variances were derived from the nlm() function in R.

$\hat{V}_{\beta_1^*}$	$\hat{V}_{\beta_2^*}$	$\hat{V}_{\beta_3^*}$
0.0103	0.0215	0.0323

The estimated sensitivity and specificity of d-dimer were: $\hat{S}_d^* = 0.797$ and $\hat{C}_d^* = 0.618$ with standard errors 0.024 and 0.024, respectively. Comparing these estimates with estimates from the 3-table analysis ($\hat{S}_d$=0.821 and $\hat{C}_d$=0.672) means that ignoring the fact that ultrasonography was not error-free underestimated the sensitivity and specificity of d-dimer. The magnitude of the difference was sizeable.

Simulations comparing the unadjusted and adjusted models

Simulations were conducted to compare the two methods: ultrasonography as imperfect reference (adjusted model) versus ultrasonography as gold standard (unadjusted model). In the simulation, analysis of the same set of marginal tables by the two methods was compared. The estimated cell probabilities from the adjusted model were used as true parameter values. In other words, the true sensitivity and specificity for simulations were 0.821 and 0.672. Bias, MSE, and 95% coverage of S_d and C_d were obtained for both models. The simulation process can be summarized below.

1. Generate a multinomial table from Multinomial(1000, $\hat{p}_{duv}$).

2. Repeat step 1 to generate marginal tables for each type: 3 DU, 4 DV, and 5 UV tables.

3. Fit the fixed effects model of imperfect ultrasonography using all the tables in step 2 and acquire estimates of sensitivity and specificity of d-dimer, S_d and C_d, and variances.

4. Construct 95% confidence intervals of $\hat{S}_d$ and $\hat{C}_d$ using the point estimates and their associated variances with the following expressions.

$$\hat{S}_d \pm 1.96\sqrt{\hat{V}_{\hat{S}_d}} \text{ and } \hat{C}_d \pm 1.96\sqrt{\hat{V}_{\hat{C}_d}}$$

5. If the 95% confidence interval included the true sensitivity(specificity), count as 1 in the coverage; if not, count as 0.

6. Fit the model where ultrasonography was the gold standard to the tables in step 2. The DU and DV marginal tables in step 2 were combined as one type of marginal tables. The UV tables were not used.

7. Obtain $\hat{S}_d$, $\hat{C}_d$ and corresponding variances.

8. Construct 95% confidence intervals of $\hat{S}_d$ and $\hat{C}_d$.

9. If the 95% confidence interval included the true sensitivity (specificity), count as 1 in the coverage; if not, count as 0.

10. Repeat steps 1-9 5000 times and calculate bias, mean square error, and the average coverage rates for sensitivity and specificity on each model.

Models	unadjusted model	adjusted model
Bias S_d (s.e.)	-0.0185 (0.00722)	-2.22e-5 (0.00753)
Bias C_d (s.e.)	-0.0378 (0.00772)	-6.42e-06 (0.00844)
MSE S_d	0.000396	5.66e-05
MSE C_d	0.00149	7.12e-05
95% coverage S_d	0.265	0.948
95% coverage C_d	0	0.944

Table 4.12: Comparison between the unadjusted and the adjusted fixed effects model on all three types of marginal tables

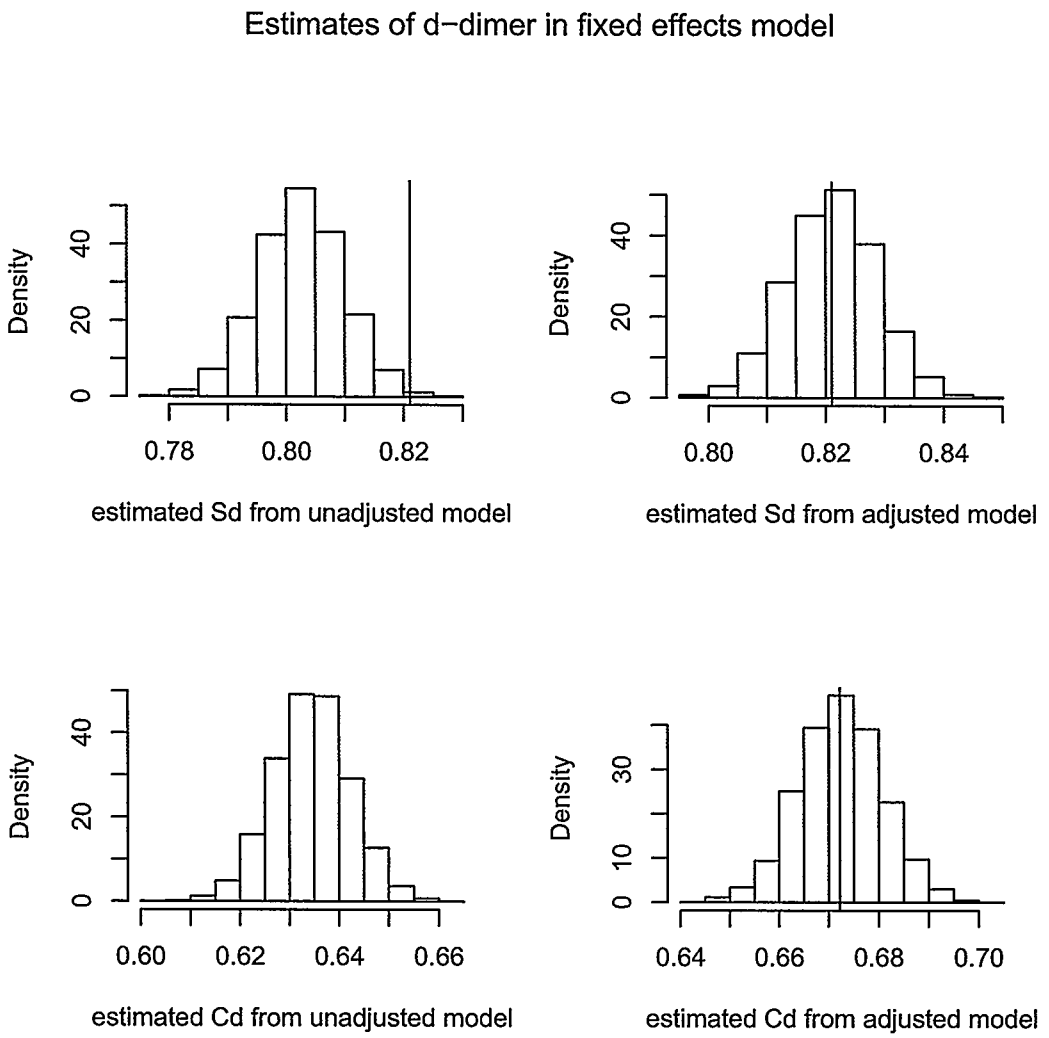

Figure 4.8: Histograms of estimated sensitivity and specificity from the unadjusted and the adjusted fixed effects model using all three types of tables

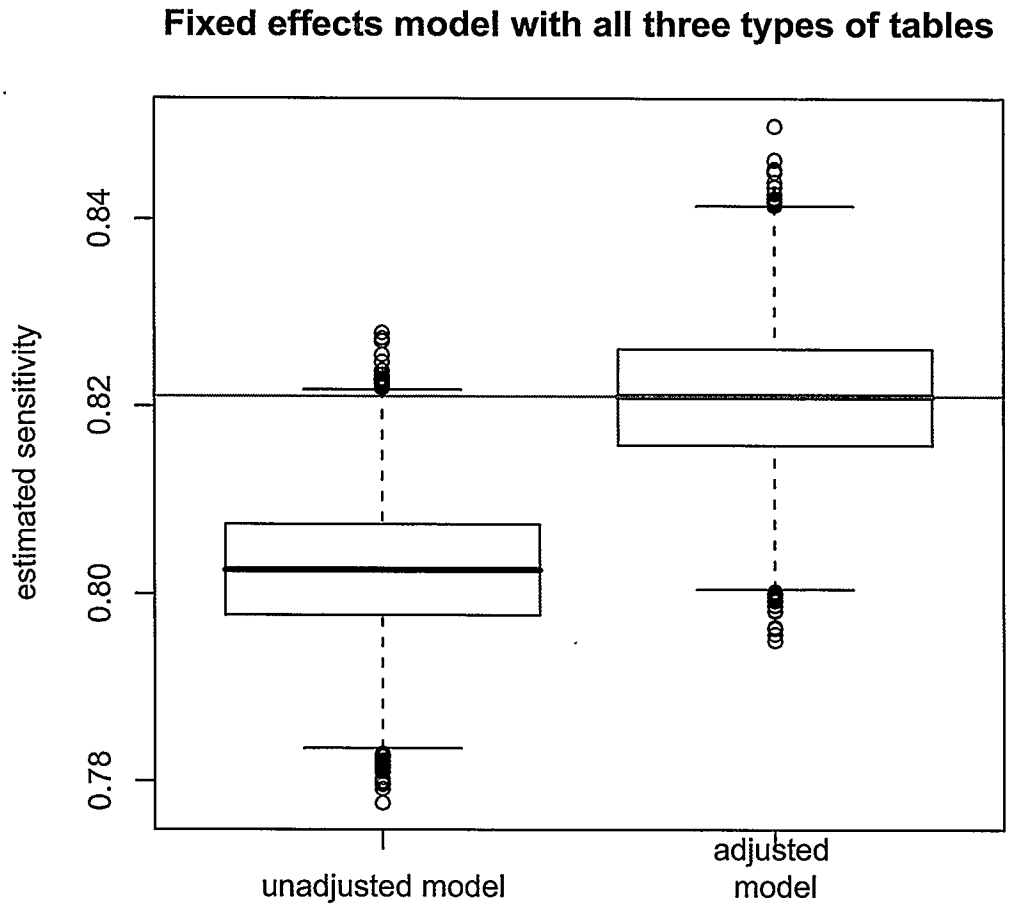

Figure 4.9: Boxplots of estimated sensitivity from the unadjusted and the adjusted fixed effects model using all three types of tables

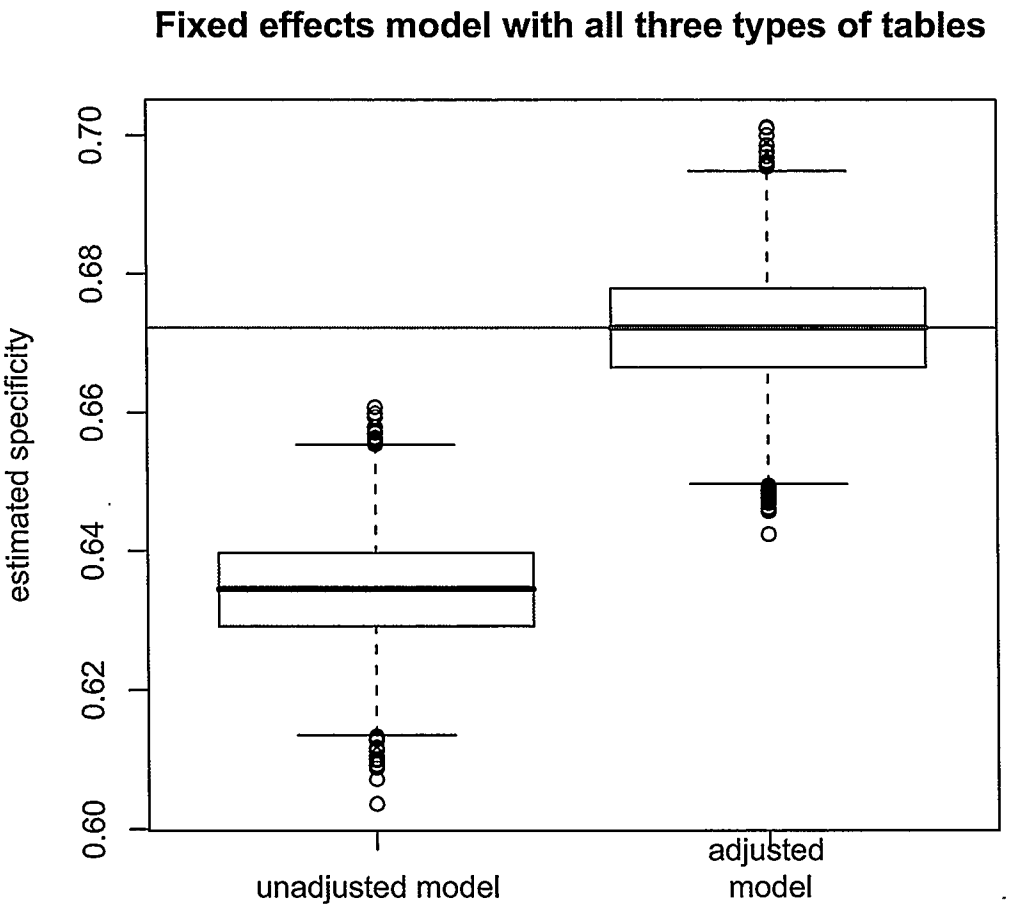

Figure 4.10: Boxplots of estimated specificity from the unadjusted and the adjusted fixed effects model using all three types of tables

Results from the two models were compared in Table 4.12. The biases from the unadjusted model were more than 10 times larger than the standard errors and the 95% coverage rates of S_d and C_d were extremely low, whereas those from the adjusted model were very close to 0.95. Figures 4.8, 4.9, and 4.10 displayed the sampling distributions of estimates from the two models. All the sampling distributions were symmetric. The estimated sensitivities from the unadjusted model were centered around 0.8, whereas those from the adjusted model were centered nicely on the parameter value. The shapes of sampling distributions from the two models were similar. With respect to estimating specificity, the estimates from the unadjusted model were centered on 0.63. In the adjusted model, estimates were centered nicely on the true parameter value. The shapes of the distributions from the two models were similar. With the small standard errors relative to the bias from the unadjusted model, the coverage of 95% confidence intervals was very low. This explanation is similar to those described in the previous section.

4.3.7 The effect of disease prevalence on the magnitude of bias in the unadjusted model

Similar to the analysis with known sensitivity and specificity of the silver standard, the unadjusted model ignoring the difference between the two references produced severely biased estimates. Simulations were performed to investigate the effect of disease prevalence on the magnitude of bias in the unadjusted model.

Figures 4.11 and 4.12 provided graphical representations of the absolute biases in estimating sensitivity and specificity against different disease prevalence. The three curves in each plot represented different diagnostic performances of ultrasonography.

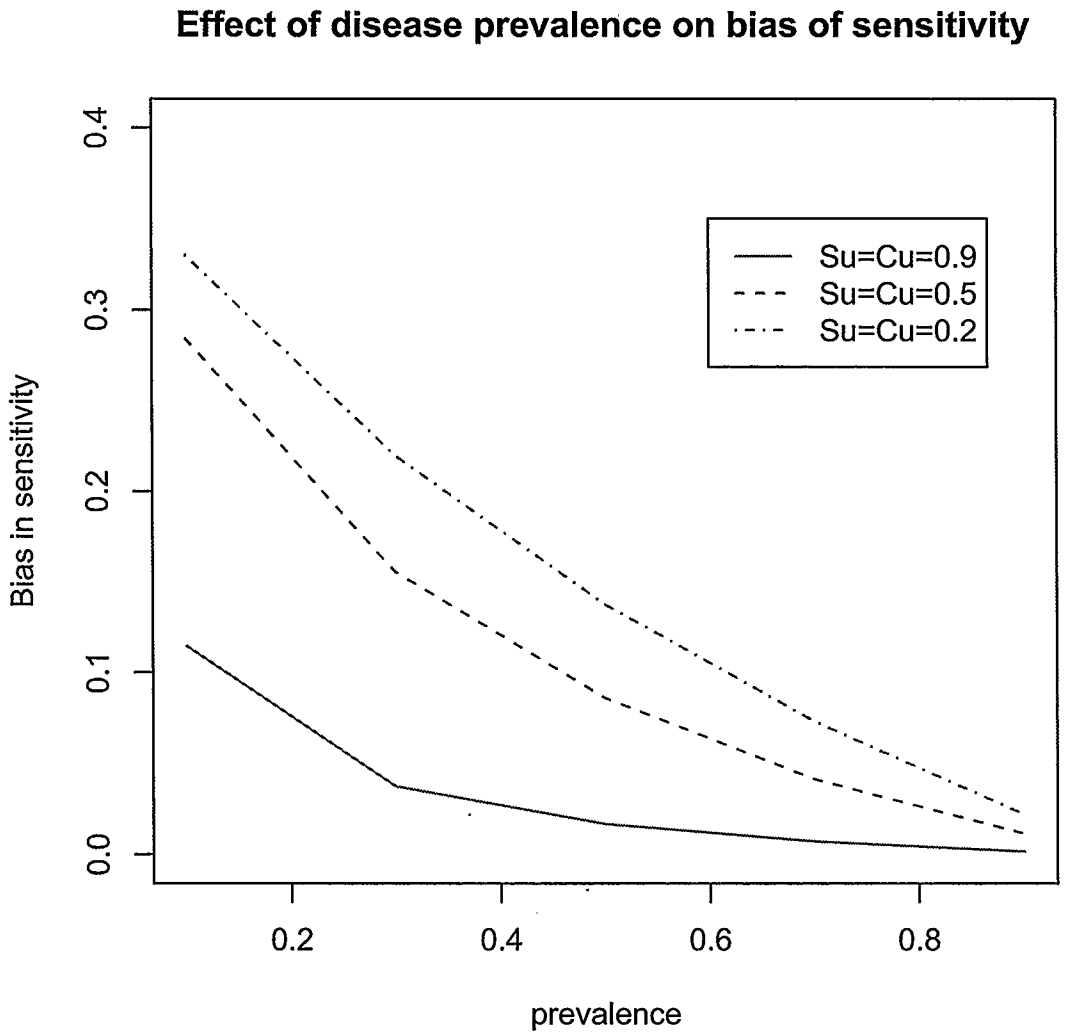

Figure 4.11: Effect of disease prevalence on the magnitude of bias in estimating sensitivity using the unadjusted fixed effects model on all three types of tables

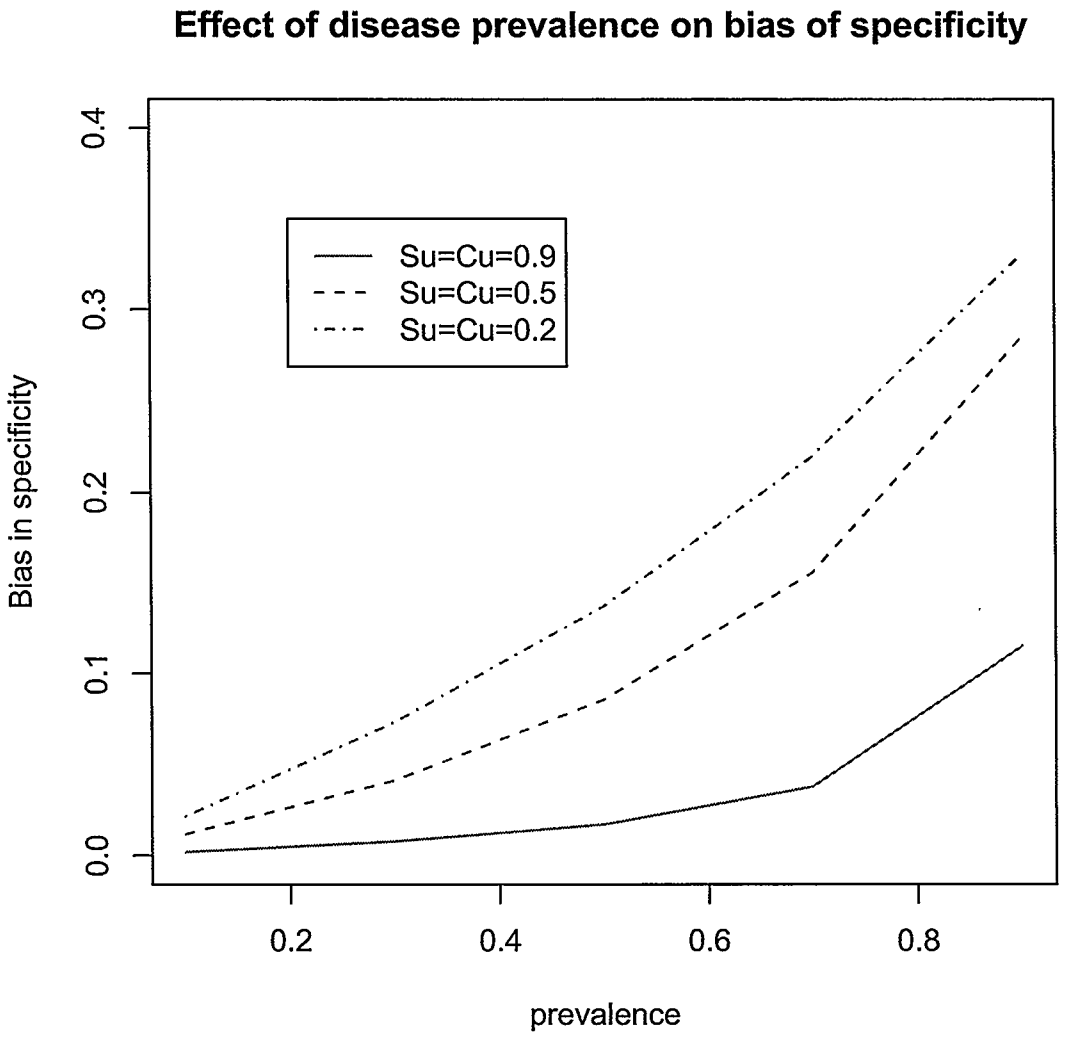

Figure 4.12: Effect of disease prevalence on the magnitude of bias in estimating specificity using the unadjusted fixed effects model on all three types of tables

The same conclusions can be arrived as those where the sensitivity and specificity of the silver standard were known. The bias in estimating sensitivity using the unadjusted model decreased as the prevalence of disease increased. The bias in specificity, in contrast, increased as the prevalence of disease increased. At the same value of disease prevalence, the biases in sensitivity and specificity were small if the silver standard had high diagnostic accuracy.

4.3.8 Comparison between the analysis using all tables and the analysis using tables from the test and gold standard only

In order to compare the efficiencies of the analysis using DV tables only and the analysis using all the tables, simulations were performed and mean squared errors from the two analyses were compared.

Analysis	DV tables only	all tables
Bias S_d (s.e.)	2.20e-05 (0.0089)	-2.67e-05 (0.0074)
Bias C_d (s.e.)	0.000203 (0.0103)	0.000117 (0.0085)
MSE S_d	7.92e-05	5.50e-05
MSE C_d	0.000106	7.25e-05
95% coverage S_d	0.948	0.953
95% coverage C_d	0.948	0.946

Table 4.13: Comparison between the fixed effects model using test versus gold standard only and the fixed effects model using all three types of tables

The two approaches produced similar bias. The standard errors from the analysis using all tables were slightly smaller than those from the analysis using DV tables only. Consequently, the mean square errors from the analysis using all tables were smaller than those from the analysis using DV tables only.

The relative efficiency of the analysis using DV tables only over that using all

tables was calculated as the ratio of mean squared error of the analysis using all tables over the mean squared error of the analysis using DV tables only. The relative efficiency of estimating sensitivity from the two models was 0.69. The same ratio in estimating specificity of d-dimer was 0.68. The analysis using all the tables in the fixed effects model was much more efficient than the analysis using the DV tables only. In other words, the meta-analysis using only tables from the gold standard resulted in loss of around 30% of the information. The additional DU tables provided diagnostic information of d-dimer as long as appropriate adjustments were taken into account. The analysis using all tables carried more information than the analysis using DV tables only.

4.4 Model accounting for the heterogeneity in disease prevalence across studies

4.4.1 Analysis of log-linear model with random disease prevalence only

As discussed in chapter 3, the disease prevalence varied from study to study. Assuming that venography was a perfect diagnostic tool of DVT, the prevalence of DVT was represented by the marginal probability of venography. In the log-linear model, it depended most strongly on the coefficient of venography. Taking into account difference in disease prevalence across studies, the random effects model can be expressed as the following.

$$\log(\mathbf{m}) = X\beta + Z\gamma = \beta_0 + \beta_1 \mathrm{D} + \beta_2 \mathrm{U} + \beta_3 \mathrm{V} + \beta_4 \mathrm{DV} + \beta_5 \mathrm{UV} + \gamma \mathrm{V}$$

In this model, $\mathbf{m}$ was the vector of cell counts, β was the vector of fixed effect coefficients and γ was the random effect of disease prevalence. Besides the conditional independence assumption between D and U, the normal distribution of γ with mean 0 and variance σ^2 was assumed. Two algorithms for the estimation were applied: the Gaussian Hermite integration and the Markov Chain Monte Carlo using Gibbs sampling. Estimation procedures using these two algorithms were presented below to analyze the d-dimer data.

Gaussian Hermite integration

As presented earlier, the joint distribution for integration took the following form

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(data|\beta, \gamma = \sqrt{2}x\sigma)e^{-x^2}dx,$$

where f was the likelihood function. This integral can be approximated by summation when using the Gaussian Hermite approximation

$$\frac{1}{\sqrt{\pi}} \sum_{l=1}^{n} w_l f(data|\beta, \gamma = \sqrt{2}\sigma a_l)$$

with abscissas, a_l, and weights, w_l. For the i^{th} DU table, $L(data|\beta, \gamma)_{du}^{i} = e^{m_{du.}^{i} \times log(p_{du.})}$, where $p_{du.}$ is the sum of p_{duv} over the levels of venography. The cell probabilities p_{duv} followed the expressions in Chapter 3 with random effects and evaluated at $\gamma = \sqrt{2}\sigma a_l$. Similarly, the likelihoods for UV and DV tables can be constructed and integrated using abscissas and weights. The joint marginal likelihoods can be obtained as the product of individual marginal likelihood. Maximization of the joint marginal likelihood provided the estimates of model coefficients and the variance of random effect.

Applying all three marginal tables from the d-dimer data, the coefficients of the log-linear model were estimated as the following.

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
-0.860	-2.53	-2.71	2.42	3.56

with estimated variances,

$\hat{V}_{\beta_1}$	$\hat{V}_{\beta_2}$	$\hat{V}_{\beta_3}$	$\hat{V}_{\beta_4}$	$\hat{V}_{\beta_5}$
0.0243	0.104	0.0669	0.0597	0.164

the estimated variance of the random effect γ was $\hat{\sigma^2} = e^{-1.9047886} = 0.149$. Integrating out the random effect, the following unconditional cell probabilities were obtained.

$\hat{p}_{000}$	$\hat{p}_{100}$	$\hat{p}_{010}$	$\hat{p}_{110}$	$\hat{p}_{001}$	$\hat{p}_{101}$	$\hat{p}_{011}$	$\hat{p}_{111}$
0.3324	0.141	0.0265	0.0112	0.0221	0.106	0.062	0.299

Based on the estimated variance matrix of β, the variance matrix of p_{duv} was derived using the delta method.

The estimated sensitivity and specificity of d-dimer were calculated using the unconditional cell probabilities.

$\hat{S}_d$	$\hat{C}_d$
0.8271	0.7026

Estimated variances of $\hat{S}_d$ and $\hat{C}_d$ were calculated using the delta method, as presented in Chapter 3. The standard errors of $\hat{S}_d$ and $\hat{C}_d$ were estimated as 0.024 and 0.033, respectively.

MCMC analysis on the model with one random effect

The analysis of three types of marginal tables from the d-dimer paper using Gibbs sampling followed the same procedures as described in Chapter 3. In brief, the procedure can be summarized below.

1. Specify initial values of $\gamma_i^{(0)}$.

2. Estimate fixed effects $\hat{\beta}$ and variance $\hat{V}_\beta$ based on the likelihood of all the data with the value of $\gamma_i^{(r)}$ from step 1.

3. Sample updated values of $\beta^{(r+1)}$ from $\mathrm{N}(\hat{\beta}, \hat{V}_\beta)$.

4. Calculate the updated values of S_d and C_d based on the updated model coefficients.

5. Update values of the variance of random effects $\sigma^{2(r)}$ based on $\gamma_i^{(r)}$.

6. Estimate the random effect $\hat{\gamma}$ based on the likelihood with $\beta^{(r+1)}$ and the predefined distribution of the random effect with variance $\sigma^{2(r)}$.

7. Use the rejection algorithm to update the random effect $\gamma^{(r+1)}$.

8. Repeat steps 2 - 7 until convergence.

According to Zeger and Karim's paper [94], the update of the variance of random effect σ^2 (step 4 above) can be derived by the following steps.

1. Let $S = \sum_{i=1}^{I} \gamma_i \times \gamma_i$.

2. Calculate the Choleski decomposition of S^{-1}, i.e., $S^{-1} = H'H$.

3. Generate W* from Wishart distribution with $I - q + 1$ degrees of freedom and parameter S, where I was the number of studies and q was the number of random effects.

4. Update $\sigma^2 = (H'W^*H)^{-1}$.

For the log-linear model with 1 random effect, γ_i is a scaler for the i^{th} study. S in the above algorithm is a scaler, which is the sum of γ_i^2 from all studies. Therefore, H is a scaler, which is the square root of S^{-1}. The updated σ^2 was a scaler, simplified as $\sigma^2 = \frac{S}{W^*}$, where W^* was a sample from Wishart distribution with I-q+1 degrees of freedom and parameter S.

Sensitivity and specificity of d-dimer were updated in each iteration. The MCMC algorithm was run for 7000 iterations. Various assessments of convergence were applied: Geweke diagnostics, Raftery and Lewis diagnostics, Heidelberg and Welch diagnostics. Statistics and plots related to these assessments were obtained from the coda package in R.. Samples from the first 2000 iterations were treated as burn-in and

parameters	MCMC sample mean	MCMC sample variance
S_d	0.8261	0.00059
C_d	0.7083	0.00100

Table 4.14: Gibbs sampling results using the model with random disease prevalence

were not included in the posterior sample. Characteristics of 5000 MCMC sample are summarized in Table 4.14.

Histograms of posterior sample of model coefficients are displayed in Figure 4.13. The posterior sampling distributions of all fixed effect coefficients appear to be normally distributed and the running averages of the sample are stable. The posterior samples of sensitivity and specificity appear to be normally distributed, see Figure 4.14. The posterior sampling distribution of the variance of random effect is positively skewed, see Figure 4.15.

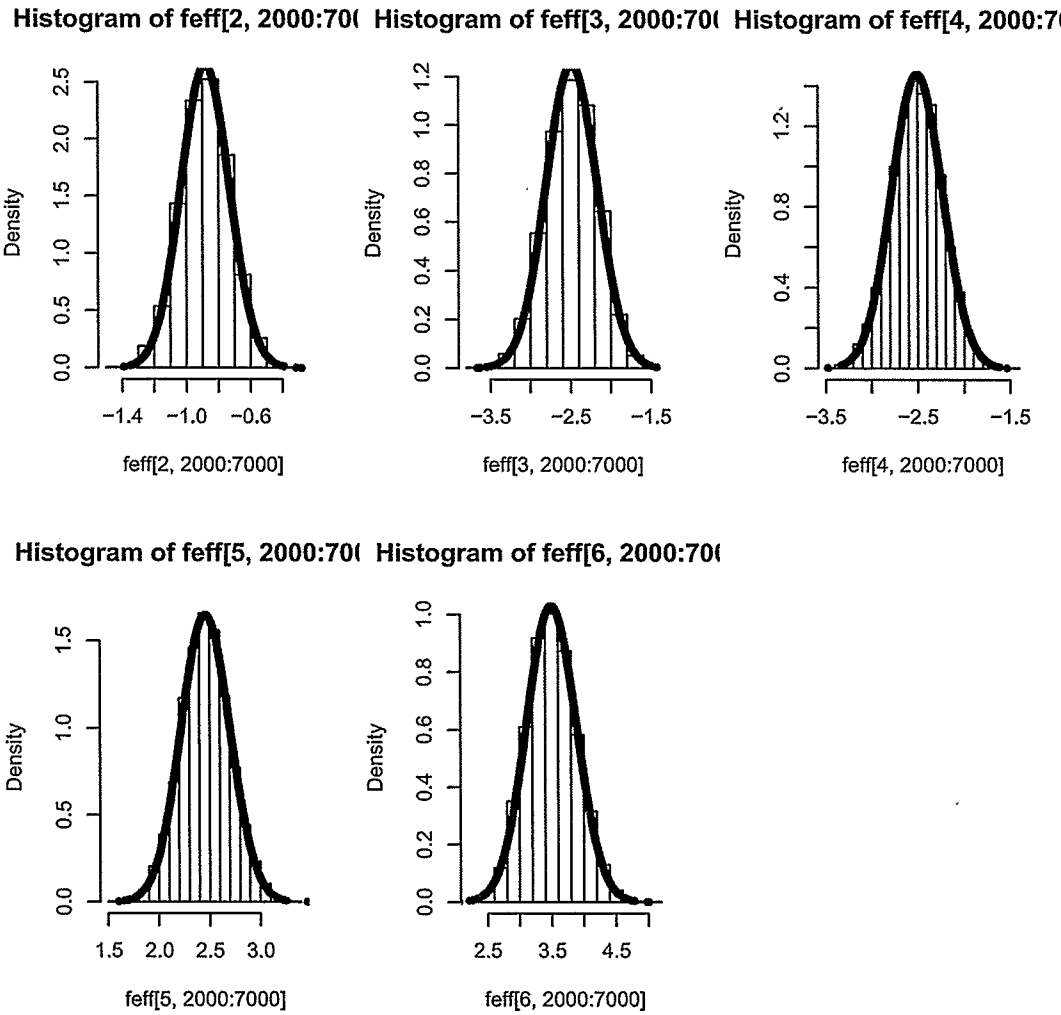

Figure 4.13: Histograms of posterior samples of coefficients in the model with random disease prevalence

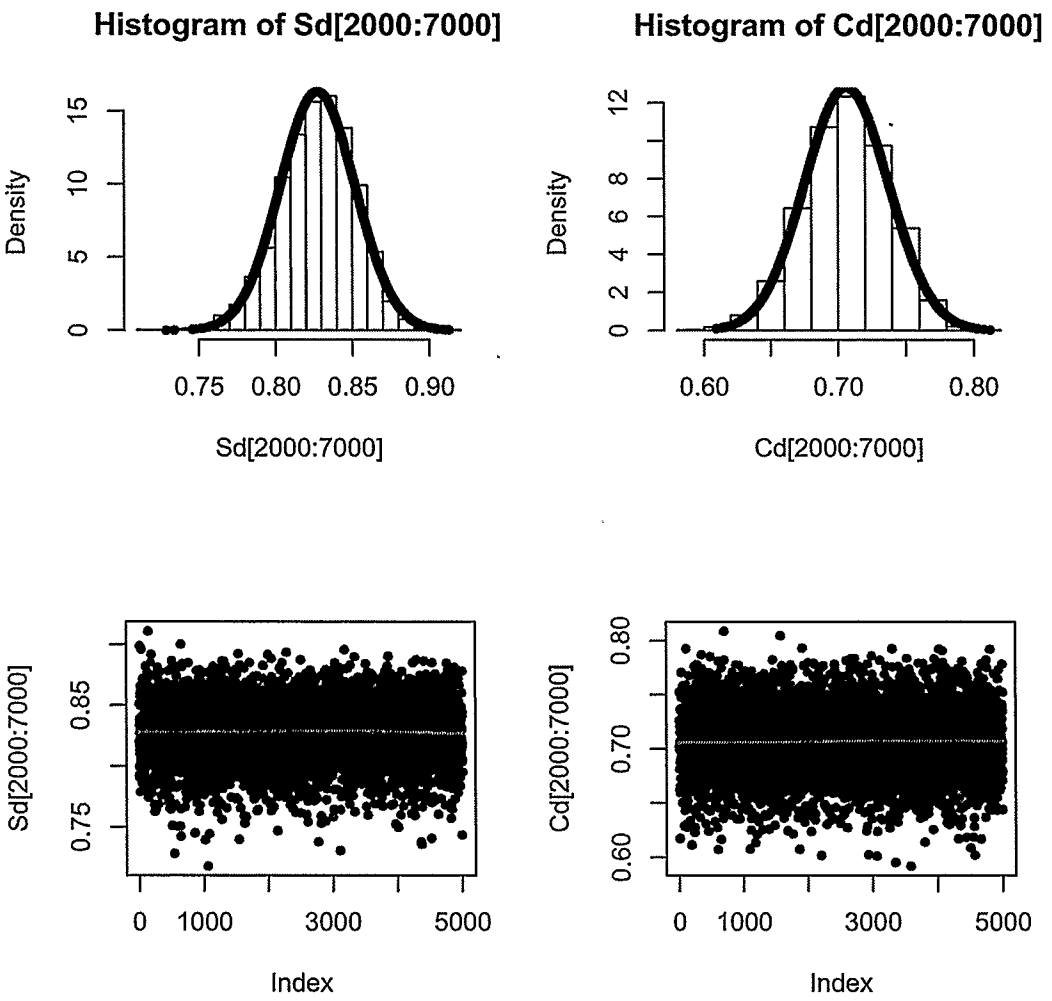

Figure 4.14: Histograms and running averages of posterior samples of sensitivity and specificity in the model with random disease prevalence

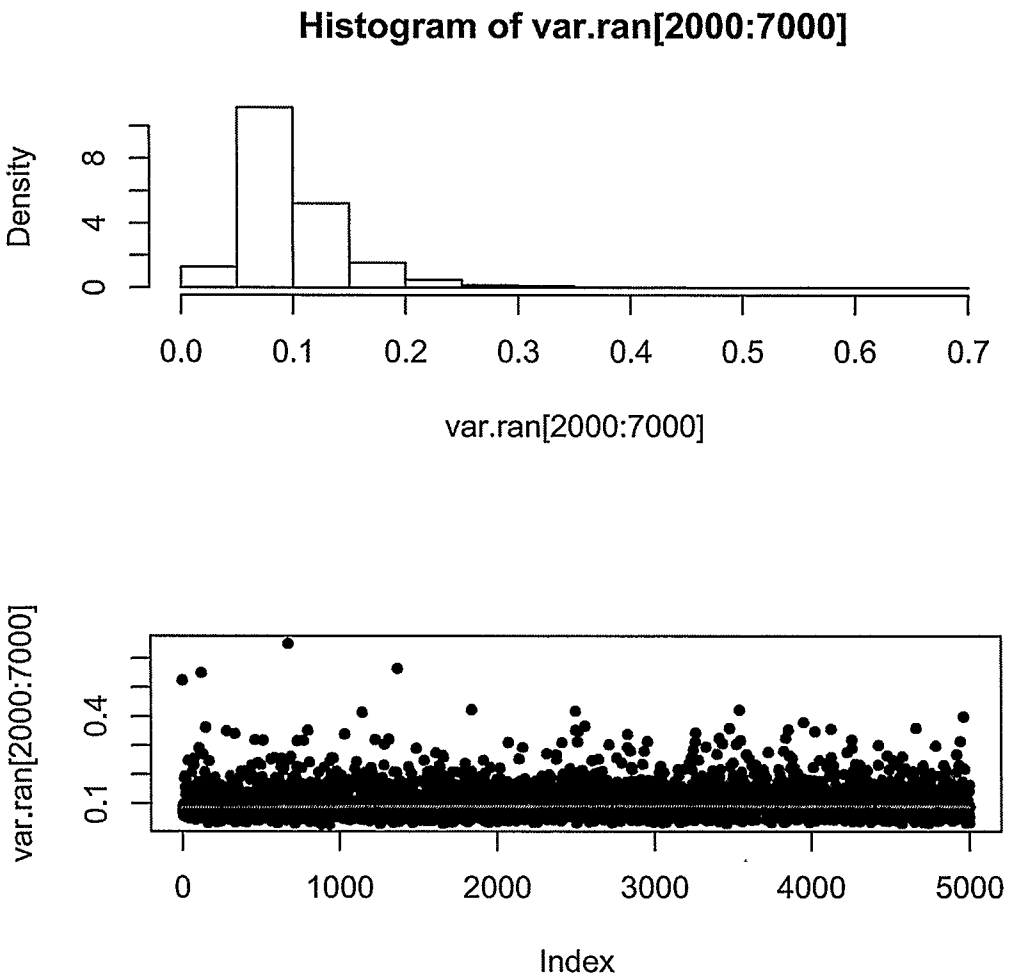

Figure 4.15: Histograms and running average of posterior samples of variance estimate of the random effect in the model with random disease prevalence

The 95% credible set was constructed by the 2.5^{th} percentile and 97.5^{th} percentile of the posterior sample. The posterior credible set of sensitivity was (0.775, 0.871). The posterior credible set of specificity values was (0.644, 0.767).

Comparing the results from Gibbs sampling and those from section 4.4.1, the estimates of S_d (0.8261) and C_d (0.7083) from the Gibbs sampling were very close to those from Gaussian Hermite integration $\hat{S}_d$ (0.8271) and $\hat{C}_d$ (0.7026). The posterior sample variances, 0.00059 and 0.001 for $\hat{S}_d$ and $\hat{C}_d$, were very close to the estimated variances obtained from Gaussian Hermite integration and the delta method, 0.00053 and 0.00116 for $\hat{S}_d$ and $\hat{C}_d$, respectively. The MCMC sample median of variance of random effect, σ^2, was 0.09, which was slightly lower than the estimate from Gaussian Hermite integration, 0.149. Overall, results from Gaussian Hermite integration and Gibbs sampling were very consistent.

4.4.2 Simulations using Gaussian Hermite integration for the model with random disease prevalence

Gibbs sampling and Gaussian Hermite integration provided consistent results in the estimation of model parameters assuming random disease prevalence. Therefore, simulations using Gaussian Hermite integration were expected to produce the same results of simulations using Gibbs sampling. In the following, simulations were conducted using estimates from the Gaussian Hermite integration as true values. The procedure is summarized below.

1. Sample a random effect value from a Normal(0, σ^2) distribution.

2. Calculate cell probabilities based on the log-linear model with known fixed

effect coefficients and the sampled value of the random effect in step 1.

3. Acquire a sample of three-dimensional table from the multinomial distribution using cell probabilities in step 2.
4. Derive the DU marginal table from the table in step 3.
5. Repeat steps 3-4 three times to obtain 3 independent marginal DU tables.
6. Acquire another three-dimension table from the multinomial distribution.
7. Calculate the UV marginal table from the table in step 6.
8. Repeat steps 6-7 five times to obtain 5 independent marginal UV tables.
9. Acquire a new three-dimensional table from multinomial distribution.
10. Calculate the DV marginal table from the table in step 9.
11. Repeat steps 9-10 four times to obtain 4 independent marginal DV tables.
12. Use the marginal tables from steps 5, 8, and 11 as available data and fit the random effects model.
13. Estimate model coefficients and variance of random effects.
14. Integrate out the random effect to obtain unconditional cell probabilities based on estimated coefficients and variance of the random effect.
15. Use the probabilities in step 14 to calculate sensitivity and specificity of d-dimer as well as 95% confidence intervals and examine whether they covered the true values.

16. Repeat steps 1-15 1000 times. Calculate bias, mean squared error, and the coverage rate of 95% confidence intervals.

In order to make the cell probabilities sum to 1, calculation of probabilities in step 2 followed an algorithm similar to that in the fixed effects model. The intercept was a function of the rest parameters in the form of: $\beta_0 = -log(\sum e^{X\beta+Z\gamma})$. The estimated sensitivity and specificity of d-dimer and 95% confidence interval were calculated at the end of each iteration.

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.827	0.828	-0.000495 (0.0134)	0.00018	0.951
C_d	0.703	0.703	0.00010 (0.0162)	0.00026	0.946

Table 4.15: Simulation results from Gaussian Hermite integration with random disease prevalence only

Comparisons of true values and estimates are summarized in Table 4.15. Results from simulations showed that the analysis using Gaussian Hermite integration produced estimates with very small biases and small mean squared errors (MSE). The coverages of confidence intervals for sensitivity and specificity were very close to 95%. The sample variances of estimated sensitivity and specificity were 0.0001769 and 0.0002528, respectively. The mean estimated variance of sensitivity was 0.000178, which was close to the empirical variance of estimated sensitivity. The mean estimated variance of specificity was 0.00025, which was close to the empirical variances of estimated specificity. The mean estimated variance of random effect was 0.368, which was higher than the true value of the pre-defined variance value of 0.149.

Treating ultrasonography as the gold standard

If the marginal tables from d-dimer versus ultrasonography were treated as true classifications, the model with random venography can be expressed as the following.

$$log(m_{dv}) = \beta_0^* + \beta_1^* D + \beta_2^* V + \beta_3^* DV + \gamma^* V$$

In this model, $\gamma^* \sim N(0, \sigma^2)$. The same procedure of estimation was applied to this model using Gaussian Hermite integration. With one random effect in the log-linear model, the likelihood function for the i^{th} study was constructed as the following.

$$L^i(m_{dv}|\beta^*, \gamma^*) = e^{table^i log(p_{dv})}$$

Note that the UV tables were not used in the analysis because ultrasonography was considered to be an error-free reference. The likelihood for the i^{th} study integrating out the random effect γ^* can be approximated by the summation below:

$$\frac{1}{\sqrt{\pi}} \sum_{k=1}^{n} w_k L^i(m_{dv}|\beta^*, \gamma^* = \sqrt{2}\sigma a_k).$$

The product of these likelihoods constituted the likelihood function which was maximized. The four cell probabilities are given by the following expressions.

$$p_{11} = \frac{e^{\beta_1^*+\beta_2^*+\beta_3^*+\gamma}}{1+e^{\beta_1^*}+e^{\beta_2^*+\gamma}+e^{\beta_1^*+\beta_2^*+\beta_3^*+\gamma}}$$

$$p_{01} = \frac{e^{\beta_2^*+\gamma}}{1+e^{\beta_1^*}+e^{\beta_2^*+\gamma}+e^{\beta_1^*+\beta_2^*+\beta_3^*+\gamma}}$$

$$p_{10} = \frac{e^{\beta_1^*}}{1+e^{\beta_1^*}+e^{\beta_2^*+\gamma}+e^{\beta_1^*+\beta_2^*+\beta_3^*+\gamma}}$$

$$p_{00} = \frac{1}{1+e^{\beta_1^*}+e^{\beta_2^*+\gamma}+e^{\beta_1^*+\beta_2^*+\beta_3^*+\gamma}}$$

The first derivatives of p_{dv} with respect to β_1^*, β_2^*, and β_3^* can be expressed as the following.

$$\frac{\partial p_{dv}}{\partial \beta_1^*} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{11} \\ p_{01} \\ p_{10} \\ p_{00} \end{pmatrix} - \int_{-\infty}^{+\infty} \begin{pmatrix} p_{11|\gamma} \\ p_{01|\gamma} \\ p_{10|\gamma} \\ p_{00|\gamma} \end{pmatrix} \times (p_{11} + p_{10}) f(\gamma) d\gamma$$

$$\frac{\partial p_{dv}}{\partial \beta_2^*} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{11} \\ p_{01} \\ p_{10} \\ p_{00} \end{pmatrix} - \int_{-\infty}^{+\infty} \begin{pmatrix} p_{11|\gamma} \\ p_{01|\gamma} \\ p_{10|\gamma} \\ p_{00|\gamma} \end{pmatrix} \times (p_{11} + p_{01}) f(\gamma) d\gamma$$

$$\frac{\partial p_{dv}}{\partial \beta_3^*} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} p_{11} \\ p_{01} \\ p_{10} \\ p_{00} \end{pmatrix} - \int_{-\infty}^{+\infty} \begin{pmatrix} p_{11|\gamma} \\ p_{01|\gamma} \\ p_{10|\gamma} \\ p_{00|\gamma} \end{pmatrix} \times p_{11} f(\gamma) d\gamma$$

The summary matrix of derivatives can be constructed below.

$$D_{p\beta^*} = \frac{\partial p_{dv}}{\partial \beta^*} = \begin{pmatrix} \frac{\partial p_{00}}{\partial \beta_1^*} & \frac{\partial p_{00}}{\partial \beta_2^*} & \frac{\partial p_{00}}{\partial \beta_3^*} \\ \frac{\partial p_{10}}{\partial \beta_1^*} & \frac{\partial p_{10}}{\partial \beta_2^*} & \frac{\partial p_{10}}{\partial \beta_3^*} \\ \frac{\partial p_{01}}{\partial \beta_1^*} & \frac{\partial p_{01}}{\partial \beta_2^*} & \frac{\partial p_{01}}{\partial \beta_3^*} \\ \frac{\partial p_{11}}{\partial \beta_1^*} & \frac{\partial p_{11}}{\partial \beta_2^*} & \frac{\partial p_{11}}{\partial \beta_3^*} \end{pmatrix}$$

The estimated variance matrix of p_{dv} can be derived by the delta method: $\hat{V}_{p_{dv}} = D_{p\beta^*} \hat{V}_{\beta^*} D'_{p\beta^*}$. The derivatives of sensitivity and specificity of d-dimer with respect to the four cell probabilities in this model were:

$$D_{scp} = \begin{pmatrix} 0 & 0 & \frac{-p_{11}}{(p_{11}+p_{01})^2} & \frac{p_{01}}{(p_{11}+p_{01})^2} \\ \frac{p_{10}}{(p_{00}+p_{10})^2} & \frac{-p_{00}}{(p_{00}+p_{10})^2} & 0 & 0 \end{pmatrix}$$

The estimated variance matrix of sensitivity and specificity of d-dimer was given by: $\hat{V}_{sc} = D_{scp}\hat{V}_{p_dv}D'_{scp}$.

Using the above expressions, estimates of model coefficients were obtained at the end of the maximization step. Integrating out the random effect, the estimated cell probabilities can then be derived.

$\hat{p}_{00}$	$\hat{p}_{10}$	$\hat{p}_{01}$	$\hat{p}_{11}$
0.363	0.224	0.0838	0.329

The estimated sensitivity and specificity of d-dimer were 0.797 and 0.618, respectively. Applying the delta method, the standard errors of $\hat{S}_d$ and $\hat{C}_d$ were 0.024 and 0.024, respectively. The estimated S_d and C_d were substantially lower than those from the model where ultrasonography was treated as an imperfect reference ($\hat{S}_d$=0.8261 and $\hat{C}_d$=0.7083).

Simulations comparing the unadjusted and adjusted models

In order to compare the difference in estimations from the two models, simulations were performed. Marginal tables were generated using the procedure of simulations on the model with ultrasonography as an imperfect reference. The estimated βs from the adjusted model were applied as true parameter values of fixed effect coefficients in the model. In other words, (β_1=-0.85958023, β_2=-2.52744121, β_3=-2.71081593, β_4=2.42450599, β_5=3.56453207) were used as true coefficients in the log-linear model and σ^2=0.149 was used as the true variance of random effect. Thus, the true sensitivity and specificity of d-dimer were set at 0.8271 and 0.7026, respectively, for simulations. The two models were applied to the same set of tables to acquire estimates and variances of S_d and C_d. The numbers of DU, DV, and UV tables were 3,

4, and 5, respectively. The table total was 300. At the end of simulations, bias, mean square error (MSE), and 95% coverage rates from the two models were compared. The procedure can be summarized below.

1. Calculate the true values of S_d and C_d based on true values of β integrating out the random effect, which was normally distributed around zero with variance σ^2 (0.149).

2. Sample a random effect value from Normal(0, 0.149).

3. Calculate cell probabilities based on the log-linear model with known fixed effect coefficients and the sampled value of the random effect in step 2.

4. Acquire a sample of three-dimensional table from the multinomial distribution with cell probabilities in step 3.

5. Calculate the DU marginal table from the table in step 4.

6. Repeat steps 4-5 three times to obtain three DU tables.

7. Acquire another three-dimension table from the multinomial distribution with cell probabilities in step 3.

8. Calculate the UV marginal table from the table in step 6.

9. Repeat steps 7-8 five times to obtain five UV tables.

10. Acquire a new three-dimensional table from multinomial distribution with cell probabilities in step 3.

11. Calculate the DV marginal table from the table in step 8.

12. Repeat steps 10-11 four times to obtain four DV tables.

13. Use the marginal tables from steps 6, 9, and 12 as available data and fit the random effects model treating ultrasonography as the silver standard. Estimate model coefficients, variance of random effects, unconditional cell probabilities, sensitivity and specificity. Calculate 95% confidence interval of S_d and C_d.

14. Use the marginal tables from steps 6, 9, and 12 as available data and fit the random effects model treating ultrasonography as the gold standard. Estimate model coefficients, variance of random effects, marginal cell probabilities, sensitivity and specificity. Calculate 95% confidence intervals for S_d and C_d.

15. Repeat steps 2-14 1000 times.

16. Calculate bias, mean squared error (MSE), and coverage rate of 95% confidence intervals for each model.

Models	unadjusted	adjusted
Bias S_d (s.e.)	-0.0199 (0.0136)	-0.000652 (0.0138)
Bias C_d (s.e.)	-0.0532 (0.0156)	3.52e-05 (0.0160)
MSE S_d	0.000579	0.000192
MSE C_d	0.00307	0.000257
95% coverage S_d	0.659	0.946
95% coverage C_d	0.039	0.957

Table 4.16: Comparison between the unadjusted and the adjusted model with random venography only using small number of tables

Note that estimations in the model where ultrasonography was treated as the gold standard did not include the UV tables generated at each iteration. Results from the two models were compared in Table 4.16. From Table 4.16, the magnitudes

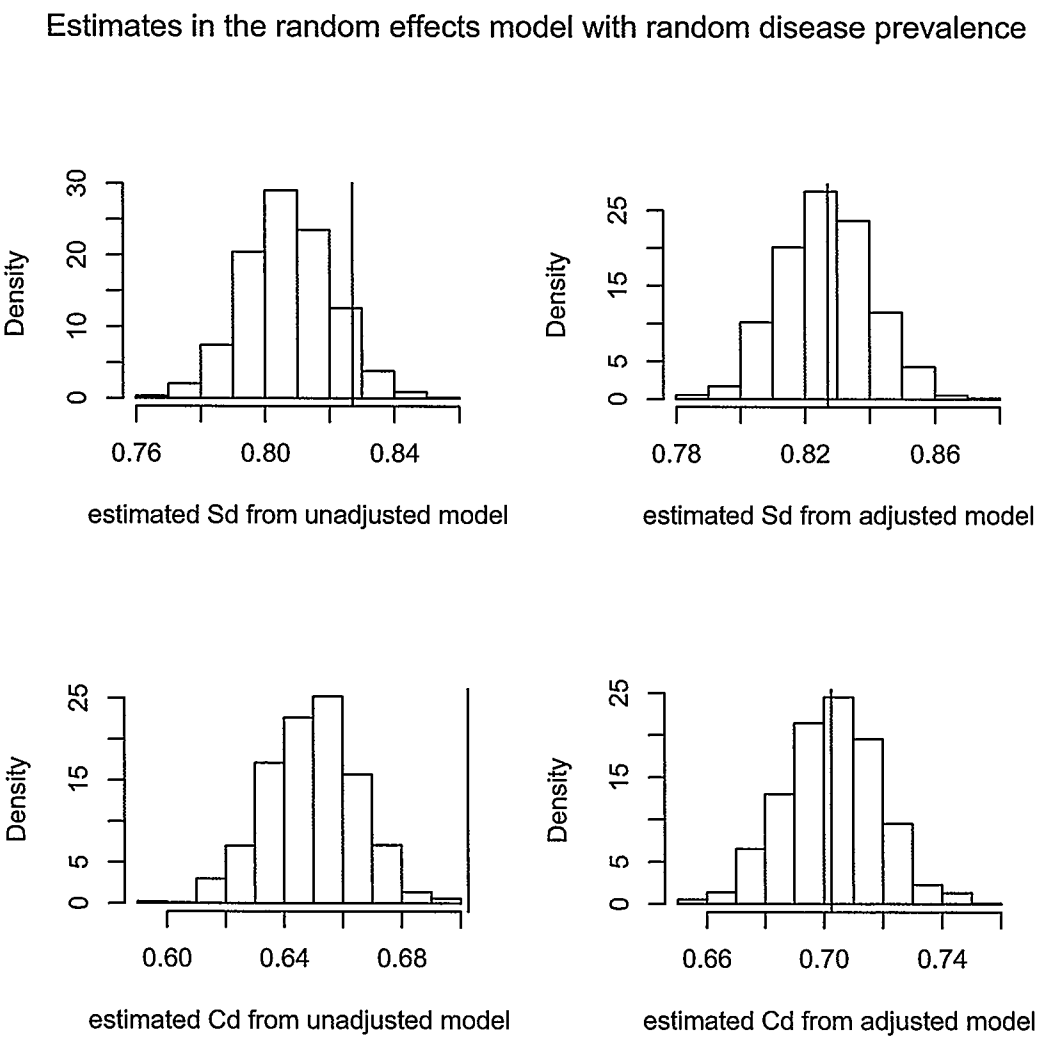

Figure 4.16: Histograms of estimated sensitivity and specificity from the two models with random disease prevalence

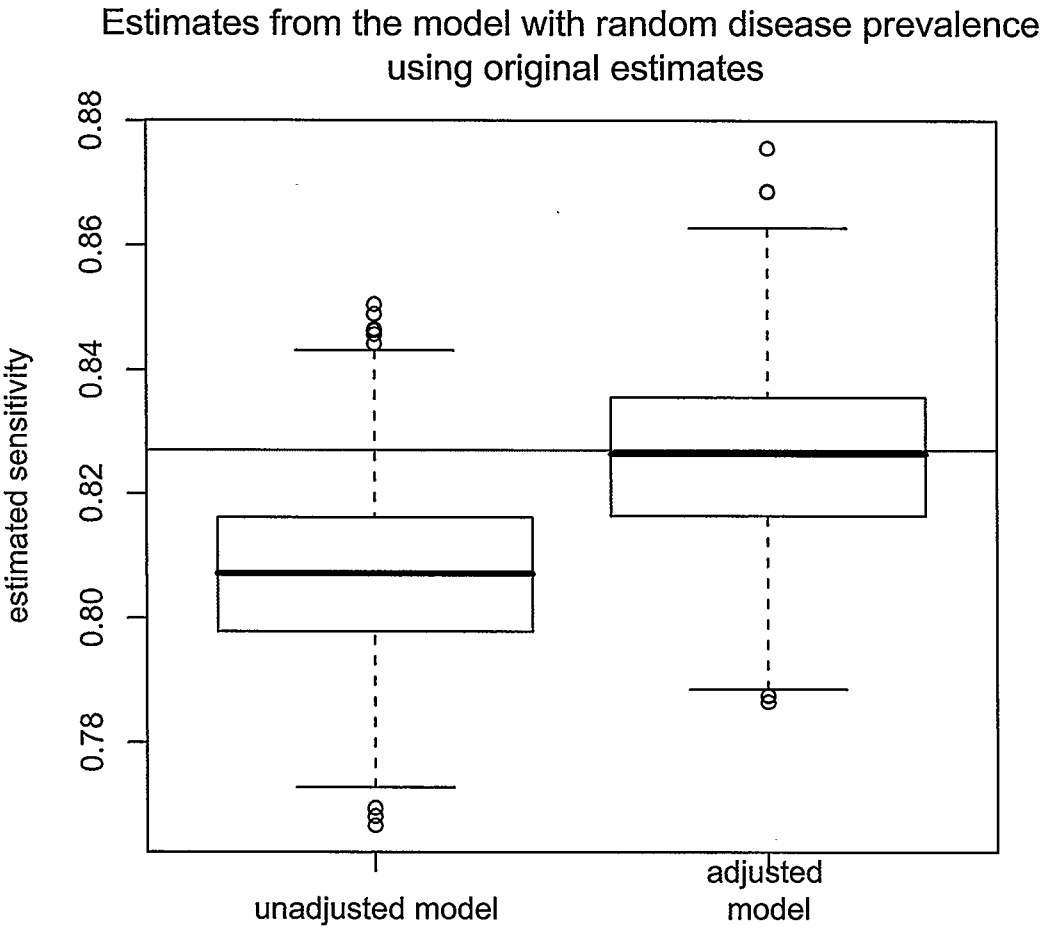

Figure 4.17: Boxplots of estimated sensitivity from the two models with random disease prevalence

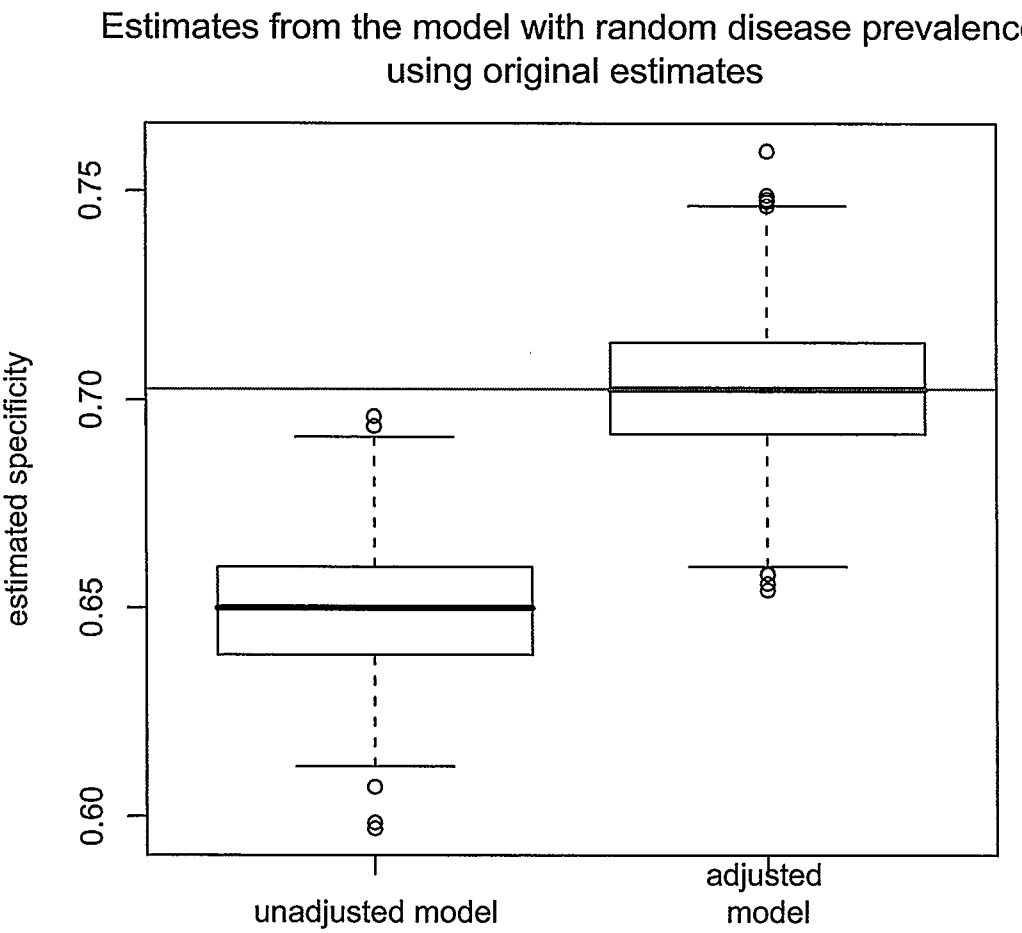

Figure 4.18: Boxplots of estimated specificity from the two models with random disease prevalence

of biases from the unadjusted model were much higher than those from the model adjusting for the difference between ultrasonography and venography. The model ignoring the error introduced by ultrasonography resulted in substantially biased estimates of sensitivity and specificity of d-dimer. The mean squared error from this model was much higher than that from the adjusted model. The coverage rate of sensitivity from the model ignoring the imperfection of ultrasonography was much lower than that from the adjusted model. The coverage rate of specificity from the model ignoring the difference between ultrasonography and venography was close to zero. Estimates of variance of random effects in both models were higher than the true variance.

Figures 4.16, 4.17, and 4.18 display the distributions of estimated sensitivity and specificity of d-dimer from the unadjusted model and the adjusted model. The distributions from the unadjusted model were not centered on the true parameter values, whereas distributions from the adjusted model were centered nicely on the true parameter values. Although the standard errors were larger than those in the fixed effects model, the biases from the unadjusted model increased, especially in specificity. Therefore, the coverage of confidence intervals was still very low in the unadjusted model.

4.4.3 Comparison between the analysis using all tables and the analysis using tables from the test and the gold standard only in the model with random disease prevalence

In this section, two models were compared: the model using the DV tables only and the model using all three types of tables. Simulations were performed to assess the

efficiency gained by including tables between d-dimer and ultrasonography in the analysis. The parameters for simulations were the same as those in the simulations comparing the unadjusted and adjusted models. The number of each type of tables was set to 10. The table total was 300.

Analysis	DV tables only	all tables
Bias S_d (s.e.)	0.000517 (0.01)	0.000414 (0.00836)
Bias C_d (s.e.)	0.000172 (0.012)	8.32e-05 (0.00974)
MSE S_d	0.000101	7.00e-05
MSE C_d	0.000137	9.49e-05
95% coverage S_d	0.939	0.949
95% coverage C_d	0.955	0.953

Table 4.17: Comparison between the analysis using the test of interest versus gold standard only and the analysis using all three types of tables with random disease prevalence

Table 4.17 showed that the coverage of 95% from both models was very close to 0.95. The standard errors and mean squared errors from the model using the DV tables only were larger than those from the model using all tables. The efficiency of using DV tables only relative to that of using all tables was 0.69 for sensitivity and 0.695 for specificity. Using the DV tables only in the analysis resulted in around 30% loss of information. In other words, removing the 10 DU tables from the analysis led to 30% loss of information on the diagnostic characteristics of d-dimer.

In order to assess the relative number of DV and DU tables, simulations were performed with the same numbers of DV and UV tables as in the simulation in the previous section, i.e., 4 DV tables and 5 UV tables. But the number of DU tables was reduced to 1 for simulations. The model coefficients and variance of random effect were set the same as previous simulations. The table total was 300. The relative efficiency of the analysis using DV tables only to the analysis using all tables

Analysis	DV tables only	all tables
Bias S_d (s.e.)	-1.9e-05 (0.0156)	-0.00030 (0.0146)
Bias C_d (s.e.)	0.0012 (0.0183)	0.0014 (0.0171)
MSE S_d	0.00024	0.00021
MSE C_d	0.00034	0.00030
95% coverage S_d	0.95	0.946
95% coverage C_d	0.95	0.952

Table 4.18: Comparison between the analysis using 4 tables of the test of interest versus gold standard and the analysis using all tables but only one misclassified table in the model with random disease prevalence

was 0.887756 in estimating sensitivity. The relative efficiency was 0.889 in estimating specificity. There was approximately 10% reduction in efficiency of the analysis using the DV tables only if 4 DV tables and 1 DU table were available.

When the number of DV tables increased to 10 and the numbers of DU table and UV table were both 1, simulations were performed to compare the efficiencies from the two analyses. Results from simulations were summarized in Table 4.19.

Analysis	DV tables only	all tables
Bias S_d (s.e.)	-5.35e-05 (0.010)	-0.000120 (0.00976)
Bias C_d (s.e.)	-4.19e-05 (0.012)	1.53e-05 (0.0114)
MSE S_d	9.96e-05	9.51e-05
MSE C_d	0.000134	0.000129
95% coverage S_d	0.948	0.942
95% coverage C_d	0.947	0.948

Table 4.19: Comparison between the analysis using 10 tables of the test of interest versus gold standard and the analysis using all tables but only one misclassified table in the model with random disease prevalence

The mean square errors from both analyses were substantially reduced compared to the analysis above with 4 DV tables. The relative efficiency of the analysis using DV tables only to analysis using all tables was 0.956 in estimating sensitivity. The

relative efficiency in estimating specificity was 0.964.

Compared to the previous simulation, the mean squared errors the analysis using DV tables only were very close to those from the analysis using all tables if the number of DV tables was ten times that of the DU tables. When the number of DV tables was reduced to 4, the efficiency gained by using all tables was slightly increased to around 11%. In the end, when only one misclassified table was collected, the loss of efficiency using the simple pooled analysis of tables from the gold standard alone was small.

4.5 Model accounting for heterogeneity in disease prevalence and association between the test and the gold standard across studies

4.5.1 Analysis of the log-linear model with two random effects

The heterogeneity of disease prevalence was represented by the random effect of venography in the log-linear model. In this section, another random effect was added to the model to represent heterogeneity of association between d-dimer and venography among studies. As discussed in Chapter 3, the log-linear model with two random effects for the i^{th} table was written as the following.

$$\log(m_i) = \mathrm{X}\beta + \mathrm{Z}\gamma_i = \beta_0 + \beta_1\mathrm{D} + \beta_2\mathrm{U} + \beta_3\mathrm{V} + \beta_4\mathrm{DV} + \beta_5\mathrm{UV} + \gamma_{1i}\mathrm{V} + \gamma_{2i}\mathrm{DV},$$

where $\begin{pmatrix} \gamma_{1i} \\ \gamma_{2i} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}\right)$. To analyze the model with two random effects, two algorithms were applied to estimate parameters: Gaussian Hermite integration and the Gibbs sampling.

Gaussian Hermite integration on the model with two random effects

In order to obtain estimates of the fixed effect coefficients β, the marginal likelihood for each table was required, which was obtained by integrating out the random effects from the full likelihood. The integral was approximated by the expression in the Gaussian Hermite integration standard format below.

$$\frac{1}{\pi}\sum_{l=1}^{n}\sum_{q=1}^{n} w_l w_q f(data|\beta, \gamma_{i1} = \sqrt{2}\sigma_1 a_{li1}, \gamma_{i2} = \sqrt{2}\sigma_2 a_{qi2}).$$

In this expression, $f(data|\beta, \gamma_{i1} = \sqrt{2}\sigma_1 a_{li1}, \gamma_{i2} = \sqrt{2}\sigma_2 a_{qi2})$ was the likelihood function and positions of γ_{i1} and γ_{i2} were replaced by $\sqrt{2}\sigma_1 a_{li1}$ and $\sqrt{2}\sigma_2 a_{qi2}$. The function f was the likelihood of each table and was in one of the following forms.

$$L_{du}^{i} = e^{table_{du}^{i} log(p_{du.})}$$
$$L_{uv}^{j} = e^{table_{uv}^{j} log(p_{.uv})}$$
$$L_{dv}^{k} = e^{table_{dv}^{k} log(p_{d.v})}$$

The marginal likelihood of the i^{th} DU table was approximated by the following expression.

$$L_{du}^{i} = \frac{1}{\pi}\sum_{l=1}^{n}\sum_{q=1}^{n} w_l w_q e^{table_{du}^{i} log(p_{du.})}$$

Similar expressions can be derived for UV and DV tables. The marginal probabilities $p_{du.}$, $p_{.uv}$, and $p_{d.v}$ were calculated from the cell probabilities p_{duv} based on the log-linear model with two random effects. Expressions of the cell probabilities can be referred to those from Chapter 3. Using data from the d-dimer paper [68], the model coefficients were estimated.

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
-0.988	-2.46	-2.63	2.60	3.35

The estimated standard deviations of random effects were $\hat{\sigma}_1$=0.0005267 and $\hat{\sigma}_2$=0.5876.

Meaningful estimates were the sensitivity and specificity of d-dimer based on cell probabilities integrating out random effects. The Gaussian Hermite approach was applied again for integration.

$\hat{p}_{000}$	$\hat{p}_{100}$	$\hat{p}_{010}$	$\hat{p}_{110}$	$\hat{p}_{001}$	$\hat{p}_{101}$	$\hat{p}_{011}$	$\hat{p}_{111}$
0.333	0.124	0.0284	0.0106	0.0239	0.123	0.0581	0.299

The resulting estimates of sensitivity and specificity of d-dimer were 0.8371 and 0.7287, respectively. The estimated variances of $\hat{S}_d$ and $\hat{C}_d$ were calculated using the delta method. Based on the estimated variance matrix of β, the estimated variance matrix of p_{duv} was derived using the delta method. The steps to obtain estimated variance covariance matrix of sensitivity and specificity were similar to those in the model with one random effect. The standard errors of sensitivity and specificity were 0.031 and 0.030, respectively.

Analysis using Gibbs sampling

In this section, the Gibbs sampling algorithm was applied to analyze the log-linear model with two random effects. The procedure using Gibbs sampling in this model was similar to that in the model with 1 random effect. Sampled values from conditional distributions were obtained at each step. An important modification in the algorithm was the derivation of the conditional distribution of variances given random effects, i.e., $f(\Sigma|\gamma)$. The update of variances σ_1^2 and σ_2^2 given γ followed the steps below.

1. Calculate $S^{(r)} = \sum_{i=1}^{I} \gamma_i^{(r)} \gamma_i^{(r)'}$.

2. Generate a sample value W^* from the Wishart distribution with $I - q + 1$ degrees of freedom and parameter S, where I was the number of studies and q was the number of random effects.

3. Calculate Choleski decomposition of $S^{(r)-1}$, namely H, so that $S^{(r)-1} = H^{(r)'}H^{(r)}$.

4. Σ was updated by $(H^{(r)'}W^*H^{(r)})^{-1}$.

5. The off-diagonal of Σ was replaced by 0 because the random effects were assumed uncorrelated.

The sensitivity and specificity were updated at each round of the Gibbs sampling process. At the end of the iteration, a posterior sample of sensitivity and specificity was obtained.

By deleting the first 2000 burn-in, posterior sampling distributions of sensitivity and specificity of d-dimer were displayed in Figure 4.19. The distributions of sensitivity and specificity were normal. The posterior sampling distributions of variances of random effect were displayed in Figure 4.20. The histograms showed a positively skewed distribution of both variances.

Parameters	MCMC Sample mean	MCMC Sample standard deviation
S_d	0.8095	0.041
C_d	0.7128	0.0375

Table 4.20: Gibbs sampling results using the model with random disease prevalence and random interaction between d-dimer and venography

Summary of the posterior samples of sensitivity and specificity of d-dimer was given in Table 4.20. Comparisons of the two sets of results from Gaussian Hermite

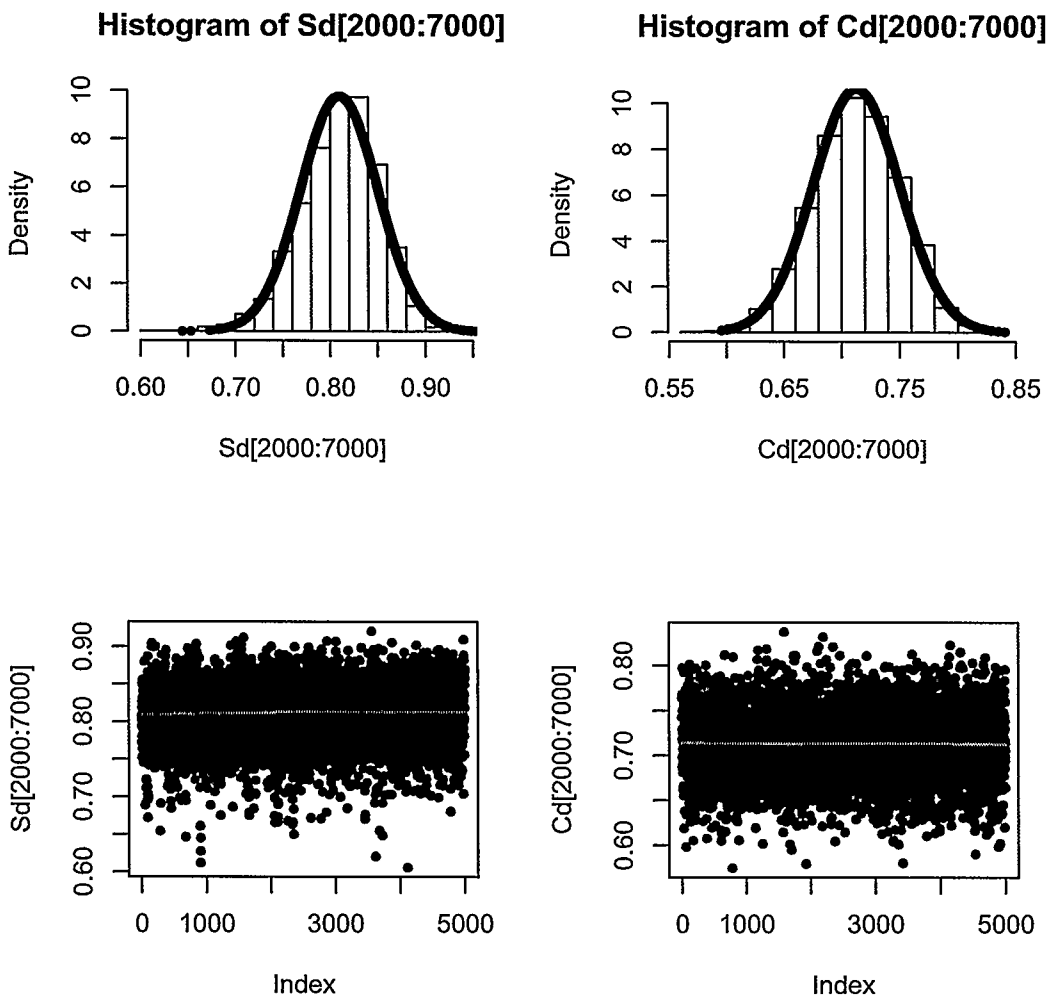

Figure 4.19: Histograms and running averages of posterior samples of sensitivity and specificity in the model with two random effects

Parameters	GHI estimates	s.e.	Gibbs sampling estimates	s.e.
S_d	0.8371	0.031	0.8095	0.041
C_d	0.7287	0.030	0.7128	0.0375

Table 4.21: Comparison of estimates from Gaussian Hermite integration and Gibbs sampling in the model with two random effects and 0-1 contrast

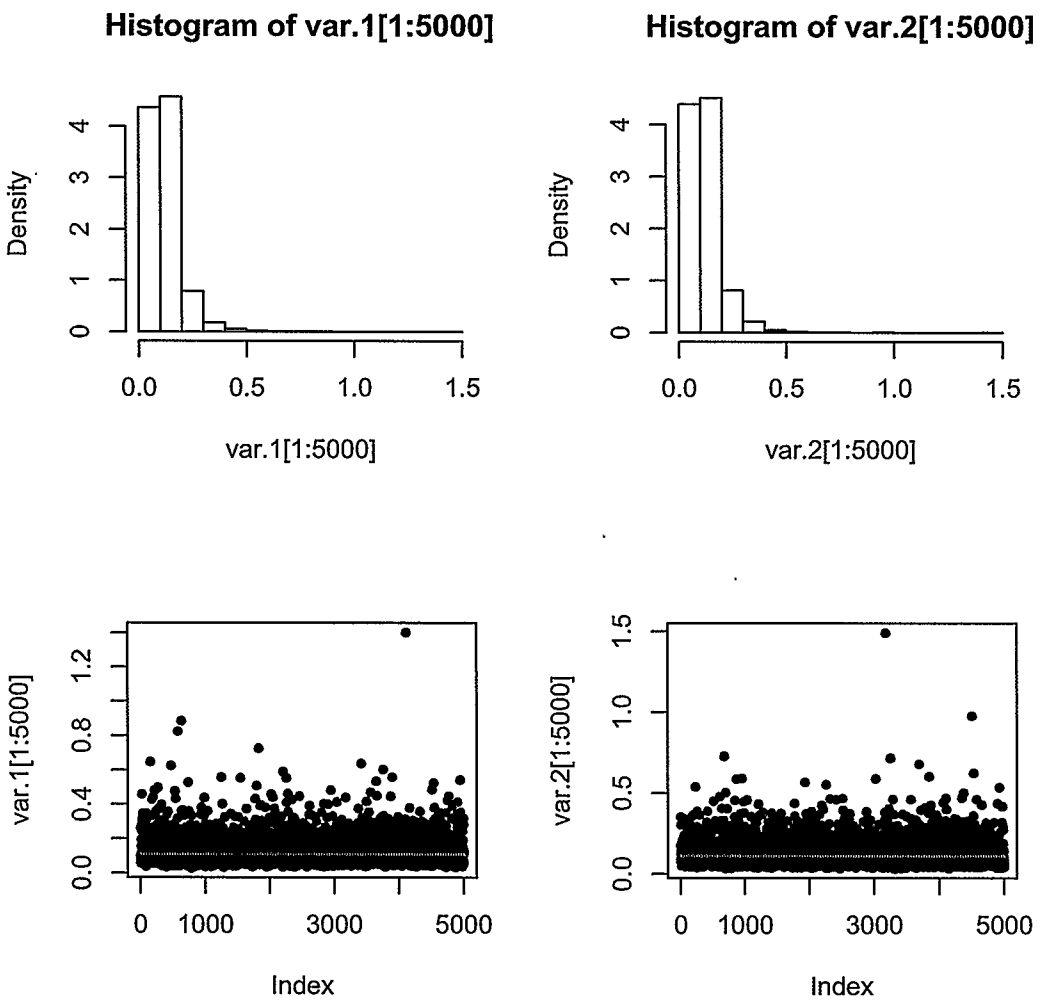

Figure 4.20: Histograms and running averages of posterior samples of variances of random effects in the model with two random effects

integration and the Gibbs sampling were summarized in Table 4.21.

Estimations of S_d and C_d from the two methods were slightly different from each other. The posterior sample standard deviations of S_d and C_d were slightly different from the standard errors from the Gaussian Hermite approach. The posterior sample median of σ_1^2 was 0.108 and that of σ_2^2 was 0.107. They were not consistent with estimates from the Gaussian Hermite integrations. Simulations for the model with two random effects using Gaussian Hermite integration did not produce satisfactory results. The two algorithms did not provide consistent results in the log-linear model with two random effects.

Gaussian Hermite integration using a new design matrix on random effects

The design matrix of random effects in the above approaches, Z, used the 0 1 contrast. As a result of zeros in the matrix, the random effect of venography affected only four cells which corresponded to the four rows of 1 in the first column of Z. Similarly, the random effect of interaction between d-dimer and venography affected only two cells which corresponded to the two rows of 1 in the interaction. In other words, by such a model, four cells with zeros in both columns of Z were not affected by the random effects if the constraint of summing to one was set aside at this point.

In order to reflect random variations across all cells, a new design matrix of Z was considered. As presented in Chapter 3, the new Z matrix was constructed below.

$$Z = \begin{pmatrix} V & DV \\ -1 & 1 \\ -1 & -1 \\ -1 & 1 \\ -1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$$

This Z matrix was applied in the log-linear model with two random effects and in the simulations using Gaussian Hermite integration.

Using the new design matrix in the log-linear model with two random effects, the model coefficients were estimated as the following.

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
−0.878	−2.5206	−2.7002	2.4493	3.538

The estimated variances of random effects were $\hat{\sigma}_1^2$=0.039 and $\hat{\sigma}_2^2$=0.001946.

Clinically meaningful measures were the sensitivity and specificity of d-dimer. The Gaussian Hermite approach was applied again to acquire unconditional cell probabilities by integrating out random effects.

$\hat{p}_{000}$	$\hat{p}_{100}$	$\hat{p}_{010}$	$\hat{p}_{110}$	$\hat{p}_{001}$	$\hat{p}_{101}$	$\hat{p}_{011}$	$\hat{p}_{111}$
0.333	0.139	0.0268	0.0111	0.0224	0.108	0.0621	0.298

Based on the estimated probabilities, the sensitivity and specificity of d-dimer were calculated as 0.8276745 and 0.705982, respectively. Based on the estimated variance

matrix of β, the variance matrix of p_{duv} was derived using the multivariate delta method. Applying the delta method again, the standard errors of $\hat{S}_d$ and $\hat{C}_d$ were calculated as 0.0246 and 0.0342.

4.5.2 Simulations of the model with two random effects using the new design matrix

The estimated model coefficients (-0.878, -2.5206, -2.7002, 2.4493, 3.538) and variances of random effects (0.039, 0.001946) were applied as true parameter values for 1000 simulations. The corresponding true values of sensitivity and specificity of d-dimer were, as stated in the estimation before, 0.8276745 and 0.705982, respectively. The number of each type of marginal tables was the same as that in the analysis. The table total was 300. The simulation procedure was summarized below.

1. Sample γ_1 and γ_2 from bivariate Normal distribution with mean $\mathbf{0}$ and variance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 0.039 & 0 \\ 0 & 0.001946 \end{pmatrix}$.

2. Calculate cell probabilities based on known model coefficients and values of random effects in step 1.

3. Acquire a sample of three-dimensional table from the multinomial distribution with cell probabilities in step 2.

4. Obtain the DU marginal table from the table in step 3.

5. Repeat steps 3-4 three times to obtain 3 independent marginal DU tables.

6. Acquire another set of three-dimension table from the multinomial distribution with the same set of cell probabilities in step 2.

7. Obtain the UV marginal table from the table in step 6.

8. Repeat steps 6-7 five times to obtain 5 independent marginal UV tables.

9. Acquire the third set of three-dimensional table from multinomial distribution with the same set of parameter values in step 2.

10. Obtain the DV marginal table from the table in step 9.

11. Repeat steps 9-10 four times to obtain 4 independent marginal DV tables.

12. Use the marginal tables from steps 5, 8, and 11 as available data and fit the random effects model.

13. Estimate model coefficients and variances of random effects.

14. Integrate out random effects based on estimated coefficients and variances.

15. Calculate sensitivity and specificity as well as correponding 95% confidence intervals.

16. Repeat steps 1-15 1000 times and calculate bias, mean squared error, and 95% coverage.

Results from simulations were summarized in Table 4.22. The mean estimated value σ_1^2 was 0.0373 and the mean estimated σ_2^2 was 0.001849. The true values for σ_1^2 and σ_2^2 were 0.039 and 0.001946, respectively. The estimated variances of random effects using the new design matrix of random effects in this model were very close

Parameters	True Values	Mean Estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.8276745	0.82716	-0.000512 (0.0143)	0.00021	0.944
C_d	0.705982	0.70512	-0.000861 (0.0177)	0.00031	0.938

Table 4.22: Simulations using the new design matrix on the model with two random effects

to the true values. Results from Table 4.22 indicated that estimates of sensitivity and specificity in this model were nearly unbiased. The 95% coverage of confidence intervals was close to 0.95. Overall, this approach using the new design matrix for random effects worked very well for the log-linear model with two random effects.

4.6 Sample size issues

In the analysis of d-dimer data, 3 DU tables, 4 DV tables, and 5 UV tables were available for the meta-analysis. The parameters being estimated were 5 fixed effect coefficients and 2 random effect variances. The number of independent studies was small relative to the number of parameters. In this section, simulations were conducted to assess the performance of estimators in the analysis of two situations using the random effects model. The first situation was the analysis when the number of tables increased, i.e., more studies were incorporated in the meta-analysis. In the second situation, the table total of each study was changed.

4.6.1 The effect of increasing the number of studies

As discussed above, the number of independent studies was small relative to the number of parameters in this project. Simulations were performed to create 10

marginal tables of each type. In other words, the total number of tables available for analysis at each iteration increased to 30. The table total was still 300.

First of all, simulations were performed on the model with 1 random effect, i.e., the model accounting for heterogeneity among studies due to prevalence of disease. The same procedure as stated in section 4.4.1 was applied except that the DU tables, DV tables, and UV tables, were generated 10 times at each iteration. Specifically, steps 5, 8, and 11 in the simulation in section 4.4.1 were changed to the following.

- Repeat steps 3-4 ten times to obtain 10 independent marginal DU tables.
- ...
- Repeat steps 6-7 ten times to obtain 10 independent marginal UV tables.
- ...
- Repeat steps 9-10 ten times to obtain 10 independent marginal DV tables.

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.827059	0.82747	0.00041 (0.00836)	7.0e-05	0.949
C_d	0.702573	0.70266	8.32e-05 (0.00975)	9.5e-05	0.953

Table 4.23: Simulations of the model with random disease prevalence on 10 marginal tables of each type

The 30 marginal tables were applied to estimate all the parameters in the log-linear model. Results from simulations were summarized in Table 4.23. The mean of estimated variances of sensitivity and specificity were very close to the sample variances of the estimated sensitivity and specificity. The mean of estimated variance

of the random effect was 0.15. With the true variance of 0.149, the estimated variance of random effect by 30 tables was greatly improved.

Furthermore, simulations were performed on the model with two random effects. Similar to the modification above, the number of marginal tables was increased to 10 for each type. Parameter values were the same as those in section 4.5.2. The table total was 300. Results from simulations were summarized in Table 4.24. The

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.8276745	0.82716	-0.000511 (0.0091)	8.30e-05	0.95
C_d	0.705982	0.70528	-0.000703 (0.0101)	0.000105	0.96

Table 4.24: Simulations of the model with two random effects on 10 marginal tables from each type

mean estimated variances of random effects were 0.0376 and 0.00185, which were very close to the true parameter values.

4.6.2 The effect of changing the number of subjects in each study

The analysis in the previous section investigated the effect of change in the number of studies on the performance of models. In this section, the change in the study size was examined. In the d-dimer study, the smallest table total was 53. In this section, the size of each study was reduced to 50 subjects. The number of studies remained at 10 for each type. The two models with random effects were applied. Parameter values were the same as those in the previous section

Results from simulations using the model with random venography only were summarized in Table 4.25. The estimated variance of random effect was 0.143. With 10 tables of each type, the estimate of variance of random effect was still very close

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.827059	0.8268344	-0.000225 (0.0195)	0.00038	0.957
C_d	0.702573	0.7028733	0.00030 (0.0240)	0.00057	0.944

Table 4.25: Simulations of the model with random disease prevalence on 10 marginal tables from each type and table total of 50

to the true parameter value, 0.149. When the table total was reduced to 50 for each table, the magnitude of bias in estimated sensitivity and specificity was similar to the analysis with larger table total, although the standard errors were increased.

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.8276745	0.8265	-0.00116 (0.0183)	0.000335	0.98
C_d	0.705982	0.70865	0.00267 (0.0223)	0.00050	0.96

Table 4.26: Simulations of the model with two random effects on 10 marginal tables from each type and table total of 50

In the model with two random effects, results from simulations were summarized in Table 4.26. The biases in estimating sensitivity and specificity were slightly larger than those in the simulation with 300 table total. The standard errors were twice those from the simulations with table total of 300. The sample variances of estimated sensitivity and specificity were 0.000334 and 0.000498. The mean variances of the estimates were 0.00042 and 0.00060, which were larger than the sample variance. In other words, the estimated variance from the delta method may slightly overestimate the true variance. Therefore, the coverages of 95% confidence intervals were slightly larger than 0.95. Besides, the estimated variances of random effects were 0.0363 and 0.00546, respectively. These two estimates were not close to the true values. The

performance of the model with two random effects with small table total was not as good as that with large table total. In reality, however, this may not be a concern for d-dimer or other diagnostic tests where the silver standard was widely applied. With a less invasive reference, it is feasible to increase sample size to overcome this.

4.7 Simulations with other parameter values

In order to assess the generalizability of the models proposed in this project, additional simulations were performed on the two models with random effects. Parameters in the log-linear random effects model can be classified into two groups: model coefficients and distribution parameters. The distribution parameters referred to the variances of random effects. Under different clinical settings, these two groups of parameters may not be the same as those used in this project. Simulations were performed to assess the performance of the two models with random effects when these two components changed. In all the simulations, 10 tables were generated in each iteration with table total of 300.

4.7.1 Different variances of random effects

In the first scenario, different variances of random effects were applied in the simulation with the model coefficients unchanged. Two models were fitted to this scenario: the model with random venography only and the model with random venography and random interaction of d-dimer and venography.

In the model with random venography only, the true variance of random effect was reduced to 0.01, as compared with 0.149 before. Simulations were summarized

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.8271	0.82722	0.000158 (0.0081)	6.51e-05	0.956
C_d	0.7026	0.70288	0.000306 (0.0097)	9.41e-05	0.951

Table 4.27: Simulations from the model with random disease prevalence and the variance of random effect at 0.01

in Table 4.27. Results showed that biases of the estimates are very small and the 95% coverages were very close to 0.95. The mean of estimated variances was 0.0087, which was very close the true parameter value 0.01. The model worked very well when the variance of random effect was reduced 10 fold.

On the other hand, the variance of random disease prevalence was increased 10 fold, i.e., 1.5, to assess the model performance. The parameter values of model coefficients were the same as those in the simulation above. Results from simulations

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.8271	0.8257	-0.00130 (0.0081)	6.72e-05	0.952
C_d	0.7026	0.7041	0.00156 (0.0094)	9.09e-05	0.948

Table 4.28: Simulations from the model with random disease prevalence and the variance of random effect at 1.5

were summarized in Table 4.28. The biases of estimated sensitivity and specificity increased as the variance of the random effect increased 10 fold. With 10 tables of each type, the standard errors of the biases were still small. The coverage of 95% confidence intervals was very close to 0.95. The mean variance of random effect was 1.23, which was slightly lower than the true value 1.5. The results indicated that the model with random venography still performed well even when the variance of

random effect increased 10 fold.

In the model with two random effects, the true variance of random interaction was increased to 0.039, the same as the variance of random venography. In other words, the true variances of both random effects in the model were 0.039 for the simulation. Results from simulations were summarized in Table 4.29. The estimates

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.8222	0.82202	-0.000225 (0.0151)	0.000229	0.928
C_d	0.6983	0.699358	0.00110 (0.0204)	0.000416	0.922

Table 4.29: Simulations from the model with two random effects on the variances of random effects at 0.039

in this model had very small biases. The coverage of 95% confidence intervals was slightly lower than 0.95. The sample variance of estimated sensitivity was 0.0002293 and that of specificity was 0.00041523. The mean estimated variance of sensitivity from each simulation was 0.0002033 and 0.0003727 for specificity. The estimated variances were slightly smaller than the sample variance of the estimates. This was the reason why the coverage of true parameter value was slightly lower than the nominal 0.95. The mean estimated variances of random effects were 0.0394 and 0.0374, which were very close to the true parameter values 0.039 and 0.039.

4.7.2 Different model coefficients

The second scenario under consideration was that changes were made to the model coefficients but not to the variances of random effects. The same two random effects models were considered. In both models, the variances of random effects were the same as those from the estimation of the d-dimer data. In other words, the variance

of random effect in the model with random venography was given as 0.149 for simulations. Accordingly, the variances of random effects in the model with two random effects were given as 0.039 and 0.001946, respectively. In each model, 10 tables of each type with table total of 300 were generated in the simulations.

Unlike in linear models, small changes to coefficients in the log-linear model had significant impacts on the cell probabilities. In this scenario, small changes were made to the model coefficients and performance of the model with two random effects was examined by simulations. The model coefficients were changed to the following values, which were very close to the original coefficients.

β_1	β_2	β_3	β_4	β_5
-1.5	-2	-2	2.5	3

Using this set of new coefficients, the true S_d and C_d were changed to 0.731 and 0.817 based on the model with two random effects, respectively. Compared with the original S_d (0.8278) and C_d (0.706), they were quite different.

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.7310586	0.7306	-0.000462 (0.0086)	7.43e-05	0.956
C_d	0.8175745	0.8175	-8.42e-05 (0.0096)	9.20e-05	0.954

Table 4.30: Simulations from the model with random disease prevalence only on a different set of model coefficients

Using the model with random venography, results from simulations were summarized in Table 4.30. The model still provided estimates of sensitivity and specificity with very small biases. The coverage rates of 95% confidence intervals were very close to 0.95. The mean estimated variance of random effect was 0.15, which was very close to 0.149, the true variance of the random effect.

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.73065	0.7305	-1.28e-04 (0.0100)	9.98e-05	0.951
C_d	0.81726	0.8173	4.73e-06 (0.0101)	0.000101	0.956

Table 4.31: Simulations from the model with random disease prevalence and association between d-dimer and venography on a different set of model coefficients

In the model with random venography and random interaction between d-dimer and venography, the same set of model coefficients as in the model with one random effect was applied in the simulation. Results are summarized in Table 4.31. The model produced estimates of sensitivity and specificity with very small biases. The coverage rates of 95% confidence intervals were very close to 0.95. The mean estimated variances of random effects were 0.038 and 0.00188, respectively. These estimates were very close to the predefined values of variances of random effects.

Another set of model coefficients was chosen, which had the values in between the above two sets. All other parameter values were the same as those in the simulations above.

β_1	β_2	β_3	β_4	β_5
-1	-2.5	-2.7	2.4	3.5

The resulting true sensitivity and specificity of d-dimer became 0.802 and 0.73, respectively, given the same set of variances of random effects.

Results from simulations were summarized in Tables 4.32 and 4.33 for the model with one random effect and two random effects, respectively. Both models provided estimates of sensitivity and specificity of d-dimer with very small biases. The coverage rates of 95% confidence intervals were very close to 0.95. The estimated variance of random effect in the model with random venography only was 0.148, which was

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.80218	0.802614	0.00043 (0.0087)	7.74e-05	0.953
C_d	0.73106	0.7312583	0.00020 (0.0093)	8.70e-05	0.946

Table 4.32: Simulations from the model with random disease prevalence only on the third set of model coefficients

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.80186	0.80168	-0.00017 (0.0095)	9.01e-05	0.944
C_d	0.73066	0.73093	0.00027 (0.0102)	0.00010	0.947

Table 4.33: Simulations from the model with random disease prevalence and association between d-dimer and venography on the third set of model coefficients

very close to the parameter value 0.149. The estimated variances of random effects in the model with two random effects were 0.03865 and 0.00181244, which were very close to the parameter values.

The simulations above provided more evidence of the applicability of the models proposed in this project. When only the variances of random effects changed, estimates from either of the random effects model had very small biases and coverage probability close to 95%. When the change was made on the model coefficients, estimates from both models had very small biases. The coverages of sensitivity and specificity in both situations were very close to the nominal 95% level.

4.7.3 Poor silver standard

Simulations in the previous section were performed by changing model coefficients and variances of random effects. In practice, these quantities did not have direct linkage to the diagnostic performance. Using the functional relationship between

model coefficients and sensitivity and specificity, model coefficients for simulations can be derived. In diagnostic tests, the imperfect reference may have poor diagnostic characteristics. In the following, the performance of models was assessed by simulations when the sensitivity and specificity of the imperfect reference were 0.2.

The derivation of model coefficients based on predetermined values of sensitivity and specificity of d-dimer and ultrasonography was similar to that in section 4.3.2. An additional step was required when calculating the bias and 95% coverages. The true sensitivity and specificity of d-dimer should be calculated by unconditional cell probabilities. In other words, integrations over random effects should be performed to derive the parameter values of sensitivity and specificity of d-dimer in the simulation.

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.59	0.5904947	0.000495 (0.00856)	7.34e-05	0.945
C_d	0.90	0.900221	0.000221 (0.00924)	8.53e-05	0.95

Table 4.34: Simulations of the model with random disease prevalence only when sensitivity and specificity of the silver standard were both 0.2

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.5896585	0.58978	0.000117 (0.00986)	9.72e-05	0.938
C_d	0.899866	0.900038	0.000172 (0.00976)	9.51e-05	0.948

Table 4.35: Simulations of the model with random disease prevalence and association between d-dimer and venography when the sensitivity and specificity of the silver standard were both 0.2

Tables 4.34 and 4.35 summarized results of simulations. In both models, the sensitivity and specificity of ultrasonography were set at 0.2. In each simulation, 10 tables of each type of marginal tables were generated with table total 300. The

true variance of random venography was 0.149 for simulations in Table 4.34. The mean estimated variance was 0.161. The true variances of random venography and interaction between d-dimer and venography in Table 4.35 were 0.039 and 0.0019, respectively. The mean estimated variances from simulations were 0.0390853 and 0.001905018, respectively. The results showed that both models performed very well even when the sensitivity and specificity of the imperfect reference were very low.

Comparison with the unadjusted model

Besides, it was interesting to compare the performances of the model that adjusted for the difference between the two references and the unadjusted model when the silver standard had low sensitivity and specificity. Simulations were performed on models with random disease prevalence only. In the simulations, 10 tables of each type with table total of 300 were generated.

Models	unadjusted	adjusted
Bias S_d (s.e.)	-0.111 (0.0115)	0.000495 (0.00856)
Bias C_d (s.e.)	-0.295 (0.0128)	0.000221 (0.00924)
MSE S_d	0.0125	7.34e-05
MSE C_d	0.0874	8.53e-05
95% coverage S_d	0	0.945
95% coverage C_d	0	0.95

Table 4.36: Comparison between the unadjusted model and the adjusted model with random disease prevalence when sensitivity and specificity of the silver standard were 0.2

Table 4.36 showed that when the sensitivity and specificity of ultrasonography (the silver standard) were 0.2, the performance of the unadjusted model was extremely poor. The biases were much larger than those from the adjusted model. The sampling distributions of sensitivity and specificity from the two models were

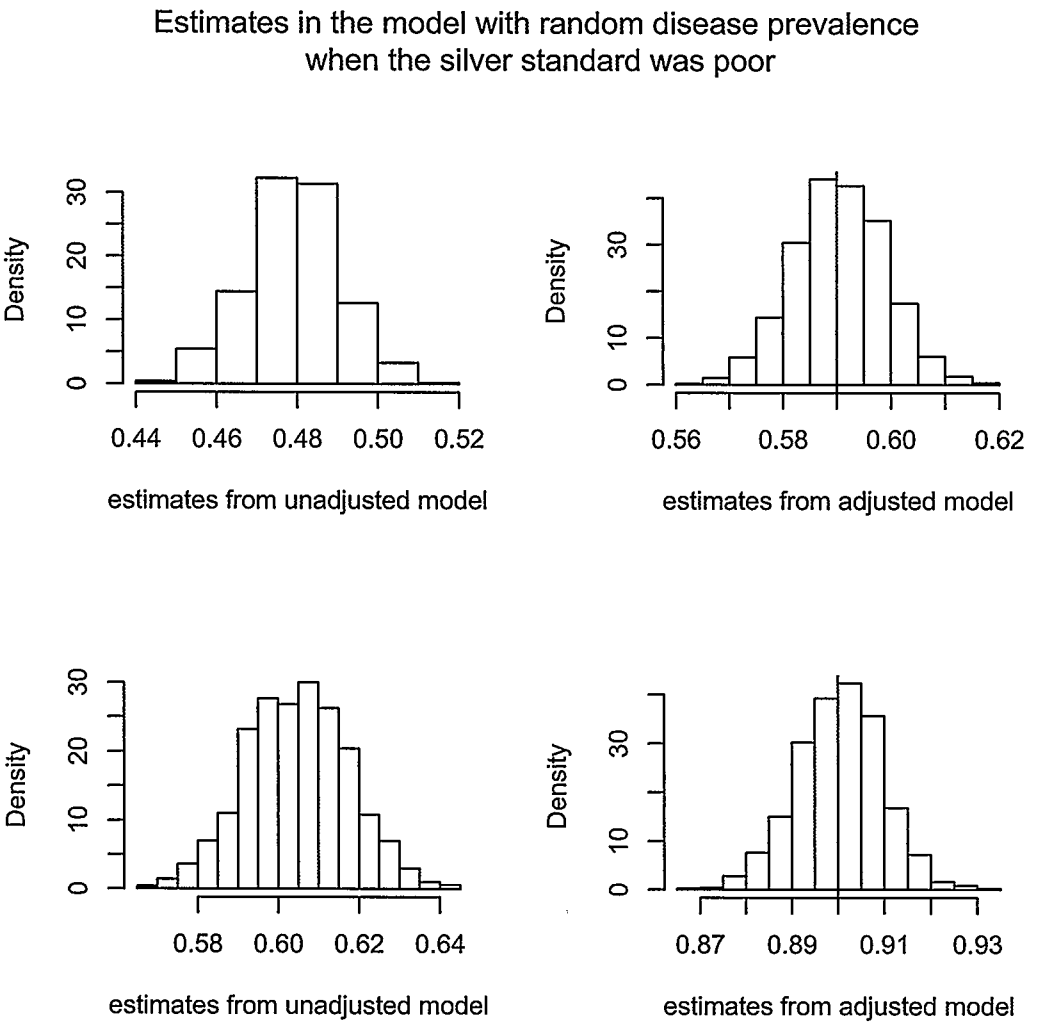

Figure 4.21: Histograms of estimated sensitivity and specificity from the unadjusted and the adjusted model with random disease prevalence when the silver standard was poor

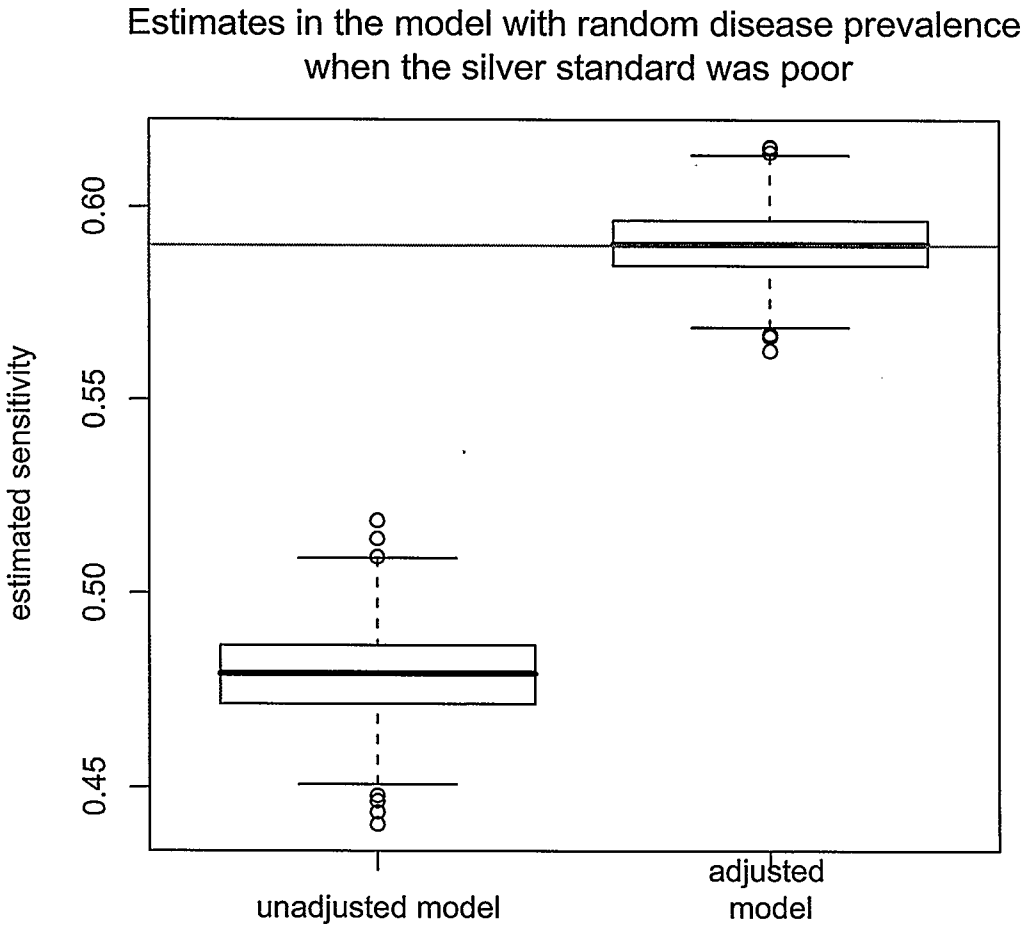

Figure 4.22: Boxplots of estimated sensitivity from the unadjusted and the adjusted model with random disease prevalence when the silver standard was poor

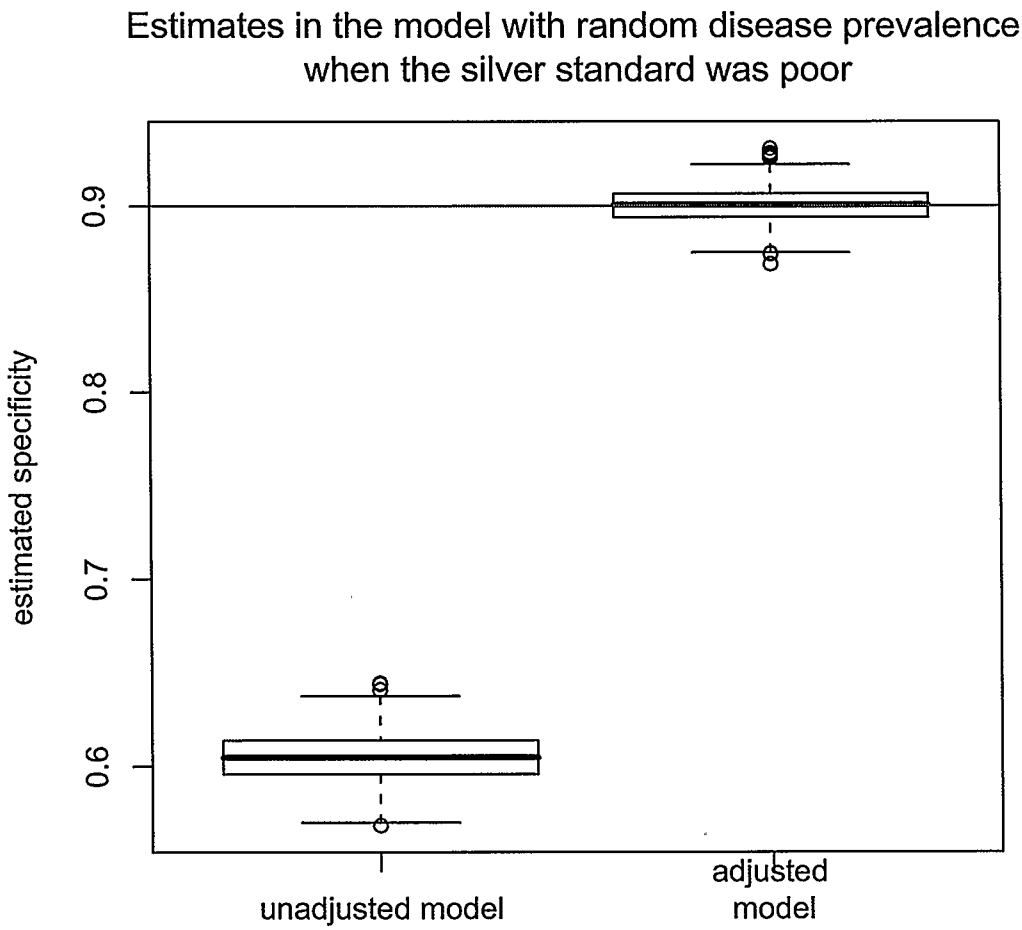

Figure 4.23: Boxplots of estimated specificity from the unadjusted and the adjusted model with random disease prevalence when the silver standard was poor

displayed in Figures 4.21, 4.22 and 4.23. When the reference test had poor diagnostic characteristics, it was clinically crucial to account for the diagnostic error from the reference test. Some researchers had shown that adjustment was important even when the sensitivity and specificity of the reference test were high [78].

Comparison with the analysis using DV tables only

When the imperfect reference was poor, it was interesting to compare the analysis using DV tables only and that using all the tables. The model with random venography only was considered. The sensitivity and specificity of d-dimer were chosen at 0.59 and 0.90 for simulations, respectively. In each iteration, 10 tables of each type were generated and the two approaches were applied. Results were summarized in

Data used	DV tables only	all tables
Bias S_d (s.e.)	-1.64e-05 (0.0107)	-8.59e-05 (0.00852)
Bias C_d (s.e.)	-0.000378 (0.00982)	-0.000398 (0.00927)
MSE S_d	0.000115	7.25e-05
MSE C_d	9.66e-05	8.60e-05
95% coverage S_d	0.96	0.949
95% coverage C_d	0.951	0.948

Table 4.37: Comparison between the analysis using DV tables only and the analysis using all tables in the model with random disease prevalence only when the sensitivity and specificity of the imperfect reference were 0.2

Table 4.37. From Table 4.37, the mean squared errors from the analysis using DV tables only were larger than those from the analysis using all tables. The relative efficiency of using the DV tables only to using all tables versus can be calculated. The relative efficiency in estimating sensitivity of d-dimer was 0.63. The relative efficiency in estimating specificity of d-dimer was 0.89. The simulations above showed that the estimations adjusting for the difference between the two references were

superior over not only the model ignoring the difference but also the analysis using the data from the gold standard only. In the first comparison, the model adjusting for the difference of reference tests reduced the bias substantially. In the latter case, the model using all the available data provided smaller mean squared errors. Removing the tables using the imperfect reference from the analysis resulted in loss of information.

4.7.4 Parameter values close to the boundary

By definition, the parameter space of sensitivity and specificity is from 0 to 1. Using the symmetric confidence interval, like the Wald-type, may result in the limits of confidence interval outside the range of 0 and 1. In the above simulations, all the limits of confidence intervals were examined. None of the lower limits exceeded 0 and none of the upper limits crossed 1. This indicated good performance of the Wald-type confidence intervals.

However, the parameter values of sensitivity and specificity of d-dimer in above simulations were 0.82 and 0.70. Worries may arise if the parameter values were close to the boundary. To examine the performance of the model when parameter values were close to the boundary, two sets of parameter values were chosen for simulations: 0.05/0.05 and 0.95/0.95. The model with random venography only and the model with two random effects were applied in the simulations. The disease prevalence was 0.7 and sensitivity and specificity of ultrasonography was 0.74 and 0.93, respectively. Ten tables of each type were generated at each iteration with table total 300. The variances of random effect were the same as those applied in previous simulations. Table 4.38 summarized results from 1000 simulations. The lower limits

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.05	0.050096	9.56e-05 (0.00434)	1.89e-05	0.944
C_d	0.05	0.050023	2.35e-05 (0.0071)	5.03e-05	0.946

Table 4.38: Simulations of the model with random disease prevalence when parameter values were 0.05

of all confidence intervals did not exceed the boundary of 0. The minimum of the lower limits of sensitivity was 0.02632 and that of specificity was 0.01761.

On the other hand, the true S_d and C_d were chosen as 0.95 for simulations to examine the performance of the model with random disease prevalence. Results

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.95	0.9499	-6.78e-05 (0.00432)	1.87e-05	0.941
C_d	0.95	0.9497	-0.000278 (0.00718)	5.16e-05	0.944

Table 4.39: Simulations of the model with random disease prevalence when parameter values were 0.95

were summarized in Table 4.39. The upper limits of all confidence intervals were examined. The maximum of the upper limits of sensitivity was 0.9723 and that of specificity was 0.9819. The model still performed well when the true parameter values were close to boundaries of the parameter space.

With respect to the model with two random effects, the parameter values were set at 0.95 for simulations. The same setting was applied as in the model with random disease prevalence only, i.e., 10 tables for each type with table total of 300. The variances of random effects were the same as the simulations before, 0.039 and 0.0019, respectively. Results from 1000 simulations were summarized under the model with

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.9498	9.5018	3.513e-04 (0.0044)	1.957e-05	0.94
C_d	0.9498	9.4972	-1.136e-04 (0.0072)	5.1392e-05	0.95

Table 4.40: Simulations of the model with random disease prevalence and association between test and gold standard when parameter values were 0.95

two random effects. The upper limits of all confidence intervals were examined. The maximum of the upper limits of sensitivity was 0.9708 and that of specificity was 0.9806. The model with two random effects maintained nice performance even when the true parameter values were close to boundaries of the parameter space.

Parameters	True values	Mean estimates	Bias (s.e.)	MSE	95% Coverage
S_d	0.05016653	5.008464e-02	-8.19e-05 (0.00433)	1.87e-05	0.954
C_d	0.05016653	5.004186e-02	-1.25e-04 (0.00715)	5.11e-05	0.951

Table 4.41: Simulations of the model with random disease prevalence and association between test and gold standard when parameter values were 0.05

Using the same model, simulations were performed on true parameter values on 0.05. Results were summarized in Table 4.41: The minimum of lower limits of sensitivity was 0.03063 and that of specificity was 0.01775.

4.8 Comparisons among the three models: fixed effects, random disease prevalence, and two random effects

The three models applied in this project were considered three approaches of the meta-analysis of d-dimer. Comparisons of the performance of these models on the

same dataset were useful because it enabled the assessment of model misclassification. On the same set of data, the information criterion was applied to compare the performances of the three models. The two commonly available criteria were Akaike's information criteria (AIC) and Bayesian information criteria (BIC). Both criteria accommodated the trade-off between the value of the log-likelihood and the number of parameters to be estimated. The BIC had additional justifications on the sample size, e.g., the number of independent clusters in the random effects model. The smaller the value of these information criteria, the better the model was. In this section, the Bayesian information criterion was applied. The log likelihood and BIC from the three models were compared.

Models	fixed effects	random venography	2 random effects
logL	-1091.396	-1087.573	-1087.561
independent samples	12	12	12
parameters	5	6	7
BIC	2188.19	2181.62	2182.68

Table 4.42: Comparison of performances of the fixed effects model and random effects models on analyzing the same dataset

In Table 4.42, the model with random venography had the lowest BIC value, which indicated that it was slightly better than the other two models. The fixed effects model had the highest BIC among the three, although the differences among the three models were not large.

In order to compare the asymptotic performances of the three models, simulations were performed. Estimates from the model with two random effects were applied as true parameter values. In other words, the model coefficients were set as the following.

$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
−0.878	−2.5206	−2.7002	2.4493	3.538

The variances of random effects were set at $\hat{\sigma}_1^2$=0.039 and $\hat{\sigma}_2^2$=0.001946.

In the simulation, 3 DU tables, 4 DV tables, and 5 UV tables with table total of 300 were generated. The three models were then applied to analyze the same dataset at each iteration. Estimates and 95% confidence intervals were obtained for each model. At the end of 1000 simulations, the mean squared errors from the three models were calculated. Table 4.43 summarized the results from simulations

Models	fixed effects	random venography	2 random effects
Bias S_d (s.e.)	-0.0010 (0.015)	-0.00126 (0.014)	-0.00116 (0.014)
Bias C_d (s.e.)	-0.0002 (0.019)	0.00061 (0.017)	0.0000766 (0.017)
MSE S_d	0.000219	0.000201	0.000201
MSE C_d	0.000366	0.000305	0.000306
95% coverage of S_d	0.918	0.933	0.948
95% coverage of C_d	0.888	0.922	0.940

Table 4.43: Simulations of the fixed effects model and random effects models when the true model was the one with two random effects

comparing the three models where the data were generated from the model with two random effects. The biases from the three models were close to each other. But the 95% coverage from the fixed effects model was the lowest. The mean estimated variances of the estimates, 0.000176 (S_d) and 0.000241 (C_d), were lower than the sampling variances of estimates, 0.000218 and 0.0003666, respectively. This indicated that the estimated variance based on the fixed effects model underestimated the true variance. Consequently, the 95% confidence intervals constructed by the fixed effects model did not have 95% coverage. Similar conclusions can be derived for the model with random disease prevalence only. The extra variation due to the association

between d-dimer and venography was 0.0019, which was relatively small. Therefore, the 95% coverages from the model with random disease prevalence only were closer to the nominal 95% level than the fixed effects model.

With respect to efficiency, the mean squared errors from the model with random venography and the model with two random effects were smaller than that from the fixed effects model. The relative efficiency of the fixed effects model over the model with random venography was 0.92 in estimating sensitivity of d-dimer and was 0.83 in estimating specificity of d-dimer. The relative efficiency of the fixed effects model over the model with two random effects was 0.92 in estimating sensitivity and was 0.84 in estimating specificity of d-dimer.

The comparisons above indicated that misspecifying the random effects model as the fixed effects model resulted in loss of coverage of 95% confidence interval and efficiency. In other words, if the fixed effects model was applied in the analysis of the data where there was truly heterogeneity, the 95% confidence intervals did not have the nominal 95% confidence and the efficiency can be lost for up to 17%.

Chapter 5

Discussion

5.1 Major findings

The results in Chapter 4 showed that ignoring the difference between the imperfect reference and the gold standard led to severely biased estimates of the diagnostic performance of d-dimer. This was consistent with conclusions in the literature [26, 73, 74, 100]. The direction of bias on diagnostic accuracy, however, was not conclusive in the literature. Some authors pointed out that inaccurate validation methods overestimated sensitivity and specificity [73, 74, 100], although some studies indicated that the direction of bias can be positive or negative [9, 26, 48]. In our study, all the biases were calculated as the mean of estimated sensitivity and specificity minus the corresponding parameter values. Under the model ignoring the difference between the two references, all the biases were negative. This indicated that using the inaccurate reference standard underestimated the diagnostic accuracy if adjustments were not made under the conditional independence assumption. In this project, all the analyses were performed under the assumption of independence between the test and the inaccurate reference. Under this assumption, the conclusion of underestimation using the inaccurate reference was consistent with that in the literature [9, 48]. When the assumption of independence was not valid, using the unadjusted model overestimated the sensitivity and specificity in the literature[9, 48], although the analysis on this situation was beyond the scope of this project.

In Chapter 4, the effect of disease prevalence on the magnitude of biases under the unadjusted model was examined. Figures showed that the magnitude of bias in sensitivity decreased as the disease prevalence increased, whereas the inaccuracy in specificity increased at the same time. In other words, when disease prevalence was high, the bias of sensitivity was small but the bias of specificity was large. When the disease prevalence was low, the bias of sensitivity was large but that of specificity was small. This finding was consistent with conclusions in the literature [22]. The estimation of sensitivity was most accurate with high disease prevalence and the estimation of specificity was most accurate with low disease prevalence. This phenomenon was observed in both the fixed effects model and the random effects model.

When the number of tables increased to 10 for each type of tables, the estimation of variance of random effects was greatly improved. On the other hand, when the table total was reduced to 50, the model still provided estimates with very small biases using 10 tables of each type. The standard errors of bias were increased (see Table 4.25). The results indicated that the model performed well even with study size as small as 50. Based on the data for analysis in this project, all studies had table total larger than 50. The model was readily applicable to studies of diagnostic tests of DVT. Furthermore, when the imperfect reference had poor diagnostic performance, i.e., low sensitivity and specificity, the model still performed well (see Table 4.34). The standard errors of bias were very close to those in the simulation with 10 tables of each type as shown in Tables 4.23.

For the model with two random effects, similar conclusions can be drawn except those on the effect of different study size. When the size of each study reduced

to 50, both the magnitude of biases and standard errors of biases increased (see Table 4.26). Compared to the simulation with table total of 300 (see Table 4.24), the magnitude of bias doubled in estimating sensitivity and that in specificity more than tripled. The standard errors doubled. Consequently, the mean squared error increased substantially when the table total reduced to 50. The impact of different study sizes on the model with two random effects was slightly larger than that on the model with random disease prevalence only. Caution should be given to the study size when applying the model with two random effects in the meta-analysis.

In general, the methods proposed by this project took into account the imperfection of the silver standard and had nice performance over various parameter settings. The estimators had very small biases and small mean squared errors. The coverage of 95% confidence intervals was very close to 0.95.

5.2 Comparisons with other approaches

The models proposed by this project were based on the log link between the outcome and the linear combination of diagnostic tests. Although logit link was commonly applied in the literature for analysis of probabilities, it is typically useful for binomial outcomes. The log-linear model is the traditional approach to analyze contingency tables [7, 89, 90, 91, 106]. The model setup was similar to that in the ANOVA-type models, which were well-known approaches for applied statistics. In particular, the log-linear model provided estimations of joint probabilities whereas the logit model provided conditional probabilities. Although both the log-linear model and the logit model provided estimates of associations, the log-linear model addressed the inter-

associations among the three tests, whereas the logit model facilitated examinations of a particular subset of associations. In this project, for example, the log-linear model specified associations among all three tests via the two-way interactions and possibly three-way interactions. If the logit model was applied, the association of interest would be the probability of one test given the rest two tests but not between the two tests in the model. The log-linear approach provided analysis of all inter-associations in one model.

In the meta-analysis of d-dimer, a different approach was attempted by researchers to analyze data in this project. In the following, the approach from Stein [68] was presented and compared with the models in this project.

5.2.1 Comparisons with the Stein paper

The data for analysis in this project were a subset of the larger dataset in the Stein paper [68]. In the paper by Stein and colleagues, the statistical methods involved a different setup due to the complexity of data. First of all, the cutoff in the paper was 500ng/mL, which was the same as one of the data extraction criteria in this project. The studies in this project, however, were restricted to those applied the SL assay and either of the references, ultrasonography or venography, but not both. Applying these criteria substantially reduced the number of studies to 12 for the analysis. More studies were available in the Stein paper, which, on the other hand, increased the complexity of analysis.

Secondly, the mixed model was applied in the Stein paper to account for heterogeneity across studies. This was similar to the approaches from this project. The mixed model was the method commonly applied in various clinical circumstances

where correlations among responses were taken into account. In Stein's paper, three random effects were considered in the analysis: assay, patient group, and study. The analysis in this project, in contrast, incorporated the prevalence of disease and the association between d-dimer and venography as random effects. As stated earlier, the data for analysis in this project were restricted to one assay, i.e., SL. Therefore, the adjustment on the heterogeneity due to assay was not applicable in this project. Besides, the prevalence of DVT was a major representation of differences among patient groups. It depended on characteristics of the patients, such as age, gender, presence of artery diseases, previous history of DVT, and so forth. Applying the prevalence of DVT as the random effect can be regarded as an approach similar to treating the patient group as the random effect in the model.

Thirdly, the analysis in the Stein's paper did not take into account the difference between the two references, ultrasonography and venography. This was the major difference between the methods in Stein's paper and the analysis in this project. The choice of collapsing tables from different references was in part due to a variety of references from different studies. The impedance plethysmography and plethysmography were applied in some studies as references instead of ultrasonography or venography. Some studies used both ultrasonography and venography as the reference. In this project, however, the difference between reference standards was taken into account, although the imperfect reference under consideration was ultrasonography only. As shown in the results from Chapter 4, accounting for the different references across studies was important.

Furthermore, the restricted maximum likelihood (REML) was applied in Stein's paper for variance estimations. The REML was a conventional approach to fixing

the problem of biased estimates of the variance in the linear mixed model [15, 30]. The approach in this project was considered the maximum likelihood estimation (MLE) and was often criticized by producing biased estimates on the variance components. Restricted maximum likelihood estimation (REML), on the other hand, provided generally unbiased variance estimates by means of an error contrast. The error contrast was a linear combination of the observed outcome such that its expected value was 0. The maximization in the REML approach was performed on the error contrast instead of the observed data vector. Harville (1977) [15] compared the two approaches, MLE and REML, and pointed out the pros and cons of each approach. He concluded that MLE generally worked for the model with small number of parameters and REML was good for the model with larger number of parameters. As the number of parameters increased, the bias from MLE increased. The model considered by Harville, however, was the random effects model with identity link function, i.e., linear mixed model. Direct generalizations of the REML approach to the logarithm link function for multinomial data were still under debate in the literature. The major concern of REML was on the bias in estimation for highly non-normal response data [4, 31, 60, 105].

In general, the major difference between the approach in the Stein paper and the models in this project was on the adjustment of the reference test. The approach in Stein's paper ignored the different references and collapsed the two types of tables. The methods proposed in this project performed appropriate adjustments for the different reference tests.

5.2.2 Comparisons with estimations using the tables between the test of interest versus the gold standard alone

In Chapter 4, simulations were performed to compare the model using DV tables only with the model using all the tables. The mean squared error from the model using DV tables alone was larger than that from the model using both DU and DV tables, Tables 4.9, 4.10, and 4.13. The relative efficiency of the model using DV tables only over the model with all tables was consistently smaller than 1, when the number of DV tables was small relative to the number of DU tables. If the number of DV tables was 10 times that of DU tables, Table 4.19, the efficiency from the two approaches was very close. The analysis using all available tables produced more efficient estimates than the approach using tables between the test and the gold standard alone. The advance in efficiency was achieved in all the analysis.

In the context of meta-analysis, studies using the imperfect reference were often excluded. In diagnostic tests, however, it was inefficient to perform this procedure because the number of studies using the gold standard may be small when the gold standard was invasive. If the gold standard test was associated with risk to patients, studies applying the silver standard may be predominant. As shown in this project, discarding these studies in the meta-analysis resulted in loss of efficiency because studies using the silver standard contributed to deriving the diagnostic characteristics of the test of interest. Furthermore, for future applications, the methods proposed in this project provided an evidence of potentially removing the gold standard in evaluating a new diagnostic test. If characteristics of the silver standard were well established in the literature, the diagnostic performance of the test of interest can be

estimated after adjusting for the imperfection of the silver standard. Therefore, the methods in this project were perceived to be solutions for estimating the diagnostic performance of the test of interest in the absence of a gold standard. Finally, it is in keeping with scientific principle to include all available evidence to derive medical conclusions. Analyses based on all well-designed studies facilitate well-powered conclusions.

5.3 Statistical issues in the analysis

5.3.1 Advantages and disadvantages of Gibbs sampling and Gaussian Hermite integration in random effects model

In the presence of random effects in the log-linear model, two algorithms were applied to obtain estimates: Gibbs sampling and Gaussian Hermite integration. In the model with random venography only, the two algorithms produced consistent estimates. This may be of interest to researchers who perform meta-analysis, in which random effects are involved. In the more complicated model with two random effects, however, estimates from the two approaches were different, especially in the variances of random effects. In this section, the advantages and disadvantages of the two algorithms are discussed.

Advantages of Gibbs sampling and Gaussian Hermite integration

The Gibbs sampling is a Bayesian approach. It requires the availability of full conditional distributions of each parameter given other parameters. The procedure is conducted by generating sequential samples from the full conditional distributions. If the conditional distributions are selected properly, it guarantees that the limit-

ing distribution of samples follows the joint posterior distribution. Gibbs sampling was developed to avoid numerical burdens of integrations and maximizations. By means of consecutive sampling from the full conditional densities, the Gibbs sampling provides a sample from the target distribution at convergence. Based on the posterior sample, estimates and standard errors can be obtained from the posterior means and standard deviations. In the present context, sensitivity and specificity of d-dimer were the parameters of interest. These two quantities are functions of cell probabilities in the three-dimensional table. At each iteration of the Gibbs sampling procedure, sensitivity and specificity of d-dimer can be updated by current values of cell probabilities. At convergence, the sequential samples of sensitivity and specificity represent their posterior target distributions. Inference on these two parameters can be derived by the posterior samples. This has advantages over the frequentist approach, in which functions of parameters have to be transformed to obtain estimates of the target parameter. Standard errors must be calculated via approximation in frequentist approaches using the delta method, whereas the posterior sample properties from the Gibbs sampling are essentially exact at convergence.

In contrast, Gaussian Hermite integration follows the conventional approach under the frequentist framework. In the presence of random effects, which are nuisance parameters, integrations of the joint distribution over random effects must be calculated in order to derive the marginal likelihood of model coefficients. The basic idea of Gaussian Hermite integration is to approximate integrals with summations. Gaussian Hermite integration applies weighted sums to approximate integrals with normal kernel. It is often used for integrations in the analysis of random effects model and is recommended for its accuracy in estimation. Accuracy increases as the

number of abscissas increases. The integrations in this project used 25 points in all the analysis. Based on the results in Chapter 4, simulations using Gaussian Hermite integrations produced estimates of sensitivity and specificity with very small biases and coverage very close to the nominal 95% level.

Disadvantages of Gibbs sampling and Gaussian Hermite integration

Despite the advantages of Gibbs sampling and Gaussian Hermite integrations discussed above, each of these two approaches had disadvantages. First of all, convergence of Gibbs sampling is always a concern. The literature does not provide completely reliable methods for the diagnosis of convergence. Conventional methods for detecting convergence use the histograms and moving-average of posterior samples. The histograms aid in detecting normality of the sample. Running-average helps to inspect stability of the posterior sample. If there is no apparent evidence of instability, one concludes stability and convergence of the posterior sample.

Gaussian Hermite integration is a frequentist approach to solve the generalized linear mixed model (GLMM). By integrating out random effects, the maximization can be performed on the marginal likelihood function. However, computational intensity from Gaussian Hermite integration increases as the number of abscissas increases. 20-point abscissas are often recommended for sufficient accuracy in random effects models. However, in our study, 20-point abscissas were not sufficient to achieve accuracy and stability in the estimation. The number of abscissas was increased to 25-point. When only one random effect was included in the log-linear model, it took 2 minutes to obtain maximum likelihood estimates on the marginal likelihood. When there were two random effects in the model, however, the maxi-

mization of the marginal likelihood took almost 5 minutes to complete. The computational burden increased substantially as the number of dimensions of integration increased. Hence, the Gaussian Hermite integration was computationally intensive when solving models with high dimensional integrals.

The Gaussian Hermite procedure in this project was non-adaptive. Several authors [36, 72] have proposed the adaptive Gaussian Hermite quadrature be used in the random effects model. The difference between adaptive and non-adaptive approaches is the center of the quadratures. The variable of integration in the non-adaptive approach centers on zero, but in the adaptive approach it centers on the posterior mode. Despite the appealing property of centering, the adaptive quadratures did not imply fewer functional evaluations [21]. When the number of quadrature points is high, e.g. 20, the log-likelihoods from adaptive and non-adaptive approaches are close to each other [21]. Laplace approximation is an alternative method to obtain estimates in the random effects model, which has been applied in the estimation for the generalized linear mixed model [96]. The 1^{st} order Gaussian Hermite quadrature is considered equivalent to the Laplace approximation [36, 72]. In general, the Laplace approximation, adaptive and non-adaptive Gaussian Hermite quadratures produce similar results.

5.3.2 Validity of the conditional independence assumption

In the log-linear model, the number of parameters for estimations depends on the assumptions. In the analysis of three-dimentional contingency table, different assumptions can be made on the log-linear model, as discussed in Chapter 1 and Chapter 2. These assumptions result in different sets of model coefficients for esti-

mations. The conditional independence between d-dimer and ultrasonography was assumed throughout this project. Clinically speaking, this assumption indicates that diagnostic results from d-dimer and ultrasonography from each subject arise from distinct sources. In other words, the error in the outcome of d-dimer is not related to the error in the diagnosis of ultrasonography, given the true disease status. In clinical practice, this indicates that the administrator for d-dimer result of a patient should not have any knowledge of the result of ultrasonography on the same patient.

If the assumption of conditional independence is violated, more fixed effect coefficients have to be added to the model. The model with all two-way interactions takes into account pair-wise associations among the three tests, although the association between each pair does not depend on the level of the third test. To reflect this, the interaction between d-dimer and ultrasonography should be added to the log-linear model. Considering the sample size issue, more tables may be required to estimate all the parameters.

The likelihood ratio test (LRT) can be conducted to compare the model with the extra coefficient and the conditional independence model on the same dataset. Twice the difference of the log likelihood values from the two models is the test statistic and it follows a Chi-square distribution with one degree of freedom. A significant test result implies that the conditional independence assumption is violated. The likelihood ratio test is appropriate to use to compare nested models. In other words, in order to apply the likelihood ratio test, the set of parameters in one model has to be a subset of the parameters in the other model. In the case where this requirement is not satisfied, the LRT cannot be applied. Information criteria (AIC or BIC), instead, are alternatives to compare performances among models, especially random

effects models.

From the clinical standpoint, the conditional independence between two diagnostic tests may not be valid. The independence between d-dimer and ultrasonography may be violated if the ultrasonography test result of a patient was known to the doctor when he assessed the d-dimer output.

5.4 Values and clinical importance

The analysis in this project applied different statistical approaches to the log-linear model using the Gaussian Hermite integration and the Gibbs sampling. This project was a novel application of these two algorithms in the analysis of incomplete contingency tables. These two algorithms were conventional frequentist and Bayesian approaches for random effects model, respectively. They produced consistent results in the model with one random effect. As convenient algorithms, they should be given more credit and attention in conducting a meta-analysis. In addition, the response in the log-linear model was the joint probability of the three tests. This was different from conventional logit models, where the response was the conditional probabilities of a particular outcome given the values of a set of predictors. The log-linear model allows investigations on the inter-association among the three tests, which is often not viable in a single logit model. Lastly, the models proposed in this project performed well even for true parameter values close to the boundary when the sample size was 300. As presented in Chapter 4, when the true sensitivity and specificity of d-dimer were both 0.05 or 0.95, the models provided estimates with very small biases and small mean squared errors. The 95% coverage of the Wald-type confidence in-

tervals was very close to the nominal 0.95 level. The Wald-type construction was the common approach for the 95% confidence interval. Although it was criticized by the possibility of limits outside the parameter space, the confidence intervals constructed in this project did not exceed the boundary of 0 or 1 for sensitivity and specificity. Even for parameter values as extreme as 0.05 or 0.95, the lower or upper limits did not fall outside the parameter space.

Besides their statistical advantages, clinical applications of the methods are promising. In the context of diagnostic tests, the likelihood ratios of a test are useful tools for clinicians to determine the disease status of a patient. As introduced in Chapter 1, the likelihood ratio is the ratio of the probability of a test result among the diseased patients over that in the healthy population. For example, the positive likelihood ratio is the ratio of the probability of test positive in the diseased population over the probability of test positive in the non-diseased individuals. The likelihood ratios are highly related to the predictive values, which provide extremely useful information in clinical practice. Given a test result, the predictive values provide the updated probability of disease or no disease, which is the major concern of patients and doctors. Furthermore, likelihood ratios are critical to update the odds of disease or non disease. By multiplying the pretest odds by likelihood ratios, the odds of disease or no disease can be updated, yielding the post-test odds. In the presence of multiple diagnostic tests, especially in a sequence, the odds of disease and no disease can be updated after each test. At the end of the diagnostic procedure, the probability of disease can be updated, incorporating the information from all tests.

The positive and negative likelihood ratios can be derived using the methods proposed in this project, since the likelihood ratios of a test are functions of cell

probabilities. According to the theory of maximum likelihood, functions of maximum likelihood estimates are maximum likelihood estimates of the corresponding functions of the parameters. Therefore, substituting the estimated cell probabilities into the diagnostic likelihood ratios yields the maximum likelihood estimates of the likelihood ratios. Using the estimated likelihood ratios, the odds of disease or non disease can then be updated. As mentioned above, the post-test odds of disease and non disease have important diagnostic indications. Furthermore, the positive and negative likelihood ratios can be summarized into one measurement, the diagnostic odds ratio. The regression model in estimating the summary ROC curves, as described in chapter 2, can then be constructed. The summary ROC curve was applied as the summary measure in the meta-analysis of diagnostic tests [57, 83]. It can also be extended by adding study-level covariates to explore the source of heterogeneity among studies [38].

5.5 Limitations

Despite the appealing properties of estimates in the models and strong clinical indications from this project, several issues should be drawn to the attention of statisticians and clinicians.

5.5.1 Statistical concerns

The analysis in this project was based on the contingency tables from the d-dimer paper [68]. In these tables, the number of people having each combination of tests was collected from each study. This was regarded as the aggregated data. In many

statistical analyses, data were typically collected on the measurement unit, rather than the aggregated level. In diagnostic tests, the measurement unit was the patient in each study. At the individual patient level, the contingency table from the first column of Table 4.1, for example, can be presented by the following structure if age and gender are covariates to be considered.

subjectID	*d − dimer*	*venography*	*age*	*gender*
1	+	+	35	*female*
2	+	+	25	*male*
...				
16	+	+	40	*male*
----	------	--------		
17	−	+	30	*male*
18	−	+	45	*female*
...				
21	−	+	38	*female*
----	------	--------		
22	+	−	28	*male*
23	+	−	33	*female*
...				
25	+	−	43	*male*
----	------	--------		
26	−	−	36	*female*
27	−	−	41	*female*
...				
53	−	−	50	*male*

If the results from each subject were available, the individual level data can be applied in the analysis. The model for individual level data was generally different from those presented in this project, especially in the random effects model. The random effects in the individual level data should correspond to heterogeneity among patients instead of study level differences. Patient specific characteristics can be considered, such as previous history of DVT, physical examinations, relevant diagnostic tests, and the development of clinical prediction rules for each patient. These variables can be included as additional columns in the data structure above. Analysis adjusting for these factors may provide more insights to the diagnostic performance of d-dimer with regard to patient characteristics.

The models in this project provided satisfactory estimations and confidence intervals. The data for analysis, however, were presented in the aggregated level instead of the individual subject level. In meta-analysis, it may be the gold standard to use individual subject level data [1, 43]. Analysis using individual level data is considered to have higher power than analysis using aggregated data. Besides, investigations of the relationship between patient characteristics and treatment effects generally require the individual subject level data [33, 49, 65]. Nevertheless, the individual level data may not be available for meta-analysis. It depends on many issues, such as the administration of each study, the confidentiality of data, cost and time to extract data, and so forth. When the individual level data are not available, analysis using the aggregated data is the only choice. In the literature of meta-analysis, both types of data were applied [25, 65]. In diagnostic tests, however, the majority of published meta-analyses continued to be based on aggregated data [1, 25, 37]. The analyses based on aggregated data continue to be the mainstay of systematic reviews

conducted by many professional societies.

Besides the concern on the data format, different distributions for the random effects may be considered. Although the normal density was common for generalized linear mixed models (GLMM), different approaches were attempted in the analysis of GLMM [31, 41]. The maximum likelihood estimation was often criticized by providing poor estimates of the variance of the random effects. Despite this criticism, studies showed that fixed effects estimates in GLMM using the maximum likelihood approach was robust to misspecifications of the distribution of random effects [31, 41, 95].

In addition, results from simulations in Chapter 4 provided evidence of good performance of the models. The coverages of 95% confidence intervals were very close to the nominal 0.95 level in the model with random disease prevalence. The coverages in the model with two random effects were also close to 0.95 except when the table total dropped to 50 (see Table 4.26), or when the variance of the random interaction between d-dimer and venography increased to 0.039 (see Table 4.29). As discussed in Chapter 4, the estimated variances in these situations were slightly different from the true variance of the estimates, which was the reason for slight departure from the nominal 0.95 coverage. Other approaches of estimating the variance may be considered. For example, the "sandwich" estimator of variance may be applied. It is constructed by the score vector and hessian matrix, which are based on the likelihood. The "sandwich" estimator is a robust estimate of variance. In order to derive the robust estimator for the variances of sensitivity and specificity, the likelihood should be re-parameterized with respect to sensitivity and specificity. This step may be challenging given the complexity of the data in this project.

Finally, the analysis in this project assumed that all the three types of marginal tables were available, especially in the random effects models. In the clinical setting, however, the test of interest may not be evaluated against the gold standard at all. In other words, there would not be any tables between the test and the gold standard. Analyzing this type of data will be an additional step that is necessary to convince clinicians to reduce or remove the use of the gold standard in evaluating a new test. In the literature, several authors had attempted to provide solutions to the analysis of this type of data [46, 24, 63]. These approaches were effective under various strong assumptions, which were often not practical in clinical settings. Meta-analysis of diagnostic studies in the absence of data between the test of interest and the gold standard may be the next challenge for future research.

5.5.2 Clinical concerns

One potential implication of this project is that extensive use of the gold standard may not be necessary for evaluating a new test. Studies of d-dimer and the silver standard provide useful information on the diagnostic characteristics of d-dimer if proper adjustment is applied. This conclusion is statistically sensible. Clinicians, however, may be skeptical if the gold standard is not used for evaluation.

In the literature of diagnostic tests, the role of a new test can be classified into three types: replacement, early filtering, and post-testing [57, 71]. The role of d-dimer in diagnosing DVT is early filtering [71]. Only patients with a particular result on d-dimer continue the testing pathway. The diagnostic characteristics of d-dimer, however, are not conclusive. There are different assays of d-dimer, each with different diagnostic properties. The range of sensitivities and specificities is wide and depends

on the cutoff value [68]. The specificity may be lower than 0.5 when the cutoff is 500 ng/mL but as high as 0.78 when the cutoff is 1000 ng/mL [68]. Based on these issues, analysis accounting for different assays of d-dimer may be helpful for the diagnostic procedure using d-dimer.

5.5.3 Future research

The methods in this project show promise in the meta-analysis of diagnostic tests. The results, however, were based on certain clinical and statistical assumptions. It may be helpful to consider random effects with non-Gaussian densities. In the analysis of diagnostic data, the Beta distribution has been applied by several authors [46, 63]. Full Bayesian analysis can be performed by specifying prior knowledge of the characteristics of d-dimer. The prior knowledge can be derived from other studies in the literature. An example was presented by Gustafson [64] to match the mean and standard deviations to parameters of the Beta prior. Properties of estimates based on different distributions of random effects can be derived and compared with those in this project.

As stated in the previous section, applications of the methods presented in this project can be extended to the situation where information between the test and the gold standard is entirely unknown. In this circumstance, the maximum likelihood estimates can be derived by the procedures similar to those provided in this project. With respect to the meta-analysis of the test of interest, however, adjustments for heterogeneities across studies may be challenging. In this project, heterogeneities across studies in disease prevalence and the association between the test and the gold standard were assumed. If the gold standard is not applied in any study, the

disease prevalence may not be estimable. Furthermore, the validity or estimation of the association between the test and gold standard may be questionable if associated tables are not collected. Alternative representations of these two random effects should be of concern for future research when the information on the gold standard test was not collected.

5.6 Accessibility to user community

Various statistical packages have been applied in medical research. As the random effects model has become important in clinical research, estimation procedures for the random effects model have been well established in widely used software programs, such as SAS, STATA, Minitab, SPSS. The built-in functions in these programs, however, were based on the likelihood of complete data. When the likelihood functions did not fit into conventional densities, none of these programs provide direct approaches to the estimation, not even a feasible approach.

The statistical package R used throughout this project is an object-oriented programming software. Unlike other statistical software, the R language does not have built-in menus for statistical analysis. Although the lack of convenient menus may prevent the widespread use of R for clinicians, its programming nature provided a much more flexible and powerful implementations of statistical analysis. In this project, for example, the problems to be solved did not directly fit into any existing statistical models. In fact, this project was a novel application of advance statistical algorithms to the meta-analysis of diagnostic tests with incomplete data. In such a circumstance, the software R may be the only choice. In recent decades, statistical

packages for medical research included more and more functions of programming, which had been shown to be more flexible and powerful than user-friendly menus.

For Bayesian analyses, the software Bayesian analysis using Gibbs sampling (BUGS) has been popular in the statistical community. BUGS provides a variety of applications and generates nice graphs of posterior samples. In BUGS, the implementation of the Gibbs sampling under conventional likelihoods, such as normal, beta, poisson, are straightforward, and diagnosis of convergence of posterior samples is possible. The programming function in BUGS, on the other hand, can be challenging. It is not as flexible as R or S-plus. Most of the applications use built-in functions in BUGS. Specifying a non-standard likelihood or user-defined function may be challenging. Researchers interested in Bayesian analysis are increasingly attracted to R because of the ease of coding algorithms to sample from posterior distributions. Besides, the significant number of packages contributed to archives is available, which provides tools for Bayesian inference. The coda package, for example, provides convergence diagnosis of the posterior sample from Markov Chain Monte Carlo.

The R programs in this project can be classified into two groups, frequentist and Bayesian programs. The frequentist programs implemented algorithms in all the models, i.e., the Newton-Raphson algorithm and the Gaussian Hermit Integration within the Newton-Raphson algorithm. The Bayesian subset of the program referred only to the Gibbs sampling algorithm for random effects models. For the convenience of end users, all the programs will be compacted into functions in R. Detailed instructions of these functions will be submitted for publication to the Comprehensive R Archive Network (CRAN) and available to user community.

5.7 Summary

In the development and assessment of a new diagnostic test, the ideal situation is that a true gold standard exists and can be applied to every patient. One could hope that such a test be completely safe, inexpensive to apply, convenient to operate in any lab condition, and most importantly, 100% accurate. In reality, however, such a perfect test usually does not exist. Most of the time, the best candidate available is the one with extremely high accuracy. In some instances, however, drawbacks of the best test hinder its widespread application. These may include the invasiveness and cost of the test, patient conditions, and lab environment, amongst others. Taking into account these disadvantages, a secondary reference test can be applied, which overcomes most drawbacks of the best test and whose accuracy has been well established.

The results and analysis proposed by this project provided potential evidence of using the silver standard in place of the gold standard. Despite the absence of complete tables of the three tests, the diagnostic characteristics of the new test are still estimable. In particular, the fixed effects model showed that using tables between the new test and the silver standard only, diagnostic performance of the new test can be estimated after proper adjustments. This conclusion is of extremely high clinical value because in many diseases, diagnostic tools with high accuracies are often harmful to the patient. This project indicates that given the existing data, one may not need to include the gold standard in assessing a new test.

Bibliography

[1] Sutton AJ, Kendrick D, and Coupland CAC. Meta-analysis of individual- and aggregate-level data. *Statistics in Medicine*, 27:651–669, 2008.

[2] Dempster AP, Lair NM, and Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Ser. B*, 39:1–38, 1977.

[3] Kosinski AS and Flanders WD. Evaluating the exposure and disease relationship with adjustment for different types of exposure misclassification: A regression approach. *Statistics in Medicine*, 18:2795–2808, 1999.

[4] Engel B. A simple illustration of the failure of pql, irreml and aphl as approximate ml methods for mixed models for binary data. *Biometrical Journal*, 40:141–54, 1998.

[5] Biggerstaff BJ, Tweedie RL, and Mengersen KL. Passive smoking in the workplace: Classical and bayesian meta-analyses. *International Archives of Occupational and Environmental Health*, 66:269–277, 1994.

[6] Mustafa BO, Rathbun SW, Whitsett TL, and Raskob GE. Sensitivity and specificity of ultrasonography in the diagnosis of upper extremity deep vein thrombosis. *Archives of Internal Medicine*, 162:401–404, 2002.

[7] Everitt BS. *The Analysis of contingency tables.* Chapman and Hall, 1997.

[8] Gatsonis C and Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: Methodologic primer. *American Journal of Roentgenology*, 187:271–281, 2006.

[9] Phelps CE and Hutson A. Estimating diagnostic test accuracy using a "fuzzy gold standard". *Medical Decision Making*, 15:44–57, 1995.

[10] Rutter CM and Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*, 20:2865–2884, 2001.

[11] Morris CN and Normand SL. *Bayesian Statistics*. Oxford University Press, 1990.

[12] Ireland CT and Kullback S. Contingency tables with given marginals. *Biometrika*, 55:179–188, 1968.

[13] Hedeker D and Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50:933–44, 1994.

[14] Rindskopf D and Rindskopf W. The value of latent class analysis in medical diagnosis. *Statistics in Medicine*, 5:21–27, 1986.

[15] Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72:320–338, 1977.

[16] Harville DA and Mee RW. A mixed-model procedure for analysing ordered categorical data. *Biometrics*, 40:393–408, 1984.

[17] Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Renie D, Moher D, Becher BJ, Sipe TA, and Thacher SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. meta-analysis of observational studies in epidemiology (moose) group. *Journal of the American Medical Association*, 283:2008–2012, 2000.

[18] Becker DM, Phillbrick JT, Bachhuber TL, and Humphries JE. D-dimer testing and acute venous thromboembolism. a shortcut to accurate diagnosis? *Archives of Internal Medicine*, 156:939–46, 1996.

[19] d'Othee BJ, Siebert U, Cury R, Jadvar H, Dunn EJ, and Hoffmann U. A systematic review on diagnostic accuracy of ct-based detection of significant coronary artery disease. *European Journal of Radiology*, 65:449–461, 2008.

[20] Brown DT. A note on approximations to discrete probability distributions. *Information and Control*, 2:386–392, 1959.

[21] Lesaffre E and Spiessens B. On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics*, 50:325–335, 2001.

[22] Boyko EJ, Alderman BW, and Baron AE. Reference test errors bias the evaluation of diagnostic tests for ischemic heart disease. *Journal of General Internal Medicine*, 3:476–81, 1988.

[23] Koch GG, Landis JR, Freeman JL, Freeman DH, and Lehnen RG. A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, 33:133–158, 1977.

[24] De Bock GH, Houwing-Duistermaat JJ, Springer MP, Kievit J, and van Houwelingen JC. Sensitivity and specificity of diagnostic tests in acute maxillary sinusitis determined by maximum likelihood in the absence of an external standard. *Journal of Clinical Epidemiology*, 47:1343–1352, 1994.

[25] Lyman GH and Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. *BioMedCentral Medical Research Methodology*, 5:14:1–7, 2005.

[26] Arana GW, Zarzar MN, and Baker E. The effect of diagnostic methodology on the sensitivity of the trh stimulation test for depression: a literature review. *Biological Psychiatry*, 28:733–7, 1990.

[27] Bounameaux H, Cirafici P, and DeMoerloose P. Measurement of d-dimer in plasma as diagnostic aid in suspected pulmonary embolism. *Lancet*, 337:196–200, 1991.

[28] Brenner H. Use and limitations of dual measurements in correcting for non-differential exposure misclassification. *Epidemiology*, 3:216–222, 1992.

[29] Goldstein H. Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, 78:45–51, 1991.

[30] Patterson HD and Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 26:545–554, 1971.

[31] Hartzel J, Agresti A, and Caffo B. Multinomial logit random effects models. *Statistical Modelling*, 1:81–102, 2001.

[32] Jansen J. On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics*, 39:74–85, 1990.

[33] Lau J, Ioannidis JP, and Schmid CH. Summing up evidence: one answer is not always enough. *Lancet*, 351:123–7, 1998.

[34] Hanley JA. Receiver operating characteristic (roc) methodology: the state of the art, critical reviews. *Diagnostic Imaging*, 29:307–335, 1989.

[35] Carlin JB. Meta-analysis for 2x2 tables: a bayesian approach. *Statistics in Medicine*, 11:141–158, 1992.

[36] Pinheiro JC and Bates DM. Approximations to the log-likelihood function in the non-linear mixed-effects model. *Journal of Computational and Graphic Statistics*, 4:12–35, 1995.

[37] Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, and van der Meulen JH et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*, 282:1061–6, 1999.

[38] Lijmer JG, Bossuyt PMM, and Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Statistics in Medicine*, 21:1525–1537, 2002.

[39] Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal*, 323:157–162, 2001.

[40] Michiels JJ, Kasbergen H, Oudega R, Van Der Graaf F, De Maeseneer M, Van Der Planken M, and Schroyens W. Exclusion and diagnosis of deep vein throm-

bosis in outpatients by sequential noninvasive tools. *International Angiology*, 21:9–19, 2002.

[41] Neuhaus JM, Hauck WW, and Kalbfleisch JD. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79:755–62, 1992.

[42] Kardaun JWPF and Kardaun OJWF. Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation. *Methods of Information in Medicine*, 29:12–22, 1990.

[43] Broeze KA, Opmeer BC, Bachmann LM, Broekmans FJ, Bossuyt PMM, Coppus SFPJ, Johnson NP, Khan KS, and et al. Individual patient data meta-analysis of diagnostic and prognostic studies in obsetetrics, gynaecology and reproductive medicine. *BioMedCentral Medical Research Methodology*, 9:22:1–11, 2009.

[44] Rothman KJ and Greenland S. *Modern Epidemiology.* Lippincott Williams and Wilkins: Philadelphia, 1998.

[45] Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, and Mosteller F. Guidelines for meta-analysis evaluating diagnostic tests. *Annals of Internal Medicine*, 120:667 – 676, 1994.

[46] Joseph L, Gyorkos TW, and Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology*, 141:263–272, 1995.

[47] Moses L, Shapiro D, and Littenberg B. Combining independent studies in diagnostic test into a summary roc curve: Data-analytic approaches and some additional considerations. *Statistics in Medicine*, 12:1293–1316, 1993.

[48] Thibodeau L. Evaluating diagnostic tests. *Biometrics*, 37:801–4, 1981.

[49] Stewart LA and Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Statistics in Medicine*, 14:2057–79, 1995.

[50] Liu LC and Hedeker D. A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics*, 62:261–268, 2006.

[51] Abramowitz M and Stegun IA. *Handbook of Mathematical Functions.* National Bureau of Standards, 1972.

[52] Cheitlin MD, Alpert JS, and Armstrong WF. Acc/aha guidelines for the clinical application of echocardiography: a report of the american college of cardiology/american heart association task force on practice guidelines (committee on clinical application of echocardiography): developed in collaboration with the american society of echocardiography. *Circulation*, 95:1686–744, 1997.

[53] Kertai MD, Boersma E, Heijenbrok-Kal MH, Hunink MGM, Ltalien GJ, Roelandt JRTC, van Urk H, and Poldermans D. A meta-analysis comparing the prognostic accuracy of six diagnostic tests for predicting perioperative cardiac risk in patients undergoing major vascular surgery. *Heart*, 89:1327–1334, 2003.

[54] Daniels MJ and Gatsonis C. Hierarchical polytomous regression models with

applications to health services research. *Statistics in Medicine*, 16:2311–2325, 1997.

[55] Morrissey MJ and Spiegelman D. Matrix methods for estimating odds ratios with misclassified exposure data: Extensions and comparisons. *Biometrics*, 55:338–344, 1999.

[56] Cowles MK and Carlin BP. Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91:883–904, 1996.

[57] Leeflang MMG, Deeks JJ, Gatsonis C, and Bossuyt PMM. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine*, 149:889–897, 2008.

[58] Siadaty MS and Shu JF. Proportional odds ratio model for comparison of diagnostic tests in meta-analysis. *BioMedCentral Medical Research Methodology*, 4:27:1–13, 2004.

[59] Siadaty MS, Philbrick JT, Heim SW, and Schectman JM. Repeated-measures modeling improved comparison of diagnostic tests in meta-analysis of dependent studies. *Journal of Clinical Epidemiology*, 57:698–710, 2004.

[60] Breslow NE and Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25, 1993.

[61] Laird NM and Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.

[62] Coursaget P, Yvonnet B, and Gilks WR. Scheduling of revaccinations against hepatitis b virus. *Lancet*, 337:1180–1183, 1991.

[63] Gustafson P, Le ND, and Saskin R. Case-control analysis with partial knowledge of exposure misclassification probabilities. *Biometrics*, 57:598–609, 2001.

[64] Gustafson P and Greenland S. Curious phenomena in bayesian adjustment for exposure misclassification. *Statistics in Medicine*, 25:87–103, 2006.

[65] Lambert P, Sutton AJ, Abrams KR, and Jones DR. A comparison of summary patient level covariates in meta-regression with individual patient data meta-analyses. *Journal of Clinical Epidemiology*, 55:86–94, 2002.

[66] McCullagh P and Nelder JA. *Generalized Linear Models.* Chapman and Hall London, second edition, 1989.

[67] Whiting P, Rutjes ASW, Reitsma JB, Glas AS, Bossuyt PMM, and Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy. *Annals of Internal Medicine*, 140:189–202, 2004.

[68] Stein PD, Hull RD, Patel KC, Olsen RE, Ghali WA, Brant R, Biel RK, Bharadia V, and Kalra NK. D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism. *Annals of Internal Medicine*, 140:589–602, 2004.

[69] Davis PJ and Rabinowitz P. *Methods of Numerical Integration.* Academic Press Inc London, 1975.

[70] Lee PM. *Bayesian statistics: an introduction.* Halsted Press New York, 3rd edition, 1992.

[71] Bossuyt PMM, Irwig L, Craig J, and Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathway. *BMJ*, 332:1089–92, 2006.

[72] Liu Q and Pierce DA. A note on gauss-hermite quadrature. *Biometrika*, 81:624–629, 1994.

[73] Detrano R, Lyons KP, Marcondes G, Abbassi N, Froelicher VF, and Janosi A. Methodologic problems in exercise testing research. are we solving them? *Archives of Internal Medicine*, 148:1289–95, 1988.

[74] Detrano R, Gianrossi R, Mulvihill D, Lehmann K, Dubach P, and Colombo A. Exercise-induced st segment depression in the diagnosis of multivessel coronary disease: a meta analysis. *Journal of American College Cardiology*, 14:1501–8, 1989.

[75] Jaeschke R., Guyatt G.H., and Sackett D.I. for the evidence-based medicine working group users' guides to the medical literature. vi. how to use an article about a diagnostic test. b: What are the results and will they help me in caring for my patient? *Journal of American Medical Association*, 271:703–707, 1994.

[76] Gibbons RD and Hedeker D. Random effects probit and logistic regression models for three-level data. *Biometrics*, 53:1527–1537, 1997.

[77] Hull RD and Pineo G. *Disorders of Thrombosis.* WB Saunders Company Philadelphia, 1996.

[78] Marshall RJ. Validation study methods for estimating proportions and odds ratios with misclassified data. *Journal of Clinical Epidemiology*, 43:941–947,

1990.

[79] Morris RK, Khan KS, Robson SC Coomarasamy A, and Kleijnen J. The value of predicting restriction of fetal growth and compromise of its wellbeing: Systematic quantitative overviews (meta-analysis) of test accuracy literature. *BioMedCentral Pregnancy and Childbirth*, 7:3:1–7, 2007.

[80] Chen S, Watson P, and Parmigiani G. Accuracy of msi testing in predicting germline mutations of msh2 and mlh1: a case study in bayesian meta-analysis of diagnostic tests without a gold standard. *Biostatistics*, 6:450–464, 2005.

[81] Geman S and Geman D. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[82] Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. *Statistics in Medicine*, 7:745–757, 1988.

[83] Mallett S, Deeks J, Halligan S, Hopewell S, Cornelius V, and Altman DG. Systematic reviews of diagnostic tests in cancer: review of methods and reporting. *British Medical Journal*, 333:413–419, 2006.

[84] Richardson S and Gilks WR. A bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology*, 138:430–442, 1993.

[85] Walter SD. Use of dual responses to increase the validity of case-control studies: a commentary. *Journal of Chronic Diseases*, 37:137–139, 1984.

[86] Walter SD. Properties of the summary receiver operating characteristic (sroc) curve for diagnostic test data. *Statistics in Medicine*, 21:1237–56, 2002.

[87] Walter SD, Irwig L, and Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. *Journal of Clinical Epidemiology*, 52:943–951, 1999.

[88] Walter SD and Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *Journal of Clinical Epidemiology*, 41:923–37, 1988.

[89] Fienberg SE. The analysis of multidimensional contingency tables. *Ecology*, 51:419–433, 1970.

[90] Fienberg SE. *The Analysis of Cross-Classified Categorical Data.* The MIT Press, 1980.

[91] Fienberg SE. Contingency tables and log-linear models: Basic results and new developments. *Journal of the American Statistical Association*, 95:643–647, 2000.

[92] Thompson SG. Systematic review: Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal*, 309:1351–1355, 1994.

[93] Press SJ. *Bayesian statistics: principles, models, and applications.* John Wiley Sons Inc New York, 1989.

[94] Zeger SL and Karim MR. Generalized linear models with random effects: A gibbs sampling approach. *Journal of American Statistical Association*, 66,

1991.

[95] Butler SM and Louis TA. Random effects models with nonparametric priors. *Statistics in Medicine*, 11:1981–2000, 1992.

[96] Raudenbush SW, Yang ML, and Yosef M. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation. *Journal of Computational and Graphical Statistics*, 9:141–157, 2000.

[97] Larsen TB, Stoffersen E, Christensen CS, and Laursen B. Validity of d-dimer tests in the diagnosis of deep vein thrombosis: a prospective comparative study of three quantitative assays. *Journal of Internal Medicine*, 252:36–40, 2002.

[98] Miller TD, Roger VL, Milavetz JJ, Hopfenspirger MR, Milavetz DL, Hodge DO, and Gibbons RJ. Assessment of the exercise electrocardiogram in women versus men using tomographic myocardial perfusion imaging as the reference standard. *American Journal of Cardiology*, 87:868–873, 2001.

[99] Dukic V and Gatsonis G. Meta-analysis of diagnostic test accuracy assessment studies with varying number of thresholds. *Biometrics*, 59:936–946, 2003.

[100] van Rijkom HM and Verdonschot EH. Factors involved in validity measurements of diagnostic tests for approximal caries–a meta-analysis. *Caries Research*, 29:364–70, 1995.

[101] Flanders WD, Drews CD, and Kosinski AS. Methodology to correct for differential misclassification. *Epidemiology*, 6:152–156, 1995.

[102] Gilks WR, Clayton DG, and Spiegelhalter DJ. Modelling complexity: applications of gibbs sampling in medicine. *Journal of Royal Statistical Society Series B*, 1:39–52, 1993.

[103] Kim WY, Danias PG, Stuber M, Flamm SD, Plein S, and Nagel E. Coronary magnetic resonance angiography for the detection of coronary stenosis. *New England Journal of Medicine*, 345:1863 – 1869, 2001.

[104] Liu X and Liang KY. Adjustment for non-differential misclassification error in the generalized linear model. *Statistics in Medicine*, 10:1197–1211, 1991.

[105] Lee Y and Nelder JA. Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B, Methodological*, 58:619–56, 1996.

[106] Bishop YMM, Fienberg SE, , and Holland PW. *Discrete Multivariate Analysis: Theory and Practice.* The MIT Press, 1975.

[107] Chen Z and Kuo L. A note on the estimation of the multinomial logit model with random effects. *The American Statistician*, 55:89–95, 2001.