Meta-Analysis of Multi-Arm Trials Using Binomial Approach

Hathaikan Chootrakool, Pichet Treewai

Department of Biostatistics, Faculty of Public Health, Mahidol University, Bangkok, Thailand
Email: hathaikan.cho@mahidol.ac.th

Abstract

Most meta-analysis has concentrated on combining of treatment effect measures based on comparisons of two treatments. Meta-analysis of multi-arm trials is a key component of submission to summarize evidence from all possible studies. In this paper, an exact binomial model is proposed by using logistic regression model to compare different treatment in multi-arm trials. Two approaches such as unconditional maximum likelihood and conditional maximum likelihood have been determined and compared for the logistic regression model. The proposed models are performed using the data from 27 randomized clinical trials (RCTs) which determine the efficacy of antiplatelet therapy in reduction venous thrombosis and pulmonary embolism.

Keywords

Meta-Analysis, Multi-Arm Trials, Binomial Approach

1. Introduction

There has been a massive growth in the number of randomized clinical trials (RCTs) since the first RCT was introduced in the well-known streptomycin trial in 1946 [1]. In making some of this information more readily available, an attempt is made to pull together the existing evidence in a form that can be used by researchers or statisticians; this is called *systematic review*. The aim of systematic reviews is to identify and summarize the findings of all possible studies addressing the clinical question of the review. Systematic review reduces the large quantities of information to a manageable size. An important aspect of most reviews is the quantitative synthesis of results; thus, *meta-analysis* is the statistical part of systematic review. The main purpose of meta-analysis is to increase the precision of the conclusions of a review. With statistical perspec-

tive, it is able to detect treatment effects with greater power and estimate these effects with greater precision than any single study. A meta-analysis from systematic reviews of Antiplatelet Trialists' Collaboration [2] was applied in the paper.

Most meta-analysis has focused on summarizing treatment effect measures based on the comparison of two treatments or called *arms*, sometimes also called *interventions* or *exposures*. In this comparison, two groups of individual studies are exposed to two different treatments. Standard two-arm RCTs are frequently used in clinical research due in part to its relative simplicity of design and interpretation. At its most basic, one power, one significance level and one magnitude of difference are analyzed for two-arm comparisons. Conclusions are straightforward: either the two arms are shown to be different or they are not. The implementation for the model is not complicated. When more than two arms are included in meta-analysis, complexity ensues. These types of dataset are called *multi-arm trials* [3] [4] [5] although some authors call it *mixed treatment comparison* (MTC) [6] [7], or *multiple treatment comparison* [4] [8]. Additionally, some authors call network *meta-analysis* for indirect head-to-head evidence for multiple evidences from RCTs [9]-[14].

For multi-arm trials data, Chootrakool and Shi [3] introduced the normal approximation model using an empirical logistic transform requiring a large number of individual observations and the probability of being the case to be not too near zero or one. If the number of individual observations is small, the normal approximation model is not suitable. The number of samples in a single study should usually be larger than 20 [4]. In this paper, an exact binomial model is introduced to fit the binary multi-arm trials data. There are two alternative maximum likelihood approaches that can be used to make inferences for the unknown parameters in the logistic regression model. These are unconditional method and conditional method. The logistic regression model has become increasingly popular with the easy availability of appropriate computer routines. Many authors have described maximum likelihood estimation procedures which turn out to be iterative [15]. There have been a large number of studies about unconditional and conditional methods, for example in Tritchler [16], Sartori [17], Lee *et al.* [18] and Caterina *et al.* [19].

Most existing methods for meta-analysis of multi arm trials use the logistic regression model with the unconditional approach. Lu and Ades [20] introduced the Bayesian hierarchical model for multi-arm trials using the unconditional method to estimate unknown parameters. Weber *et al.* [21] used random effect model using the unconditional maximum likelihood method for the meta-analysis. They conducted a simulation study comparing two zero-cell corrections under the ordinary random-effects model. Logistic regression model is introduced using the conditional approach in the paper.

There has the direct and indirect comparison for RCTs in a meta-analysis [9] [20], this has led to network meta-analysis. Seide *et al.* [12] perform simulation

study for sparse network of trial including multi-arm trials. They compared Bayesian and Frequentist methods in random effects network meta-analysis. Jenkins *et al.* [13] introduced methods for inclusion of evidence in network meta-analysis. More applications of network meta-analysis can be found in Greco *et al.* [22], Wang [23] and Zhao *et al.* [14].

The structure of the paper is arranged as follows. The data structure of multi-arm trials is firstly introduced in Section 2. Fitting the logistic regression model is proposed in the next section. Unconditional maximum likelihood approach for the model including the standard error of maximum likelihood estimators (MLEs) are described in Section 4. Similarly, conditional maximum likelihood approach for the model is presented in Section 5. In Section 6, the logistic regression model is illustrated with the unconditional and conditional approaches with the data. The final section is discussion and conclusion including the advantages and the limitations of the two approaches.

2. The Data Structure of Multi-Arm Trials

Let the indices $i = 1, \cdots, M$ and $j = 0, 1, \cdots, K$ stand for the studies and the treatments respectively, where the index $j = 0$ stands for the control group. To make multi-arm comparisons, suppose that there is M RCTs comparing $K + 1$ treatments. For the ith study, let r_{ij} represent the number of cases on treatment j and let n_{ij} be the total number in the corresponding group. Let π_{ij} be the probability of being the case for a patient received the treatment j in the ith study. The r_{ij} has a binomial distribution

$$r_{ij} \sim Bin\left(\pi_{ij}, n_{ij}\right); \ i = 1, \cdots, M; \ \text{and} \ j = 0, 1, \cdots, K.$$

A data structure of multi-arm trials shall be defined by introducing an index set J_i comprising the treatments involved in the ith study. The data structure is shown as

$$\mathcal{D} = \left\{\left(r_{ij}, n_{ij}\right) : i = 1, \cdots, M; j \in J_i\right\}.$$

Some studies of meta-analysis might not have all the treatments available. For example, some studies might compare fewer than $K + 1$ treatments, or some baseline treatments may be different, or both cases could occur simultaneously. The data structure is analogous to an incomplete-blocks design. For example, from Antiplatelet data, the 8th - 17th studies are composed of two treatment comparison unlike the 1st - 7th studies comparing three treatments.

Let $b(i)$ denote *baseline treatment* or called *study-specific reference treatment* [4] corresponding to the ith study, which can be the control group or any other treatments. As mentioned earlier about indirect comparison, in a situation that the treatments in some studies cannot be compared directly to the control group, we need to use evidence from the external studies. To make it clear, if $b(i)$ = 0 then the direct comparison is involved in this study. Conversely, if $b(i) \neq 0$ then the study makes indirect comparison. Let $J_{(i)} = J_i \setminus \left\{b(i)\right\}$ represent the

set of treatments involved in the ith study but excluding the baseline treatment $b(i)$. Let k_i and k_{i+1} denote the number of treatments in the sets J_i and J_i respectively. The $r_{ib(i)}$ and r_{ij} are binomially distributed, respectively as $Bin\left(n_{ib(i)}, \pi_{ib(i)}\right)$ and $Bin\left(n_{ij}, \pi_{ij}\right)$ for $i = 1, \cdots, M$ and $j \in J_{(i)}$. Let D be set of studies that make the direct comparison of treatment thus let I be set of indirect comparison of treatment.

The multi-arm trials data is consisted of 27 RCTs from systematic reviews of Antiplatelet Trialists' Collaboration [2] in total as shown in **Table 1**. The studies compare three treatments: aspirin plus dipyridamole (A), aspirin alone (B) and control group (C), where seven trials compare aspirin plus dipyridamole, aspirin alone and control group (*i.e.* comparing all A, B and C), ten trials compare aspirin plus dipyridamole and control group (*i.e.* comparing A and C) and ten trials compare aspirin alone and control group (*i.e.* comparing B and C) The "event" in **Table 1** represents the number of patients in whom deep venous thrombosis was detected by systematic fibrinogen scans or venography, or both, after general and orthopedic surgery and in high risk medical patients. The "total" represents the number of patients controlled in each group.

3. Fitting the Logistic Regression Model

This section illustrates how to fit the logistic regression model to the binary data related to multi-arm trials including the direct and indirect comparisons. Logistic regression is a regression model for a binomially distributed response/dependent variable. It is useful for modelling the probability of an event occurring as a function of other factors. Logistic regression is part of a category of statistical models called *generalized linear models* and uses the logit as its link function. Logistic regression can be used only with two types of dependent variables: one is a categorical dependent variable that has exactly two categories (*i.e.* a binary or dichotomous variable). The other is a continuous dependent variable that has values in the range 0 to 1 representing the probability values or the proportions. The names for logistic regression used in various other application areas are *logistic model* or *logit model*. Logistic regression is similar to linear regression. In linear regression, ultimate objective for the study may be either estimation of the coefficient values, or prediction of the response value. One significant difference between logistic and linear models is that the linear model has a continuous response variable and the logistic model uses a binary or dichotomous response. Logistic regression models for the ith study can be defined by

$$\log\left(\frac{\pi_{ib(i)}}{1 - \pi_{ib(i)}}\right) = \alpha_i \tag{1}$$

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \alpha_i + \delta_{i,b(i)j}, \quad j \in J(i) \tag{2}$$

where α_i and $\delta_{i,b(i)j}$ are trial effect and treatment effect, they will be detailed next.

Table 1. RCTs of Antiplatelet data.

Study number	Number of patients		
	Aspirin + Dipyridamole (A) even/total	Aspirin (B) even/total	Control (C) even/total
1	3/31	7/30	13/35
2	6/12	6/9	4/9
3	3/30	9/32	13/34
4	0/100	4/100	5/100
5	6/18	8/16	8/25
6	1/11	2/10	4/11
7	0/11	2/14	1/14
8	13/75		35/75
9	12/85		24/75
10	3/38		14/66
11	1/30		11/36
12	20/32		21/32
13	10/20		8/20
14	8/21		8/22
15	3/13		6/15
16	1/19		7/19
17	6/40		14/40
18		42/153	33/150
19		5/702	11/679
20		9/56	11/49
21		9/357	32/357
22		16/50	12/50
23		7/138	17/140
24		27/66	29/63
25		16/44	20/44
26		7/26	4/25
27		11/58	23/59

3.1. Trial Effect

Two assumptions are usually made about the trial effect α_i. The first one is that the trial effects are assumed to be *study-level effects*, which means that the α_i's are different parameters and are treated as nuisance parameters in the model. The M different unknown parameters are needed to include in the model. The

second one is that the model may be assumed for the α_i's. A special case is to assume that the trial effect is a fixed effect, defined by $\alpha_i = \alpha_0$. Conversely, it may be assumed to be a random effect, given by $\alpha_i \sim N\left(\mu_{\alpha 0}, \tau_{\alpha 0}^2\right)$, where $\mu_{\alpha 0}$ is the overall mean of the trial effect and $\tau_{\alpha 0}^2$ measures the magnitude of the variation between the studies. Most of the existing methods therefore used the first assumption. However, the number of unknown parameters is the same as the number of studies if the first assumption of the trial effect is made. It will result in some theoretical and computational problems. The accuracy of the estimation depends on the sample size of each study not the overall sample size of the pool in the meta-analysis. Therefore, the α_i's are assumed to be different in the paper.

3.2. Treatment Effect

The treatment effect $\delta_{i,b(i)j}$ can be direct treatment effect if $i \in D$ or indirect treatment effect if $i \in I$: they are defined as follows.

$$\delta_{i,b(i)j} = \begin{cases} \delta_{i,0j} \sim N\left(\mu_{0j}, \tau_{0j}^2\right) & \text{if } i \in D, \\ \delta_{i,0j} - \delta_{i,0b(i)} \sim N\left(\mu_{0j} - \mu_{0b(i)}, \tau_{0j}^2 + \tau_{0b(i)}^2 - 2\rho_{jb(i)}\tau_{0j}\tau_{0b(i)}\right) & \text{if } i \in I. \end{cases}$$

where $\rho_{jb(i)}$ is the correlation coefficient between $\delta_{i,0j}$ and $\delta_{i,0b(i)}$. For example, if the data, includes 3 treatments such as treatment A, B, C and C stands for control group. The baseline treatment for some studies exclude the control group C (such as A and B) thus the indirect treatment effect can be written as

$$\delta_{i,AB} = \delta_{i,AC} - \delta_{i,BC} \sim N\left(\mu_{AC} - \mu_{BC}, \tau_{AC}^2 + \tau_{BC}^2 - 2\rho_{AB}\tau_{AC}\tau_{BC}\right), \ i = I.$$

Next, we shall consider the treatment effect in a matrix form. Let $\boldsymbol{\delta}_{i,0}$ and $\boldsymbol{\mu}_0$ represent the vectors of $\left(\delta_{i,0j}, j = 1, \cdots, K\right)^t$ and $\left(\mu_{0j}, j = 1, \cdots, K\right)^t$ respectively where the superscript t stands for matrix transposition and let $\boldsymbol{\Omega}_0$ represent the $K \times K$ covariance matrix. The model $\boldsymbol{\delta}_{i,0}$ can be written as

$$\boldsymbol{\delta}_{i,0} \sim MVN\left(\boldsymbol{\mu}_0, \boldsymbol{\Omega}_0\right).$$

This is called *the basic model of random treatment effect*. Let $\boldsymbol{F}_{ij}$ be the index vector of length K consisting of elements 0 and 1 corresponding to $\delta_{i,b(i)j}$, given by

$$\boldsymbol{F}_{ij} = \begin{cases} \left(0, \cdots, \underset{b(i)\text{th}}{0}, \cdots \underset{j\text{th}}{1}, \cdots, 0\right) & \text{if } i \in D, \\ \left(0, \cdots, \underset{b(i)\text{th}}{-1}, \cdots, \underset{j\text{th}}{1}, \cdots, 0\right) & \text{if } i \in I. \end{cases}$$

Now, the random effect $\delta_{i,b(i)j}$ can be written in the form of $\boldsymbol{\delta}_{i,0}$ and $\boldsymbol{F}_{ij}$:

$$\delta_{i,b(i)j} = \boldsymbol{F}_{ij}\boldsymbol{\delta}_{i,0} \sim N\left(\boldsymbol{F}_{ij}\boldsymbol{\mu}_0, \boldsymbol{F}_{ij}\boldsymbol{\Omega}_0\boldsymbol{F}_{ij}^t\right).$$

As before, the covariance between the treatment effects $\delta_{i,b(i)j}$ and $\delta_{i,b(i)k}$ for $j \neq k$ and $j, k \in J_{(i)}$ may be dependent. For the ith study, let $\boldsymbol{F}_i$ be the

following $k_i \times K$ matrix

$$F_i = \left(F_{ij} \right)_{k_i \times K} \quad \text{for } j \in J_{(i)},$$

Let δ_i denote the vector $\left(\delta_{i,b(i)j}, j \in J_{(i)} \right)^t$ then we have

$$\delta_i = F_i \delta_{i,0} \sim MVN \left(\mu_i, \Omega_i \right),$$

where $\mu_i = F_i \mu_0$ and $\Omega_i = F_i \Omega_0 F_i^t$.

The α_i are assumed to be different and the $\delta_{i,b(i)j}$ are assumed to be a random effect as δ_i. The models in the Equation (1) and Equation (2) can be used for both treatment comparisons. From model in Equation (2), log $\pi_{ij} / 1 - \pi_{ij}$ is called the *logistic transform of probability* π_{ij} or alternatively *log odds* π_{ij} or *logit* π_{ij}. Having considered the properties of logit π_{ij}, the term $\pi_{ij} / 1 - \pi_{ij}$ is the odds of an unsuccessful outcome from a patient treated with treatment *j* and so logit π_{ij} is the log odds of being the case. It is easily seen that a value of π_{ij} in the range (0, 1) corresponds to a value of logit π_{ij} in $(-\infty, \infty)$. As π_{ij} approaches to 0, logit π_{ij} approaches to $-\infty$; as π_{ij} approaches to 1, logit π_{ij} approaches to ∞ and for $\pi_{ij} = 0.5$, logit $\pi_{ij} = 0$. After some rearrangement, the logistic regression models in Equation (1) and Equation (2) have respectively equivalent formulations as

$$\pi_{ib(i)} = \left(\frac{e^{\alpha_i}}{1 + e^{\alpha_i}} \right) \quad \text{and} \quad \pi_{ij} = \left(\frac{e^{\alpha_i + \delta_{i,b(i)j}}}{1 + e^{\alpha_i + \delta_{i,b(i)j}}} \right). \tag{3}$$

There are two alternative maximum likelihood (ML) approaches, the unconditional and conditional approaches that can be used to estimate the unknown parameters in a logistic regression model.

4. Unconditional Maximum Likelihood Approach

Generally, unconditional ML estimation is preferred if the number of parameters in the model is small relative to the number of studies in a meta-analysis [25].

4.1. Probability Functions

To demonstrate the unconditional ML estimation, let $p\left(r_{ib(i)} \mid \alpha_i \right)$ and $p\left(r_{ij} \mid \alpha_i, \delta_{i,b(i)j} \right)$ denote the probability functions associated with the distributions of $r_{ib(i)} \mid \alpha_i$ and $r_{ij} \mid \alpha_i, \delta_{i,b(i)j}$ respectively for $i = 1, \cdots, M$ and $j \in J_{(i)}$, defined as follows.

For the baseline treatment,

$$p\left(r_{ib(i)} \mid \alpha_i \right) = \binom{n_{ib(i)}}{r_{ib(i)}} \pi_{ib(i)}^{r_{ib(i)}} \left(1 - \pi_{ib(i)} \right)^{n_{ib(i)} - r_{ib(i)}} = \binom{n_{ib(i)}}{r_{ib(i)}} \frac{e^{\alpha_i r_{ib(i)}}}{\left(1 + e^{\alpha_i} \right)^{n_{ib(i)}}} \tag{4}$$

For the treatments *j*, $j \in J_{(i)}$

$$p\left(r_{ij} \mid \alpha_i, \delta_{i,b(i)j} \right) = \binom{n_{ij}}{r_{ij}} \pi_{ij}^{r_{ij}} \left(1 - \pi_{ij} \right)^{n_{ij} - r_{ij}} = \binom{n_{ij}}{r_{ij}} \frac{e^{\left(\alpha_i + \delta_{i,b(i)j} \right) r_{ij}}}{\left(1 + e^{\alpha_i + \delta_{i,b(i)j}} \right)^{n_{ij}}}. \tag{5}$$

The combination in Equation (4) represents the number of possible combinations of $n_{ib(i)}$ taken $r_{ib(i)}$ at a time. The $\pi_{ib(i)}$ in the middle term of Equation (4) is substituted from Equation (3) and $\left(1 - \pi_{ib(i)}\right)$ becomes $1/1 + e^{\alpha_i}$. The combination in Equation (5) can be analyzed in the same way.

4.2. The Unconditional Likelihood

From the probability functions $p\left(r_{ib(i)} \mid \alpha_i\right)$ and $p\left(r_{ij} \mid \alpha_i, \delta_{i,b(i)j}\right)$, the trial effects α_i's are study-level effects. They are assumed to be different and also included in both probability functions. While the $\delta_{i,b(i)j}$ is a random effect, thus the $p\left(r_{ij} \mid \alpha_i, \delta_{i,b(i)j}\right)$ involves the vector of random effects, $\boldsymbol{\delta}_i$. The standard method of handling a probability function which involves random variables that have a fully specified probability is to integrate the probability function with respect to the distribution of those variables. To deal with the random effects $\boldsymbol{\delta}_i$, let $r_{(i)}$ be the vector $\left(r_{ij}, j \in J_{(i)}\right)^t$. The probability function $p\left(r_{ij} \mid \boldsymbol{\delta}_i\right)$ will be integrated with respect to $\boldsymbol{\delta}_i$. The $p\left(\boldsymbol{r}_{(i)}\right)$ contains k_i integrals, which is the number of treatments in the set $J_{(i)}$, and is given by

$$p\left(\boldsymbol{r}_{(i)}\right) = \int_{\boldsymbol{\delta}_i} \prod_{j \in J_{(i)}} p\left(r_{ij} \mid \boldsymbol{\delta}_i\right) \phi\left(\boldsymbol{\delta}_i; \boldsymbol{\mu}_i, \boldsymbol{\Omega}_i\right) \mathrm{d}\boldsymbol{\delta}_i, \tag{6}$$

where $\phi\left(\boldsymbol{\delta}_i; \boldsymbol{\mu}_i, \boldsymbol{\Omega}_i\right)$ is the probability density function of the normal distribution with mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Omega}_i$ given by

$$\phi\left(\boldsymbol{\delta}_i; \boldsymbol{\mu}_i, \boldsymbol{\Omega}_i\right) = \frac{1}{\left(2\pi\right)^{k_i/2} \left|\boldsymbol{\Omega}_i\right|^{1/2}} e^{-\left(\boldsymbol{\delta}_i - \boldsymbol{\mu}_i\right)' \boldsymbol{\Omega}_i^{-1} \left(\boldsymbol{\delta}_i - \boldsymbol{\mu}_i\right)/2} \tag{7}$$

The integral in Equation (6) can be calculated numerically; one way to do it is to use the Gauss-Hermite method. To apply Gauss-Hermite approximation, the probability function $p\left(\boldsymbol{r}_{(i)}\right)$ for the ith study can be estimated by

$$p\left(\boldsymbol{r}_{(i)}\right) \approx \pi^{k_i/2} \sum_{n_1=1}^{l_1} w_{n_1}^{(1)} \cdots \sum_{n_{k_i}=1}^{l_{k_i}} w_{n_{k_i}}^{(k_i)} \left\{ \prod_{j \in J_{(i)}} \binom{n_{ij}}{r_{ij}} \frac{e^{\left(\alpha_i + \left(\boldsymbol{\mu}_i + \sqrt{2}\boldsymbol{\Omega}_i^{1/2} d_{i,n}\right)\right) r_{ij}}}{\left(1 + e^{\alpha_i + \left(\boldsymbol{\mu}_i + \sqrt{2}\boldsymbol{\Omega}_i^{1/2} d_{i,n}\right)}\right)^{n_{ij}}} \right\}, \tag{8}$$

where the sampling nodes are at $\boldsymbol{\mu}_i + \sqrt{2}\boldsymbol{\Omega}_i^{1/2} d_{i,n}$ and $\boldsymbol{d}_{i,n} = \left(x_{n_1}^{(1)}, \cdots, x_{n_{k_i}}^{(k_i)}\right)$. The vector $\boldsymbol{d}_{i,n}$ depends on the number k_i, which is the number of treatments comprising in the ith study. The resulting function $p\left(\boldsymbol{r}_{(i)}\right)$ does not depend on the $\boldsymbol{\delta}_i$. For most practical purposes, l_{k_i} need not be greater than 20, although some authors suggest using even smaller values [26].

As before, let $\boldsymbol{\theta}$ be the collection of all unknown parameters for the meta-analysis including all trial effects $\left(\alpha_i, \cdots, \alpha_M\right)$, $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ and let $\boldsymbol{r}_i$ be the vector $\left(r_{ij}, j \in J_i\right)$. The likelihood function for the ith study can be written as

$$L\left(\boldsymbol{\theta} \mid \boldsymbol{r}_i\right) = \prod_{j \in J_i} p\left(r_{ij}\right) = p\left(r_{ib(i)} \mid \alpha_i\right) p\left(\boldsymbol{r}_{(i)}\right). \tag{9}$$

where $p\left(r_{ib(i)} \mid \alpha_i\right)$ and $p\left(\boldsymbol{r}_{(i)}\right)$ are given in Equation (4) and Equation (8) respectively. Let $l_{u,i} = \log L\left(\boldsymbol{\theta} \mid \boldsymbol{r}_i\right)$ standing for the unconditional log-likelihood function of the logistic regression model for the ith study. The log-likelihood

function of $\boldsymbol{\theta}$ for the logistic regression models is given by

$$l_u(\boldsymbol{\theta}) = \sum_{i=1}^{M} l_{u,i}. \tag{10}$$

The number of α_i's is the same as the number of studies. The computation of MLEs may be quite unstable if the number of studies is large while the sample size of each study is small. As discussed earlier, this may also result in a biased or misleading estimate. A conditional approach is suggested to eliminate all nuisance parameters.

4.3. Asymptotic Variance-Covariance Matrix

This section shows how to calculate the standard errors for the MLEs of the logistic regression model using the unconditional approach. Since there are random effects in the model, some integrals are involved in the likelihood function. The unconditional log-likelihood function $l_u(\boldsymbol{\theta})$ in Equation (10) can be written as

$$\begin{aligned} l_u(\boldsymbol{\theta}) &= \sum_{i=1}^{M} \log p\left(r_{ib(i)}\right) + \sum_{i=1}^{M} \log p\left(\boldsymbol{r}_{(i)}\right) \\ &= \sum_{i=1}^{M} \log p\left(r_{ib(i)}\right) + \sum_{i=1}^{M} \log \int_{\boldsymbol{\delta}_i} \prod_{j \in J_{(i)}} p\left(r_{ij} \mid \boldsymbol{\delta}_i\right) \phi\left(\boldsymbol{\delta}_i; \boldsymbol{\mu}_i, \boldsymbol{\Omega}_i\right) \mathrm{d}\boldsymbol{\delta}_i. \end{aligned} \tag{11}$$

Let l_1 and l_2 stand for the first and second terms of the above log-likelihood function, given by $l_u(\boldsymbol{\theta}) = l_1 + l_2$. Three types of unknown parameters are involved in $\boldsymbol{\theta}$; the trial effects, α_i's, the overall mean effects μ's (for $\boldsymbol{\mu}$), and the variances τ's and the correlation coefficients ρ's in the covariance matrix $\boldsymbol{\Omega}$. Let τ represent a parameter (either τ or ρ) involved in $\boldsymbol{\Omega}$. There is no random effect involved in l_1.

First, the second-order partial derivative $\partial^2 l_1 / \partial \alpha_i^2$ can be calculated in the usual way; while the other terms are

$$\frac{\partial^2 l_1}{\partial \alpha_i \alpha_j} = \frac{\partial^2 l_1}{\partial \alpha_i \mu} = \frac{\partial^2 l_1}{\partial \alpha_i \tau} = 0, \quad i \neq j \quad \text{and} \quad i, j \in \{1, \cdots, M\}.$$

Next, consider the second term of Equation (11), for notational convenience, let $P_i(\boldsymbol{\delta}_i)$ represent the function $\prod_{j \in J_{(i)}} p\left(r_{ij} \mid \boldsymbol{\delta}_i\right)$ in l_2. Now the term l_2 takes the form

$$l_2\left(\alpha_1, \cdots, \alpha_M, \boldsymbol{\mu}, \boldsymbol{\Omega}\right) = \sum_{i=1}^{M} \log \int_{\boldsymbol{\delta}_i} p\left(\boldsymbol{\delta}_i\right) \phi\left(\boldsymbol{\delta}_i\right) \mathrm{d}\boldsymbol{\delta}_i = \sum_{i=1}^{M} l_{2i},$$

where $\phi(\boldsymbol{\delta})$ is the density of the multivariate normal distribution with mean $\boldsymbol{\mu}_i$ and variance matrix $\boldsymbol{\Omega}_i$, and l_{2i} is a summand of the log-likelihood involving the integrals ($\log \int_{\boldsymbol{\delta}_i} P_i(\boldsymbol{\delta}_i) \phi(\boldsymbol{\delta}_i) \mathrm{d}\boldsymbol{\delta}_i$). The first-order partial derivatives relating to l_2 are shown as follows

$$\frac{\partial l_2}{\partial \alpha_i} = \sum_{i=1}^{M} \mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} \frac{\partial P_i(\boldsymbol{\delta}_i)}{\partial \alpha_i} \phi(\boldsymbol{\delta}_i) \mathrm{d}\boldsymbol{\delta}_i,$$

$$\frac{\partial l_2}{\partial \mu} = \sum_{i=1}^{M} \mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} P_i(\boldsymbol{\delta}_i) \frac{\partial \phi(\boldsymbol{\delta}_i)}{\partial \mu} \mathrm{d}\boldsymbol{\delta}_i,$$

$$\frac{\partial l_2}{\partial \tau} = \sum_{i=1}^{M} \mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} P_i(\boldsymbol{\delta}_i) \frac{\partial \phi(\boldsymbol{\delta}_i)}{\partial \tau} \mathrm{d}\boldsymbol{\delta}_i.$$

Similarly, the second-order partial derivatives are

$$\frac{\partial^2 l_2}{\partial \alpha_i^2} = \sum_{i=1}^{M} \left(\mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} \frac{\partial^2 P_i(\boldsymbol{\delta}_i)}{\partial \alpha_i^2} \phi(\boldsymbol{\delta}_i) \mathrm{d}\boldsymbol{\delta}_i - \left(\mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} \frac{\partial P(\boldsymbol{\delta}_i)}{\partial \alpha_i} \phi(\boldsymbol{\delta}_i) \mathrm{d}\boldsymbol{\delta}_i \right)^2 \right), \quad (12)$$

$$\frac{\partial^2 l_2}{\partial \mu^2} = \sum_{i=1}^{M} \left(\mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} P_i(\boldsymbol{\delta}_i) \frac{\partial^2 \phi(\boldsymbol{\delta}_i)}{\partial \mu^2} \mathrm{d}\boldsymbol{\delta}_i - \left(\mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} P_i(\boldsymbol{\delta}_i) \frac{\partial \phi(\boldsymbol{\delta}_i)}{\partial \mu} \mathrm{d}\boldsymbol{\delta}_i \right)^2 \right), \quad (13)$$

$$\frac{\partial^2 l_2}{\partial \tau^2} = \sum_{i=1}^{M} \left(\mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} P_i(\boldsymbol{\delta}_i) \frac{\partial^2 \phi(\boldsymbol{\delta}_i)}{\partial \tau^2} \mathrm{d}\boldsymbol{\delta}_i - \left(\mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} P_i(\boldsymbol{\delta}_i) \frac{\partial \phi(\boldsymbol{\delta}_i)}{\partial \tau} \mathrm{d}\boldsymbol{\delta}_i \right)^2 \right), \quad (14)$$

$$\frac{\partial^2 l_2}{\partial \mu \partial \tau} = \sum_{i=1}^{M} \left(\mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} P_i(\boldsymbol{\delta}_i) \frac{\partial^2 \phi(\boldsymbol{\delta}_i)}{\partial \mu \partial \tau} \mathrm{d}\boldsymbol{\delta}_i - \left(\mathrm{e}^{-l_{2i}} \int_{\boldsymbol{\delta}_i} P_i(\boldsymbol{\delta}_i) \frac{\partial \phi(\boldsymbol{\delta}_i)}{\partial \mu} \mathrm{d}\boldsymbol{\delta}_i \right)^2 \right). \quad (15)$$

Note that the second-order partial derivative $\partial^2 l_2 / \partial \alpha_i \partial \alpha_j$ is equal to zero. The second-order partial derivatives of $\partial^2 l_2 / \partial \mu_i \partial \mu_j$ and $\partial^2 l_2 / \partial \tau_i \partial \tau_j$ can be expressed in similar equations to Equation (13) and Equation (14) respectively. The integrals in the first-order and second-order partial derivatives can be approximated by Gaussian quadrature.

From the log-likelihood $l_u(\boldsymbol{\theta})$ in Equation (11), the second-order partial derivatives for the observed Fisher information matrix can be calculated as

$$\frac{\partial^2 l_u}{\partial \alpha_i^2} = \frac{\partial^2 l_1}{\partial \alpha_i^2} + \frac{\partial^2 l_2}{\partial \alpha_i^2}, \quad \frac{\partial^2 l_u}{\partial \mu^2} = \frac{\partial^2 l_2}{\partial \mu^2}, \quad \frac{\partial^2 l_u}{\partial \mu_i \partial \mu_j} = \frac{\partial^2 l_2}{\partial \mu_i \partial \mu_j},$$

$$\frac{\partial^2 l_u}{\partial \tau^2} = \frac{\partial^2 l_2}{\partial \tau^2}, \quad \frac{\partial^2 l_u}{\partial \tau_i \partial \tau_j} = \frac{\partial^2 l_2}{\partial \tau_i \partial \tau_j}, \quad \frac{\partial^2 l_u}{\partial \mu \partial \tau} = \frac{\partial^2 l_2}{\partial \mu \partial \tau}.$$

As set earlier, the second partial derivatives of $\partial^2 l_u / \partial \rho^2$ and $\partial^2 l_u / \partial \mu \partial \rho$ (and $\partial^2 l_u / \partial \tau \partial \rho$) can be calculated in similar equations to $\partial^2 l_u / \partial \tau^2$ and $\partial^2 l_u / \partial \mu \partial \tau$ respectively. Notice that the second-order derivative of l_1 is only related in $\partial^2 l_u / \partial \alpha_i^2$. The matrix of second partial derivatives can be partitioned into a block matrix with null matrices in the off diagonals:

$$\mathcal{H}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{H}_\alpha(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{H}_{\mu \tau \rho}(\boldsymbol{\theta}) \end{pmatrix},$$

where $\mathcal{H}_\alpha(\boldsymbol{\theta})$ and $\mathcal{H}_{\mu \tau \rho}(\boldsymbol{\theta})$ are the second-order partial derivatives about α_j, and μ, τ and ρ respectively. By multiplying $\mathcal{H}(\boldsymbol{\theta})$ by -1, the observed Fisher information matrix $\mathcal{J}(\boldsymbol{\theta})$ is obtained. The inverse of $\mathcal{J}(\boldsymbol{\theta})$ is the asymptotic variance-covariance matrix of MLEs and their standard errors are the square roots of the diagonal of $\mathcal{J}(\theta)^{-1}$.

5. Conditional Maximum Likelihood Approach

Conditional likelihood is widely used in logistic regression models with binary data. In particular, this leads to accurate inferences for the parameters of interest

and eliminates all nuisance parameters [25]. The conditional likelihood will be defined and the maximum likelihood estimation will be described in this section.

5.1. Conditional Likelihood

From the logistic regression models in Equation (1) and Equation (2), the conditional likelihood r_i given that $C_i = \sum_{j \in J_i} r_{ij} = c_i$ for the ith study, is given by

$$f\left(r_i \mid C_i = c_i; \delta_i\right) = f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i; \delta_i\right) = \frac{f\left(r_i \mid \delta_i\right)}{f\left(\sum_{j \in J_i} r_{ij} = c_i \mid \delta_i\right)}. \tag{16}$$

The conditional likelihood reflects the probability of the observed data configuration relative to the probability of all possible configurations of the given data. The numerator $f\left(r_i \mid \delta_i\right)$ is exactly the same as the unconditional likelihood obtained from Equation (4) and Equation (5). The denominator is what makes the conditional likelihood different from the unconditional likelihood; it sums the joint probability for all possible configurations. To derive the Equation (16), the conditional likelihood r_i given C_i can be simplified as

$$f\left(r_i \mid C_i = c_i; \delta_i\right) = \frac{\prod_{j \in J_i} \binom{n_{ij}}{r_{ij}} e^{\delta_{ij} r_{ij}}}{\sum_{u_i \in U_i} \binom{n_{i0}}{c_i - \sum_{j \in J_{(i)}} u_{ij}} \prod_{j \in J_{(i)}} \binom{n_{ij}}{u_{ij}} e^{\delta_{ij} u_{ij}}}, \tag{17}$$

where $u_i = \left(u_{ij}, j \in J_{(i)}\right)^t$ and

$$U_i = \left\{ u_i : 0 \le u_{ij} \le n_{ij}, j \in J_{(i)} \text{ and } c_i - n_{i0} \le \sum_{j \in J_{(i)}} u_{ij} \le c_i \right\}.$$

Notice that this likelihood function does not involve any nuisance parameters α_i's and is a function of δ_i alone. The removal of the trial effects from the conditional likelihood is important because when the conditional likelihood is used, estimates are obtained only for the parameters of interest in the model and not for the α_i's.

5.2. Estimation

The conditional likelihood in Equation (17) has k_i random effects so the likelihood $f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i\right)$ involves k_i integrations:

$$f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i\right) = \int_{\delta_i} f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i; \delta_i\right) \phi\left(\delta_i; \mu_i, \Omega_i\right) d\delta_i, \tag{18}$$

where $\phi\left(\delta_i; \mu_i, \Omega_i\right)$ is the probability density function of multivariate normal distribution with mean μ_i and covariance Ω_i given in Equation (7). Gauss-Hermite approximation is applied to Equation (18) and obtain:

$$f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i\right) \approx \pi^{-k_i/2} \sum_{n_1=1}^{l_1} w_{n_1}^{(1)} \cdots \sum_{n_{k_i}=1}^{l_{k_i}} w_{n_{k_i}}^{(k_i)} f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i; \delta_{i,n}\right), \tag{19}$$

where $f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i; \delta_{i,n}\right)$ is obtained from Equation (17) where the sampling nodes is $\delta_{i,n} = \mu_i + \sqrt{2}\Omega_i^{1/2} d_{i,n}$ and $d_{i,n} = \left(d_{n_1}^{(1)}, \cdots, d_{n_{k_i}}^{(k_i)}\right)$. The likelihood for the ith study $L(\theta \mid r_i)$ can be written as

$$L(\theta \mid r_i) = f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i; \delta_i\right).$$

The log-likelihood function of the logistic regression models using the conditional approach is

$$l_c(\theta) = \log L(\theta \mid r) = \sum_{i=1}^{M} \log L(\theta \mid r_i) \tag{20}$$

By maximizing the conditional likelihood function over θ, an exact parameter estimate is obtained for θ, called *the conditional maximum likelihood estimate*. To calculate the standard error of their MLEs, the log-likelihood function in Equation (20) can be written as

$$
\begin{aligned}
l_c(\theta) &= \sum_{i=1}^{M} \log f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i\right) \\
&= \sum_{i=1}^{M} \log \int_{\delta_i} f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i, \delta_i\right) \phi(\delta_i; \mu_i, \Omega_i) \, \mathrm{d}\delta_i.
\end{aligned} \tag{21}
$$

Let $P_i(\delta_i)$ represent $f\left(r_i \mid \sum_{j \in J_i} r_{ij} = c_i, \delta_i\right)$ in the above equation. The second-order partial derivatives of $\partial^2 l_2 / \partial \mu^2$, $\partial^2 l_2 / \partial \tau^2$ and $\partial^2 l_2 / \partial \mu \partial \tau$ are similar to Equation (13) - Equation (15) respectively. In a similar way to the previous section, the standard errors for the MLEs are obtained.

6. Application to Antiplatelet Therapy Data

From the data given in Table 1, the total number of individual observations is small thus the empirical log-odds model is not appropriate. The logistic regression model will be applied using the unconditional and conditional approaches with the data.

6.1. Unconditional Inference

From the data, there are 27 studies investigating the use of aspirin plus dipyridamole or aspirin alone in comparison with the control group. The studies compare three treatments: aspirin plus dipyridamole (A), aspirin alone (B) and the control treatment (C). Seven studies compare A, B and C, ten studies compare A and C and ten studies compare B and C. There is no indirect comparison for this dataset, so the set D is $\{1, \cdots, 27\}$. The baseline treatment for all studies is the control group ($b(i) = 0$).

The indices $i = 1, \cdots, 27$ and $j = 0, 1, 2$ stand for the studies and the treatments C, A and B, respectively. The data is partitioned into three groups: $G_1 = \{1, \cdots, 7\}$, $G_2 = \{8, \cdots, 17\}$ and $G_3 = \{18, \cdots, 27\}$. The sets J_i and $J_{(i)}$ are given by

$$
\begin{cases}
J_i = \{0, 1, 2\}, \ J_{(i)} = \{1, 2\} & \text{for } i \in G_1, \\
J_i = \{0, 1\}, \ J_{(i)} = \{1\} & \text{for } i \in G_2, \\
J_i = \{0, 2\}, \ J_{(i)} = \{2\} & \text{for } i \in G_3.
\end{cases} \tag{22}
$$

Let r_{i0}, r_{i1} and r_{i2} be the numbers of patients who suffered reocclusions on treatments C, A and B respectively, where the ith study is in $G_1 \cup G_2 \cup G_3$, $G_1 \cup G_2$ and $G_1 \cup G_3$, respectively. The total numbers of patients are n_{i0}, n_{i1} and n_{i2} Let π_{i0}, π_{i1} and π_{i2} be the probabilities that patients have reocclusions on treatments C, A and B respectively in the ith study. The r_{i0}, r_{i1} and r_{i2} are binomially distributed as

$$r_{i0} \sim Bin(\pi_{i0}, n_{i0}), \quad i \in G_1 \cup G_2 \cup G_3,$$
$$r_{i1} \sim Bin(\pi_{i1}, n_{i1}), \quad i \in G_1 \cup G_2,$$
$$r_{i2} \sim Bin(\pi_{i2}, n_{i2}), \quad i \in G_1 \cup G_3.$$

The treatment effect δ_i for G_1 is defined as

$$\begin{pmatrix} \delta_{i,01} \\ \delta_{i,02} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{01} \\ \mu_{02} \end{pmatrix}, \begin{pmatrix} \tau_{01}^2 & \rho\tau_{01}\tau_{02} \\ \rho\tau_{01}\tau_{02} & \tau_{02}^2 \end{pmatrix} \right) \tag{23}$$

Logistic regression models for the data can be fitted using the Equation (1) and Equation (2) where $b(i) = 0$ and $J_{(i)}$ is given in Equation (22). Note that the trial effects are assumed to be different in each study. To define the unconditional likelihood function, let $r_{(i)}$ represent the vector (r_{i1}, r_{i2}). The probability functions $p(r_{i0})$ and $p(r_{(i)})$ are formulated from Equation (4) and Equation (8) respectively.

From δ_i for G_1, the correlation coefficient ρ between $\delta_{i,01}$ and $\delta_{i,02}$ is in the form $\tau_0^2/\tau_{01}\tau_{02}$, where is obtained from $\delta_{i0} \sim N(\mu_0, \tau_0^2)$. Note that μ_0 and τ_0^2 are not estimable unless some other information is used. The assumption of homogeneity variance will be considered here. Suppose that all heterogeneity parameters are the same: $\tau_{01} = \tau_{02} = \tau$ and the correlation coefficient takes the value 1/2. The unknown parameter θ for the models is $\{\alpha_1, \alpha_2, \cdots, \alpha_{27}, \mu_{01}, \mu_{02}, \tau^2\}$. The log-likelihood function $l_u(\theta)$ is obtained from Equation (10). By maximizing the log-likelihood function, the MLEs can be estimated. Also their standard errors are given by the observed Fisher information matrix.

The results for the treatment effects δ_{01} and δ_{02} are given in Table 2. SD and CI stand for standard deviation and confidence interval respectively. The trial effects are presented in Table 3. The overall means on the log odds ratio

Table 2. The results of the treatment effects for the model using the unconditional method.

	δ_{01}		δ_{02}	
	μ_{01}	τ_{01}	μ_{01}	τ_{01}
LOR scale (SD)	−1.17849 (0.08499)	0.00372 (0.04752)	−0.63700 (0.03728)	0.00372 (0.04752)
95% CI	(−1.33, −1.00)	(−0.08, 0.09)	(−0.64, −0.62)	(−0.08, 0.09)
OR scale	0.30774		0.52800	
95% CI	(0.26, 0.36)		(0.52, 0.53)	

Table 3. The trial effect of the model using the unconditional method.

	Study 1	Study 2	Study 3	Study 4	Study 5
Trial Effect	−0.72387	1.20619	−0.54688	−3.01061	0.55283
(SD)	(0.01021)	(0.00934)	(0.00993)	(0.0117)	(0.00926)
95% CI	(−0.74, −0.70)	(1.18, 1.22)	(−0.56, −0.52)	(−3.03, −0.98)	(0.53, 0.57)
	Study 6	**Study 7**	**Study 8**	**Study 9**	**Study 10**
Trial Effect	−0.85773	−1.69947	−0.34480	−0.65231	−1.29308
(SD)	0.01206)	(0.01264)	(0.00735)	(0.00770)	(0.01087)
95% CI	(0.88, −0.83)	(−1.72, −1.67)	(−0.35, −0.33)	(−0.66, −0.63)	(−1.31, −1.27)
	Study 11	**Study 12**	**Study 13**	**Study 14**	**Study 15**
Trial Effect	2.18147	1.68130	1.17724	0.68567	−0.14132
(SD)	(0.01231)	(0.00636)	(0.00811)	(0.00869)	(0.01102)
95% CI	(−2.20, −2.15)	(1.66, 1.69)	(1.16, 1.19)	(0.66, 0.70)	(−0.16, −0.11)
	Study 16	**Study 18**	**Study 18**	**Study 19**	**Study 20**
Trial Effect	−1.53114	−0.57320	−0.33486	−4.24972	−1.05748
(SD)	(0.01214)	(0.00941)	(0.00661)	(0.01367)	(0.01199)
95% CI	(−1.55, −1.50)	(−0.59, −0.55)	(−0.34, −0.32)	(−4.27, −4.22)	(−1.08, −1.03)
	Study 21	**Study 22**	**Study 23**	**Study 24**	**Study 25**
Trial Effect	−3.01727	−0.11773	−2.23388	0.24853	0.04007
(SD)	(0.01184)	(0.00977)	(0.01252)	(0.00802)	(0.00987)
95% CI	(−3.04, −2.99)	(−0.13, −0.09)	(−2.25, −2.20)	(0.23, 0.26)	(0.02, 0.05)
	Study 26	**Study 27**			
Trial Effect	−0.37573	−0.76995			
(SD)	(0.01277)	(0.01098)			
95% CI	(−0.40, −0.35)	(−0.79, −0.74)			

(LOR) scale for δ_{01} and δ_{02} are −1.17849 (SD 0.08499) and −0.63700 (SD 0.03728), and the heterogeneity parameter is 0.0372 (SD 0.04752). On the odds ration (OR) scale, the means are 0.30774 and 0.52800 respectively. Their confidence interval (CI) can be calculated from the related CI on the LOR scale. To be concluded, treatments aspirin plus dipyridamole and aspirin only in antiplatelet therapy reduce deep venous thrombosis by over 70% and 45% respectively. The average of both treatments reduces deep venous thrombosis by over 55%.

6.2. Conditional Inference

The models and other parameters are similar to those defined in the unconditional method. The function C_i for the data can be defined by

$$
\begin{cases}
C_i = r_{i0} + r_{i1} + r_{i2} & \text{for } i \in G_1, \\
C_i = r_{i0} + r_{i1} & \text{for } i \in G_2, \\
C_i = r_{i0} + r_{i2} & \text{for } i \in G_3.
\end{cases}
\tag{24}
$$

Let r_i denote the vector (r_{i0}, r_{i1}, r_{i2}). The conditional likelihood $f(r_i | C_i)$ for the ith study is given in Equation (18). To handle the random treatment effect δ_i, the likelihood function is approximated by Gaussian-Hermite approximation as defined in Equation (19). The unknown parameter θ for the models is $\{\mu_{01}, \mu_{02}, \tau\}$. By using the log-likelihood function in Equation (20), the results of the models are given in Table 4. On the LOR scale, the overall mean effects for both treatment effects are −0.87516 (SD 0.04340) and −0.39000 (SD 0.31160) while their variation between studies is 0.37000 (SD 0.03900). Those means on the OR scale are 0.41679 and 0.67434. As before their confidence intervals are obtained from the related CI on the LOR scale. The results indicate that treatments aspirin plus dipyridamole and aspirin only produce a reduction in deep venous thrombosis by over 55% and 30% respectively. The average of both treatments in antiplatelet therapy reduces deep venous thrombosis by over 40%.

As seen from Table 2 and Table 4, the results from using the unconditional likelihood (on the LOR scale) are smaller than from using conditional likelihood. Note that those results are negative. That is to say that estimation with unconditional likelihood may cause underestimation or bias. Collaboration [2] summarized that antiplatelet therapy produced a highly significant (2p ≤ 0.00001) reduction in deep venous thrombosis of about 40%. The results from the model using the conditional likelihood support this.

7. Conclusions and Discussion

The logistic regression model was introduced for the exact binomial distribution. Two alternative approaches for making inferences were presented. The unconditional likelihood involves nuisance parameters (from the trial effects). If the number of studies (M) is large, it may lead to inconsistent estimators. Cox and Snell [15] concluded for the unconditional likelihood that if the number of studies (M) is large and the number of individual observations (n_{ij}) is small then it makes estimation inaccurate and inconsistent. The conditional maximum likelihood approach was introduced for the model to eliminate all nuisance parameters. In making a choice between the two approaches, it is needed to consider the number of studies and the number of individual observations. However, the use

Table 4. The results of the treatment effects for the model using the conditional method.

	δ_{01}		δ_{02}	
	μ_{01}	τ_{01}	μ_{01}	τ_{01}
LOR scale	−0.87516	0.37000	−0.39000	0.37000
(SD)	(0.04340)	(0.03900)	(0.31160)	(0.03900)
95% CI	(−0.96, −0.79)	(0.29, 0.44)	(−1.00, 0.22)	
OR scale	0.41679		0.67434	
95% CI	(0.38, 0.45)		(0.36, 1.24)	

of this method can be expensive in term of the cost of computer running time, especially if the number of individual observations is large. Using the unconditional maximum likelihood approach, note that if the number of studies is large and the number of individual observations is small then the estimate may be biased or misleading.

Some other methods can be used in the logistic regression model, for example, using a pseudo-loglikelihood, (see Severini [27]); or the modified profile likelihood (see Bellio and Sartori [28], Hamza, Houwelingen and Stijnen [29] and Caterina *et al.* [19]). Gaussian-Hermite quadrature was used to calculate the integral forms of the probabilities including random effects in the likelihood functions for both approaches [30]. The approximation is reasonably effective for low-order integrations [31]. Implementing Gaussian-Hermite approximation, we used the function "gauss.quad" in the software R to estimate MLEs for the model. The number of integrands depends on the number of treatments involved in those studies. If this number is large then it makes the dimensionality of the integral large and so it cannot be approximated accurately. Other approximations such as Laplace approximation or Monte Carlo method can be used, see Shi and Copas [32] and Caterina [19]. Laplace approximation could make the calculation of second-order derivatives for the observed Fisher information matrix easier than using Gaussian approximation since there is no weight term in the approximation [33] [34]. Additionally, simulation studies could be conducted in further work to compare these two approaches in different scenarios including improving the approximation if multi-arm trials are more than three treatment comparisons.

The established binomial tree method could be applied in the further work [24].

8. Conclusion

Simulation studies could be conducted in further work to compare these two approaches in different scenarios including improving the approximation if multi-arm trials are more than three treatment comparisons.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Hill, S.A.B. (1990) Memories of the British Streptomycin Trial in Tuberculosis: The First Randomized Clinical Trial. *Controlled Clinical Trials*, **11**, 77-79.

https://doi.org/10.1016/0197-2456(90)90001-I

[2] Antiplatelet Trialists' Collaboration (1994) Collaborative Overview of Randomised Trials of Antiplatelet Therapy—III: Reduction in Venous Thrombosis and Pulmonary Embolism by Antiplatelet Prophylaxis among Surgical and Medical Patients. *British Medical Journal*, **308**, 235-246. https://doi.org/10.1136/bmj.308.6923.235

[3] Chootrakool, H. and Shi, J.Q. (2008) Meta-Analysis of Multi-Arm Trials Using Empirical Logistic Transform. *The Open Medical Informatics Journal*, **2**, 112-116. https://doi.org/10.2174/1874431100802010112

[4] Chootrakool, H., Shi, J.Q. and Yue, R. (2011) Meta-Analysis and Sensitivity Analysis for Multi-Arm Trials with Selection Bias. *Statistics in Medicine*, **30**, 1183-1198. https://doi.org/10.1002/sim.4143

[5] Rücker, G. (2017) Methods for Including Information from Multi-Arm Trials in Pairwise Meta-Analysis Multi-Arm Trials in Pairwise Meta-Analysis. *Research Synthesis Methods*, **8**, 392-403. https://doi.org/10.1002/jrsm.1259

[6] Lu, G. and Ades, A.E. (2004) Combination of Direct and Indirect Evidence in Mixed Treatment Comparisons. *Statistics in Medicine*, **23**, 3105-3124. https://doi.org/10.1002/sim.1875

[7] Lee, A.W. (2014) Review of Mixed Treatment Comparisons in Published Systematic Reviews Shows Marked Increase since 2009. *Journal of Clinical Epidemiology*, **67**, 138-143.

[8] Salanti, G. (2012) Indirect and Mixed-Treatment Comparison, Network, or Multiple-Treatments Meta-Analysis: Many Names, Many Benefits, Many Concerns for the Next Generation Evidence Synthesis Tool. *Research Synthesis Methods*, **3**, 80-97. https://doi.org/10.1002/jrsm.1037

[9] Lumley, T. (2002) Network Meta-Analysis for Indirect Treatment Comparisons. *Statistics in Medicine*, **21**, 2313-2324. https://doi.org/10.1002/sim.1201

[10] Mbuagbaw, L., *et al.* (2017) Approaches to Interpreting and Choosing the Best Treatments in Network Meta-Analyses. *Systematic Reviews*, **6**, Article No. 79. https://doi.org/10.1186/s13643-017-0473-z

[11] Veroniki, A.A., *et al.* (2018) Is Providing Uncertainty Intervals in Treatment Ranking Helpful in a Network Meta-Analysis? *Journal of Clinical Epidemiology*, **100**, 122-129. https://doi.org/10.1016/j.jclinepi.2018.02.009

[12] Seide, S.E., Jensen, K. and Kieser, M. (2019) Simulation and Data-Generation for Random-Effects Network Meta-Analysis of Binary Outcome. *Statistics in Medicine*, **38**, 3288-3303. https://doi.org/10.1002/sim.8193

[13] Jenkins, D.A., *et al.* (2021) Methods for the Inclusion of Real-World Evidence in Network Meta-Analysis. *BMC Medical Research Methodology*, **21**, Article No. 207. https://doi.org/10.1186/s12874-021-01399-3

[14] Zhao, Y.P., *et al.* (2021) A Network Meta-Analysis for Neoadjuvant and Adjuvant Treatments for Resectable Squamous Cell Carcinoma of Esophagus. *Scientific Reports*, **11**, Article No. 6800. https://doi.org/10.1038/s41598-021-86102-8

[15] Cox, D.R. and Snell, E.J. (1989) Analysis of Binary Data. 2nd Edition, Chapman & Hall/CRC, Boca Raton.

[16] Tritchler, D. (1984) An Algorithm for Exact Logistic Regression. *Journal of the American Statistical Association*, **79**, 709-711. https://doi.org/10.2307/2288420

[17] Sartori, N. (2003) Modified Profile Likelihoods in Models with Stratum Nuisance Parameters. *Biometrika*, **90**, 533-549. http://www.jstor.org/stable/30042064 https://doi.org/10.1093/biomet/90.3.533

[18] Lee, W.J., Shi, J.Q. and Lee, Y.J. (2010) Approximate Conditional Inference in Mixed-Effects Models with Binary Data. *Computational Statistics & Data Analysis*, **54**, 173-184. https://doi.org/10.1016/j.csda.2009.07.027

[19] Di Caterina, C., Cortese, G. and Sartori, N. (2018) Monte Carlo Modified Profile Likelihood for Panel Data Models. arXiv:1801.02597v1.

[20] Lu, G.B. and Ades, A.E. (2006) Assessing Evidence Inconsistency in Mixed Treatment Comparisons. *Journal of the American Statistical Association*, **101**, 447-459. https://doi.org/10.1198/016214505000001302

[21] Weber, F., *et al.* (2020) Zero-Cell Corrections in Random-Effects Meta-Analyses. *Research Synthesis Methods*, **11**, 913-919. https://doi.org/10.1002/jrsm.1460

[22] Greco, T., *et al.* (2015) The Attractiveness of Network Meta-Analysis: A Comprehensive Systematic and Narrative Review. *Heart Lung and Vessels*, **7**, 133-142. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4476767/

[23] Wang, F.F., *et al.* (2018) Pharmacologic Therapy to Induce Weight Loss in Women Who Have Obesity/Overweight with Polycystic Ovary Syndrome: A Systematic Review and Net-Work Meta-Analysis. *Obesity Reviews*, **19**, 1424-1445. https://doi.org/10.1111/obr.12720

[24] Zhu, L.K., *et al.* (2019) A New Binomial Tree Method for European Options under the Jump Diffusion Model. *Journal of Applied Mathematics and Physics*, **7**, 3012-3021. https://doi.org/10.4236/jamp.2019.712211

[25] Kleinbaum, D.G. (1994) Logistic Regression. Springer, New York. https://doi.org/10.1007/978-1-4757-4108-7

[26] Collett, D. (1991) Modelling Binary Data (Hardback). Taylor & Francis, Oxfordshire.

[27] Severini, T.A. (1998) Likelihood Functions for Inference in the Presence of a Nuisance Parameter. *Biometrika*, **85**, 507-522. https://doi.org/10.1093/biomet/85.3.507

[28] Bellio, R. and Sartori, N. (2003) Extending Conditional Likelihood in Models for Stratified Binary Data. *Statistical Methods and Applications*, **12**, 121-132. https://doi.org/10.1007/s10260-003-0055-1

[29] Hamza, T.H., van Houwelingen, H.C. and Stijnen, T. (2008) The Binomial Distribution of Meta-Analysis Was Preferred to Model Within-Study Variability. *Journal of Clinical Epidemiology*, **61**, 41-51. https://doi.org/10.1016/j.jclinepi.2007.03.016

[30] Adeniran, A.T., *et al.* (2020) Derivation of Gaussian Probability Distribution: A New Approach. *Applied Mathematics*, **11**, 436-446. https://doi.org/10.4236/am.2020.116031

[31] Crouch, E.A.C. and Spiegelman, D. (1990) The Evaluation of Integrals of the Form $\int_{-\infty}^{+\infty} f(t)\exp(-t^2)\, dt$: Application to Logistic-Normal Models. *Journal of the American Statistical Association*, **85**, 464-469. https://doi.org/10.1080/01621459.1990.10476222

[32] Shi, J.Q. and Copas, J. (2002) Publication Bias and Meta-Analysis for 2×2 Tables: An Average Markov Chain Monte Carlo EM Algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 221-236. https://doi.org/10.1111/1467-9868.00334

[33] Liu, Q. and Pierce, D.A. (1994) A Note on Gauss-Hermite Quadrature. *Biometrika*, **81**, 624-629. https://doi.org/10.2307/2337136

[34] Shaobo, J. and Björn, A. (2020) A Note on the Accuracy of Adaptive Gauss-Hermite Quadrature. *Biometrika*, **107**, 737-744. https://doi.org/10.1093/biomet/asz080