Original Research Paper

A Distribution-Free Bayesian Approach for Indirect Comparisons

Johanna S. van Zyl and Jack D. Tubbs

Department of Statistical Science, Baylor University, USA

Abstract: The objective of this paper is to consider an indirect comparison between treatments B and C when each have been compared directly to treatment A in separate studies. Problems of this type are common in Network Meta-Analysis (NMA). A commonly used method assumes that the underlying data, Y, are normally distributed and $\mu = E(Y)$ is the measure of clinical effectiveness. The normal assumption is often violated. In addition, the sample sizes are not necessarily large. These conditions challenge the concept that a single location parameter, such as, the mean or median should be used as the measure of clinical effectiveness in the analysis. In this paper, we present an alternative approach where the Area Under the ROC Curve (AUC) is used as the measure of clinical effectiveness. Since the normal distribution may be uncertain, we use a distribution-free Bayesian mixtures of Finite Polya Trees (MFPT) model with the AUC in order to make the indirect comparison.

Keywords: AUC Regression, Simple Network Meta-Analysis, Mixture of Finite Polya Trees, ROC

Introduction

In a standard Meta-Analysis (MA), aggregate effects from different studies of a single treatment or the comparison between a single treatment and a common control are combined using a statistical model. The basic elements of the meta-analysis model include two parts, a model for the within study variability and a model for the between-study variability. In a fixed effects analysis, the between-study variability is assumed to be 0. A very commonly used MA is based on a normal-normal model for the two components (Viechtbauer, 2010). Models of this type allow one to estimate a pooled effect by combining the results from the individual studies.

Suppose the objective is to compare different treatments from several studies that are linked through common comparisons, then a Network Meta-Analysis (NMA) is used instead of a meta-analysis. In a NMA, indirect comparisons between treatments are possible whenever two or more studies compare different treatment effects with a common control. Li *et al*. (2018) presents a NMA for advanced or metastatic non-small cell lung cancer where each of the studies involve various therapeutics when compared with a common control, similar dosage of Docetaxel. In this paper, we use two studies, KEYNOTE-010 and CHECKMATE 057, from the larger NMA as illustrated in Fig. 1 as a motivating example where the objective is to make an indirect comparison between treatments Nivolumab and Pembrolizumab, when each have been compared directly to Docetaxel. An indirect comparison of treatments Nivolumab and Pembrolizumab is possible if the clinical measure of effectiveness, Y, is used for each direct comparison and if the assumption of consistency holds. It should be noted that the studies considered in Li *et al*. (2018) had the Overall Survival rate (OS) as a primary endpoint. A more common NMA assumes that $\mu = E[Y]$ is the measure of clinical effectiveness where the aggregate data for the study, $\bar{Y}$, have a normal distribution.

In a frequentist analysis, a common practice is to estimate the combined effect between all the studies using inverse weighting of the mean of the study effects (Sutton *et al*., 2000; Der Simonian and Laird, 2015). In a Bayesian analysis, a typical approach uses a hierarchical model with a normal distribution specified on the study effects (Sutton and Abrams, 2001). When the above assumption holds and $E[Y]$ is used as the measure of clinical effectiveness, then an indirect comparison of B and C is consistent since:

$$\mu_{CA} - \mu_{BA} = \left(\mu_C - \mu_A\right) - \left(\mu_B - \mu_A\right) = \mu_C - \mu_B. \tag{1}$$

Suppose the normal assumption does not hold, then the parameter $\mu = E(Y)$ may not be an appropriate measure of clinical effectiveness. This situation often happens when the density functions for the continuous random variable Y in the NMA are skewed or multi-modal.

The purpose of this paper is to investigate an alternative method for making the indirect comparison when the normal assumption does not hold. We propose two modifications to the traditional NMA for making an indirect comparison between treatments B and C with continuous data. The first modification involves the use the Area Under the receiver operating Curve (AUC) as the measure of clinical effectiveness when making a direct comparison between treatments A and B and A and C. The AUC_{AB} is chosen since the $AUC_{AB} = \Pr[Y_A > Y_B]$ does not depend upon the distribution of Y yet, involves the entire distribution of Y rather than a single location parameter, such as the mean or median. The second modification involves the use a distribution-free Bayesian approach when constructing the hierarchical model. A Mixture of a Finite Polya Tree (MFPT) model is used.

The outline for the paper is as follows. A description of the MFPT model is given in section 2. Section 3 provides a brief summary of the ROC and AUC for treatments A and B. The special cases when the response variable Y are normally distributed and the distribution-free MFPT estimates of the ROC and AUC are also presented in Section 3. The network models for an indirect comparison of treatments B and C are given in Section 4. A simulation study comparing the performance of the proposed method with the traditional normal model is presented in Section 5. The proposed method is illustrated using data from two Diabetic Macular Edema (DME) studies in Section 6. A summary and discussion is given in section 7.

Mixtures of Finite Polya Trees

A flexible class of methods given in Bayesian nonparametric literature have been proposed for describing distributions that exhibit features such as multi-modality and skewness. One such model is a mixture of finite Polya Trees. As in any Bayesian approach, the posterior density function is proportional to the product of the likelihood and a joint prior distribution. These components for the Polya tree model are presented in the next two sections.

The Likelihood

Christensen *et al.* (2008) illustrate mixtures of finite Polya trees that are centered on a $N(\mu,\sigma^2)$ family of distributions. Flexibility from the normal distribution family is achieved by reweighting sections of the distribution through stagewise partitioning of the support. At each successive partition, the conditional probability of being above and below the split is altered through the introduction of a new set of parameters. The procedure given in

Christensen *et al.* (2008) enables one to express the density of Y at the Mth stage as:

$$f(y \mid \mu,\sigma^2,\Theta_m) = \phi(y;\mu,\sigma^2)r(y \mid \mu,\sigma^2), \qquad (2)$$

where, $\Theta_m = \{\boldsymbol{\theta}_{ms}; \ m = 1,...,M, \ s = 1,...,2^m\}$ and the weighting function is:

$$r\left(y \mid \mu,\sigma^2\right) = 2^M \prod_{\theta_{ms} \in \Theta(y)} \theta_{ms}$$

and $\Theta(y)$ is the set of all probabilities, θ_{ms}, associated with the bins that contain y. The likelihood based on different base distributions follows a similar derivation where $\square(\cdot; \mu, \sigma^2)$ in Equation 2 is replaced by the density function for the new base distribution.

Prior Distributions

The parameters of the Polya tree model are divided into two sections. The first, denoted by Ψ, consists of the indexing parameters for the base distribution. When $Y \sim N(\mu, \sigma^2)$, then Ψ consists of μ and σ^2. The second section consists of the bin probabilities $\theta_{ms} \in \Theta$ for $m = 1,..., M$ and $s = 1,...,2^m$ where Θ contains $2^{M+1} - 2$ parameters. The flexibility of the model is achieved through the choice of the bin probabilities.

Suppose that one has expert opinion that one wishes to incorporate into the model. For example, suppose at the first stage that one has expert knowledge concerning $\Pr[Y \leq \mu]$. At the second stage, one can specify $\Pr[Y \leq q_1 \mid Y \leq \mu]$ and $\Pr[Y > q_1 \mid Y \leq \mu]$ where q_1 is the first quantile for Y. This process of eliciting information can be performed for each of the M stages.

If expert opinion is not available then an alternative approach uses the beta distribution as a reference prior. That is, let:

$$\theta_{m,2s} \sim Beta\left(\alpha_{m,2s},\alpha_{m,2s+1}\right)$$

and $\theta_{m,2s+1} = 1 - \theta_{m,2s}$ for $m = 1,..., M$ and $s = 1,3,5,..., 2^m - 1$ where the pairs of bin probabilities are assumed to be independent. In which case, one exchanges specifying priors on Θ with the hyperparameters denoted by α. If the prior information is unknown, a reference prior in the form of $\alpha_{ms} = c\rho(m)$ is commonly used where c reflects the strength of the prior belief in the original parametric family and $\rho(\cdot)$ is a nondecreasing function. In which case, prior specification on each α_{ms} is replaced by specifying a single prior on c. This formulation ensures that the beta prior is symmetric with decreasing variability as the tree level m increases. In which case, the conditional bin probabilities at lower tree levels are less likely to deviate from the base distribution than higher tree levels.

Fixed priors specified on the parameters of the base distribution result in a finite Polya tree. A Finite Polya Tree (FPT) with M levels, reference prior $c\rho(\cdot)$, parametric centering family F_Ψ^* and priors on the θ_{ms} is denoted as:

$$F \sim PT_M(F_\Psi^*, c, \rho). \tag{3}$$

The posterior density using a finite Polya tree model is potentially discontinuous at the boundaries of the bins. A solution to this problem is to specify a prior distribution on the parameters of the base distribution (Christensen *et al.* (2008)). That is, by introducing a prior on $\Psi = (\mu, \sigma^2)$ when $Y \sim N(\mu, \sigma^2)$, the median, quartiles and all subsequent splits are no longer fixed but are random. A Mixture of Finite Polya Trees (MFPT) is given by:

$$F \sim \int PT_M(F_\Psi^*, c, \rho) dp(\Psi) \tag{4}$$

which is obtained after generating a set of $\{\theta_{ms}\}$ from the prior distribution on $\{\theta_{ms}\}$ and averaging over the prior on the support Ψ. The MFPT model for the distribution, F, is denoted as:

$$F \sim MFPT_M(F_\Psi^*, c, \rho). \tag{5}$$

There are situations when $E[Y] = \mu$ may not be an appropriate measure of clinical effectiveness, in which case, we propose using the AUC. In the next section, a non-parametric estimate of the AUC is presented when the model described in Equation 5 holds.

3 ROC and AUC

The Receiver Operating Curve (ROC) and area under the ROC (AUC) can be used to compare two distributions. In the MFPT model, the relationship between the ROC and survival functions for each treatment are of special interest as given by:

$$ROC_{AB}(t) = S_A\left(S_B^{-1}(t)\right) = 1 - F_A\left(F_B^{-1}(t)\right), \tag{6}$$

where, $S_g(\cdot) = 1 - F_g(\cdot)$ is the survival function for treatment g, $g \in \{A, B\}$ and $t \in [0,1]$ (Pepe *et al.*, 2009). The AUC is:

$$AUC_{AB} = \int_0^1 ROC_{AB}(t) dt. \tag{7}$$

Bayesian non-parametric models in ROC analysis can be found in Branscum *et al.* (2008; 2015) and Hanson, (2006).

ROC and AUC for the Binormal Model

In the special case when $Y_A \sim N(\mu_A, \sigma_A)$ and $Y_B \sim N(\mu_B, \sigma_B)$, the ROC and AUC have a closed form expression (Dorfman and Alf (1968) and Faraggi and Reiser (2002)) given by:

$$ROC_{AB}(t) = \Phi\left[a + b \cdot \Phi^{-1}(t)\right],$$

and:

$$AUC_{AB} = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right),$$

where Φ is CDF for the standard normal distribution function:

$$a = \frac{\mu_A - \mu_B}{\sigma_A}, \ and \ b = \frac{\sigma_B}{\sigma_A}.$$

ROC and AUC for the MFPT Model

The MFPT model provides distribution-free estimates for the ROC by replacing F_g and F_g^{-1} in Equation 6 with the posterior distribution function and inverse of the posterior distribution function. The posterior distribution function, F_g, for treatment g is:

$$\begin{aligned} F_g(y \mid \Theta, \Psi) &= \sum_{i=1}^{k(y)-1} p_g(i) \\ &+ p_g(k(y))\left[2^M F_{\Psi,g}^*(y) - k(y) + 1\right], \end{aligned} \tag{8}$$

where, $k(y)$ indicates the bin at level M that contains observation y from group g, $F_\Psi^*(y)$ is the base distribution function and:

$$p_g(i) = \prod_{\theta_{gms} \in \Theta_g(i)} \theta_{gms}$$

is the unconditional posterior probability of bin i at level M for $i = 1, ..., k(y)$ (Hanson, 2006). $\Theta_g(i)$ is the set of probabilities, θ_{gms}, for group g, $m = 1, ..., M$ and where s indicates the bin at the mth level that contains the ith bin at level M.

The inverse posterior distribution function, F_g^{-1}, for treatment g is:

$$F_g^{-1}(t \mid \Theta, \Psi) = F_{\Psi,g}^{-1}\left[\frac{t - \sum_{i=1}^Q p_g(i) + Q \cdot p_g(Q)}{2^M \cdot p_g(Q)}\right] \tag{9}$$

where, Q is such that $\sum_{i=1}^{Q-1} p_g(i) < t < \sum_{i=1}^{Q} p_g(i)$ (Hanson, 2006).

The posterior density function is given by:

$$f_g(y \mid \Theta, \Psi) = 2^M p_g(k(y)) f_{\Psi,g}^*(y), \tag{10}$$

which is similar to the likelihood in Equation 2 where $p_g k(y)$ is the posterior unconditional probability of the bin at level M and $f_{\Psi,g}^*(y)$ is the density function associated with the base distribution $F_{\Psi,g}^*(y)$.

Numerical integration of the ROC is used to provide an estimate for the AUC.

Network Models with Random Effects

The two primary methods used to combine information from studies are fixed and random effects models (Stangl and Berry, 2000). A fixed effects model is used when the studies are homogeneous, in which case the studies do not differ in terms of the underlying study population, patient selection criteria and application of therapies. The fixed effects model is highly restrictive, in that, it assumes that patients are exchangeable between studies. A random effects model is appropriate under a less rigid assumption of homogeneity, where each study effect is modeled as a random effect from a common distribution of study effects.

A random effects model is used in the scenario described in Fig. 1 where the direct comparisons with effect A are estimated from two studies, one with B and one with C.

Bayesian Normal Network Model

In this section, the commonly used Bayesian parametric normal model is described where one assumes that the observations for the treatment and control groups are random samples from normal populations. That is, let:

$$y_{ijk} \sim N(\mu_{jk}, \sigma_k) \tag{11}$$

for subject $i = 1,..., N_{jk}$, study $j = 1,..., J_k$ and treatment $k \in \{A,B,C\}$ where $J_A = 2$ and $J_B = J_C = 1$. The mean, μ_{jk}, in a random effects model is:

$$g\left[E\left(y_{ijk} \mid \beta_{jA}, \beta_k\right)\right] = \beta_{jA} + \beta_k \tag{12}$$

where, β_{jA} is the random effect for the control group and β_k is the treatment effect for $k \in \{B,C\}$. The random effects model assumes that the control group effect, β_{jA}, is randomly selected from a normal population of effects with mean m_A and standard deviation, s_A. Independent uniform priors, $\sigma_{jk} \sim U(0, B_{\sigma k})$, are specified on the standard deviations in Equation 11 and independent normal priors on the regression coefficients, β_A and β_k, in Equation 12. A hyperprior, $m_A \sim N(0, \tau_{m_A})$, is specified on the mean of the random effects on A and a uniform hyperprior, $S_A \sim U(0, B_{S_A})$, on the standard deviation on A. The Bayesian normal random effects model can be written as:

$$y_{ijk} \mid \mu_{jk}, \sigma_k \sim N(\mu_{jk}, \sigma_k) \tag{13}$$

$$g\left[E\left(y_{ijk} \mid \beta_{jA}, \beta_k\right)\right] = \beta_{jA} + \beta_k \tag{14}$$

with priors on the model parameters:

$$\beta_{jA} \mid m_A, s_A \sim N(m_A, s_A) \tag{15}$$

$$\beta_k \sim N(m_k, s_k) \tag{16}$$

and hyperpriors on the parameters of the random effects:

$$\sigma_k \sim U\left(0, B_{\sigma_k}\right) \tag{17}$$

$$m_A \sim N\left(0, \tau_{m_A}\right) \text{ and } s_A \sim U\left(0, B_{s_A}\right). \tag{18}$$

MFPT Network Model

In this section, the MFPT model is specified when the effect on the control group, A, is assumed to be random. That is, let:

$$y_{ijA} \sim MFPT_M\left(N\left(\mu_{jA}, \sigma_A\right), c_{jA}, \rho\right). \tag{19}$$

and treatment effects for groups, $k \in \{B,C\}$:

$$y_{ik} \sim MFPT_M\left(N\left(\mu_k, \sigma_k\right), c_k, \rho\right). \tag{20}$$

Independent normal priors are specified on the means, $\mu_{jA} \sim N(m_A, s_A)$ and $\mu_k \sim N(m_k, s_k)$ and independent uniform priors on the standard deviations, $\sigma_A \sim U(0, B_{\sigma_A})$ and $\sigma_k \sim U(0, B_{\sigma_k})$. Similarly to the binormal model, hyperpriors are specified on the mean of the random effects on A, $m_A \sim N(0, \tau_{m_A})$ and a uniform hyperprior on the standard deviation, $s_A \sim U(0, B_{s_A})$. The weighting factors, c_{jA} and c_k, are fixed constants and a non-decreasing function, $\rho(m) = 2^{-vm}$ with $v = -1$ or -2 is used.
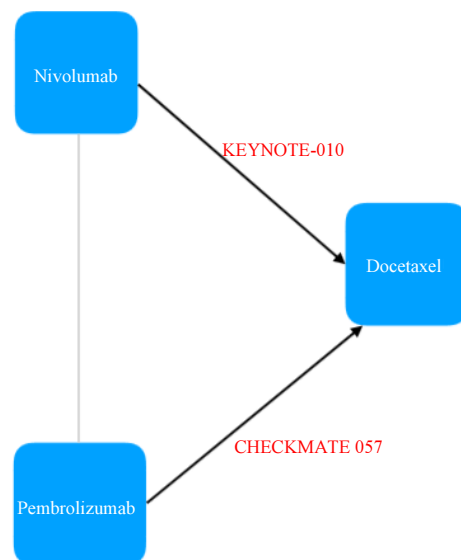

Fig.1: Illustration of a simple network configuration of treatments, B = Nivolumab and C = Pembrolizumab with A = Docetaxel

The Bayesian MFPT random effects model on A is written as:

$$Y_{ijA} \mid G_{jA\theta_{jA}}, c_{jA} \sim MFPT_M\left(G_{jA\theta_{jA}}, c_{jA}\right) \tag{21}$$

$$G_{jA\theta_{jA}} \mid \mu_{jA}, \sigma_A \sim N\left(\mu_{jA}, \sigma_A\right), \tag{22}$$

with priors on the model parameters:

$$\mu_{jA} \sim N\left(m_A, s_A\right) \ and \ \sigma_A \sim U\left(0, B_{\sigma_A}\right), \tag{23}$$

and hyperpriors on the random effects parameters:

$$m_A \sim N\left(0, \tau_{m_A}\right) \ and \ s_A \sim U\left(0, B_{s_A}\right). \tag{24}$$

The model on the kth treatment effect is written as:

$$Y_{ijk} \mid G_{k\theta_k}, c_k \sim MFPT_M\left(G_{k\theta_k}, c_k\right) \tag{25}$$

$$G_{k\theta_k} \mid \mu_k, \sigma_k \sim N\left(\mu_k, \sigma_k\right), \tag{26}$$

with prior structure:

$$\mu_k \sim N\left(m_k, s_k\right) \ and \ \sigma_k \sim U\left(0, B_{\sigma_k}\right) \tag{27}$$

Simulation Study

A simulation is performed to evaluate the performance of the MFPT model using the AUC as a measure of clinical effectiveness. The type I error in testing the indirect comparison of treatments B and C will be of particular interest.

The null hypothesis is:

$$H_0 : R_{CB} = \frac{AUC_{CA}}{AUC_{BA}} = 1, \tag{28}$$

when the AUC is the measure of effectiveness in the MFPT model. The ratio of AUCs is chosen instead of a difference in AUCs for the sake of interpretability. For example, if $R_{CB} = 1.2$ we can interpret this as $P(C>A)$ is 1.2 times larger than $P(B>A)$.

The simulations involved two data types, normal and Gumbel (extreme value). The pdf for the Gumbel distribution is:

$$g(v) = \lambda^{-1} exp\left\{\frac{-v+\mu}{\lambda}\right\} exp\left\{-exp\left\{\frac{-v+\mu}{\lambda}\right\}\right\} \tag{29}$$

for $v \in R$, the mode, μ and scale parameter, λ. When the data are normal, the indirect comparison of B and C using the MFPT model is compared with the traditional binormal model given in Section 4.1 using μ as a measure of clinical effectiveness and the null hypothesis given by:

$$H_0 : \mu_B = \mu_C. \tag{30}$$

Table 1 summarizes the design parameters and true values for the AUC. The design parameters for the three designs using the extreme value data are summarized in Table 2. The performance of the MFPT model is evaluated using both data types.

Rejection of the null hypotheses in Equation 30 and 28 are based on the $(1-\alpha)$ 100% equal-tailed credible set from the Markov Chain Monte Carlo simulation to control the type I error at $\alpha = 0.05$. For design D1 and D3, the null hypothesis of equality of B and C should not be rejected; whereas in design D2 the null hypothesis should be rejected. The percentage of times the null hypothesis are rejected for the normal and extreme value simulations using $N = 250$ iterations are reported in Table 3 and 4.

Results with Normal Data

In this section, the comparison of B and C through the null hypothesis in Equation 30 is appropriate using the estimates from the binormal random effects model. In D1 where the null hypothesis should not be rejected more than 5% of the times, the rejection rate of Equation 30 is 5% using the binormal model. Using the ratio of AUC's in Equation 28 has the same rejection rate as that found in the binormal model. The MFPT model performs similarly with a rejection rate of 6% and 4% testing Equation 28 and 30, respectively. In D2 where the null hypothesis is false, the null hypotheses in Equation 28 and 30 are correctly rejected 100% of the times using the binormal and MFPT models. In D3, the rejection rate of Equation 30 and 28 is 3% and 4%, respectively, using the binormal model. The MFPT model has a rejection rate of 3% and 5% testing Equation 30 and 28, respectively. The type I error is controlled close to the nominal rate using each measure of clinical effectiveness with the binormal and MFPT models.

Results with Extreme Value Data

In this section, it is no longer appropriate to use the binormal model or difference in means to compare B and C. Instead, the results for the rejection rate of the null hypothesis with the MFPT model using a ratio of AUC's is reported in Table 4. In D1, the rejection rate of Equation 28 is 1%. The null hypothesis is correctly rejected 100% of the times in D2. In D3, where the null hypothesis should not be rejected, the rejection rate is 2% using the ratio of AUC's from the MFPT model. The random effects MFPT model using the ratio of AUC's is slightly conservative testing Equation 28.

The results presented here are very representative of the results found when doing a more exhaustive simulation study.

In the next section, we illustrate the methods given above with data from two published clinical trial with diabetic patients with macular edema.

Table 1: Normal simulation design parameters: $n = 100$ and $\sigma = 1$

Design	μ_A	μ_B	μ_C	AUC_{BA}	AUC_{CA}	AUC_{CB}
D1	0	0	0	0.5	0.5	0.5
D2	0	0	2.5	0.5	0.96	0.96
D3	0	2.5	2.5	0.96	0.96	0.5

Table 2: Extreme value simulation design parameters: $n = 100$ and $\lambda = 1$

Design	μ_A	μ_B	μ_C	AUC_{BA}	AUC_{CA}	AUC_{CB}
D1	0	0	0	0.5	0.5	0.5
D2	0	0	3.18	0.5	0.96	0.96
D3	0	3.18	3.18	0.96	0.96	0.5

Table 3: Indirect comparison of B and C in the normal simulation

Model	Parameter	Iterations	D1	D2	D3
Binormal	$\mu_C - \mu_B$	250	5	100	3
	AUC_{CA}/AUC_{BA}	250	5	100	4
MFPT	$\mu_C - \mu_B$	250	6	100	3
	AUC_{CA}/AUC_{BA}	250	4	100	5

Table 4: Indirect comparison of B and C in the extremevalue simulation

Model	Parameter	Iterations	D1	D2	D3
MFPT	AUC_{CA}/AUC_{BA}	250	1	100	2

Example with Diabetic Macular Edema

Diabetic Macular Edema (DME) is a form of diabetic retinopathy that impairs the central vision. The current standard treatment is focal/grid photocoagulation (laser therapy) from which vision stabilization is an optimistic outcome where a primary endpoint is the change from baseline in the visual acuity score and a secondary endpoint is the change from baseline in central retinal thickness (CRT). Two randomized controlled trials, RESTORE (NCT00687804) and VIVID (NCT01331681), compared the efficacy of Ranibizumab (R) and Intravitreal Aflibercept (IAI) to laser therapy (control). A description of RESTORE and VIVID and results for the secondary endpoint, change from baseline in CRT, are given in Sections 6.1 and 6.2. A comparison of the efficacy of R and IAI is of interest. However, there are no available trials that directly compare the two therapeutics. The random effects MFPT model for an indirect comparison between R and IAI is given in Section 6.3.

RESTORE Trial

Novartis, (2013) performed a randomized, double-blind, multi-center, phase 3 trial with a laser control arm and two treatment arms. The laser photocoagulation treatment was administered at the commencement of treatment. If necessary, additional treatments were administered at intervals of at least 3 months. Ranibizumab 0.5 mg was administered monthly by intravitreal injection in the study eye for 3 months. The treatment was suspended following the third injection if one of the specified criteria reported by Novartis, (2013) was met. The study results for the change from baseline

at month 12 in central retinal thickness of the study eye is given in Table 5.

VIVID Trial

A double-blind, randomized, phase 3 study of the efficacy and safety of IAI administration of VEGF TRap-Eye with a laser control arm in patients with DME was completed by Bayer and Regeneron Pharmaceuticals (2016). The laser control arm was administered at baseline, with additional treatments at intervals of at least 3 months when necessary. Patients in the IAI 2Q4 group received 2mg Intravitreal aflibercept injection every 4 weeks. The study results for the change from baseline at month 12 in CRT of the study eye is given in Table 6.

Indirect Comparison of RESTORE and VIVID Trials

The summary statistics from RESTORE and VIVID trials were used to simulate the data presented in this section. The treatment arms of interest are Ranibizumab 0.5mg from the RESTORE trial and IAI 2Q4 from the VIVID trial. Data are adjusted so that a larger response correspond to a more effective treatment. Summary statistics for the simulated data are given in. Table 7 note that the data are skewed with the median less than the mean.

Figure 2 illustrates the sample densities of the four treatment arms. The objective of the study is to make an indirect comparison between Ranibizumab 0.5mg and IAI 2Q4. The random effects MFPT model used to make an indirect comparison between B and C has two components. The model for the random effects on A and the model for the treatment effects.
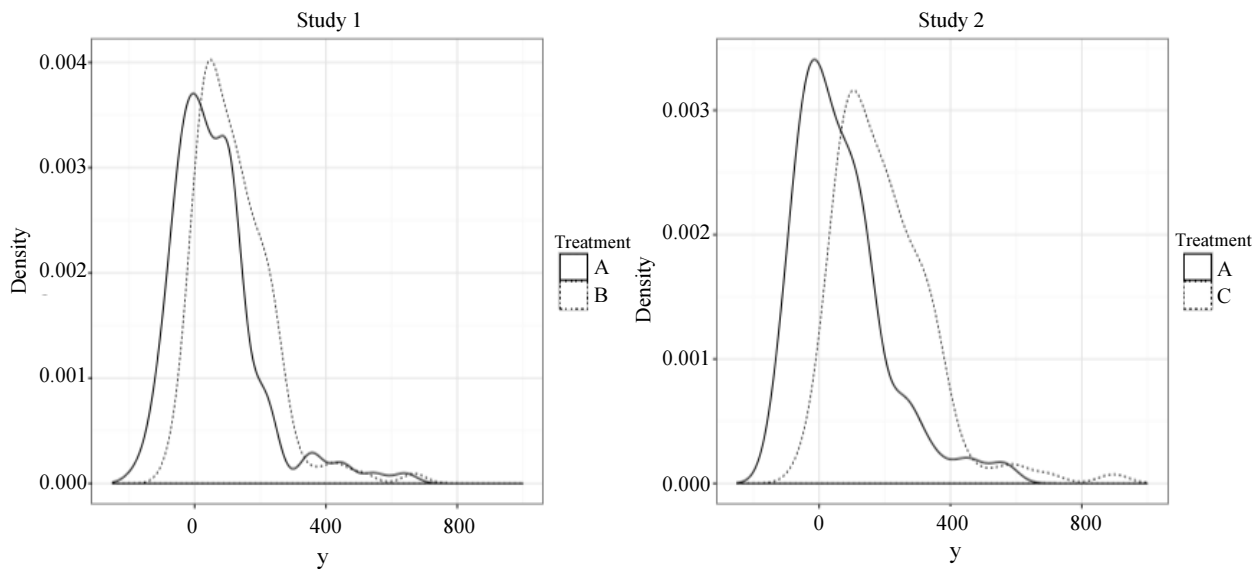
Fig. 2: Sample densities for simulated change from baseline in CRT in the RESOLVE (Study 1) [laser (A), Ranibizumab 0.5mg (B)] and in the VIVID (Study 2) [laser (A), IAI 2Q4 (C)]

Table 5: Mean change from baseline at month 12

Variable	Laser	Ranibizumab 0.5 mg
Number of participants	113	116
Mean Baseline (SE)	412.4 (124.53)	427.1 (118.42)
Value at Month 12 (SE)	351.1 (139.91)	308.4 (112.26)
Change from Baseline (SE)	-61.3 (132.29)	-118.7 (115.07)

Table 6: Mean change from baseline at week 52.

Variable	Control	IAI 2Q4
Number of participants	132	135
Change from Baseline (SE)	-66.2 (138.99)	-195.0 (146.59)

Table 7: Summary statistics of simulated data

Treatment	N	Mean (SE)	Median (IQR)
A1: Laser (RESTORE)	120	61.30 (132.29)	35.97 (135.42)
B: Ranibizumab	120	118.70 (115.08)	100.10 (139.25)
A2: Laser (VIVID)	120	66.19 (138.95)	36.53 (161.70)
C: IAI 2Q4	120	195.01 (146.59)	171.31 (177.40)

on B and C. The random effects component is summarized as:

$$Y_{ijA} \mid G_{jA\theta_{jA}}, c_{jA} \sim MFPT_M\left(G_{jA\theta_{jA}}, c_{jA}\right) \tag{31}$$

$$G_{jA\theta_{jA}} \mid \mu_{jA}, \sigma_A \sim N\left(\mu_{jA}, \sigma_A\right) \tag{32}$$

$$\mu_{jA} \mid m_A, s_A \sim N\left(m_A, s_A\right) \tag{33}$$

$$\sigma_A \sim U\left(0, 250\right) \tag{34}$$

$$m_A \sim N\left(58.4, 177.60\right) \tag{35}$$

$$s_A \sim U\left(0, 250\right) \tag{36}$$

and the model on the treatment effects as:

$$Y_{ijk} \mid G_{k\theta_k}, c_k \sim MFPT_M\left(G_{k\theta_k}, c_k\right) \tag{37}$$

$$G_{k\theta_k} \mid \mu_k, \sigma_k \sim N\left(\mu_k, \sigma_k\right) \tag{38}$$

$$\mu_B \sim N\left(187.6, 150\right) \tag{39}$$

$$\mu_C \sim N\left(227.4, 150\right) \tag{40}$$

$$\sigma_k \sim U\left(0, 250\right) \tag{41}$$

for $k \in \{B,C\}$. Prior selection on the random effects component, m_A and treatment means, μ_B and μ_C are based on summary statistics from two additional trials; RESOLVE (NCT00284050) and DA VINCI (NCT00789477). These trials are Phase II studies for the safety and efficacy of the treatments under consideration (Novartis, 2011; Bayer and Regeneron Pharmaceuticals, 2016) A sensitivity analysis for the specification of the upper bound on the uniform prior on s_A and σ_k in Equation 36 and 41 is illustrated in Fig. 3. The MFPT model for the network of treatments is fitted with a range

of upper bounds, B, on the uniform priors. Figure 3 presents the 95% credible set on R_{CB}, the ratio of AUCs. This example is an interesting scenario where the choice of maxB potentially influences the inference illustrating the importance of conducting a sensitivity analysis on the prior selection. The width of the credible sets on the ratio of AUCs conservatively has stabilized beyond $maxB = 230$ indicating that the choice of a U(0, 250) prior does not influence the posterior results in terms of the variability around the estimate of the ratio of AUCs.
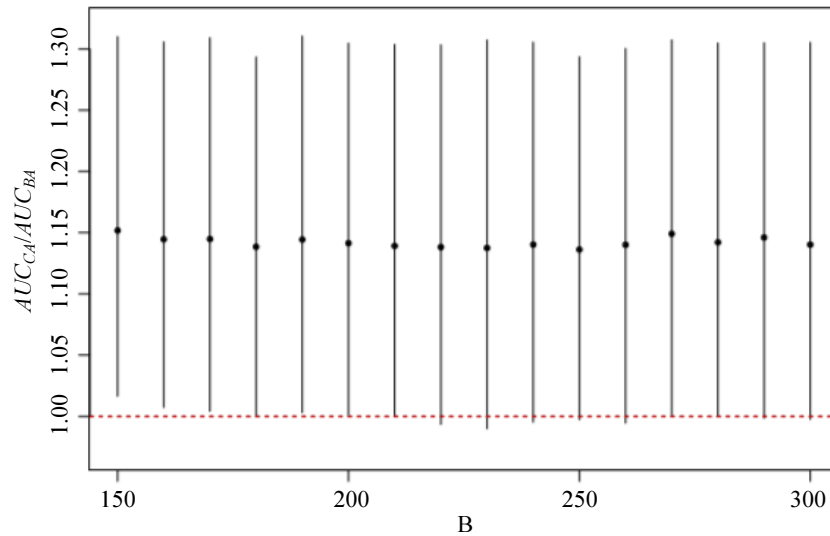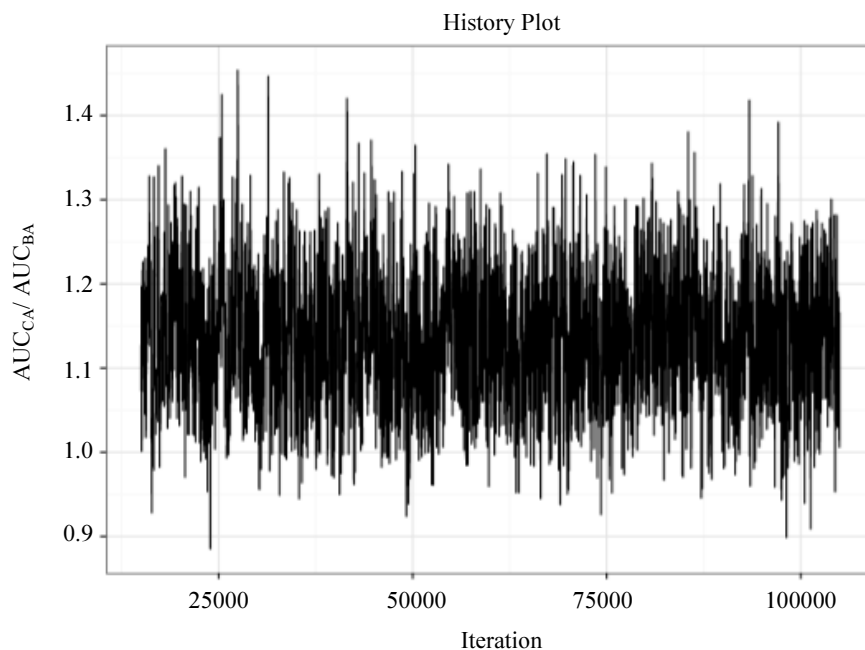

Fig. 3: Sensitivity analysis of the prior specification on the upper bound of the uniform prior on the standard deviation of s_A, σ_B and σ_C.
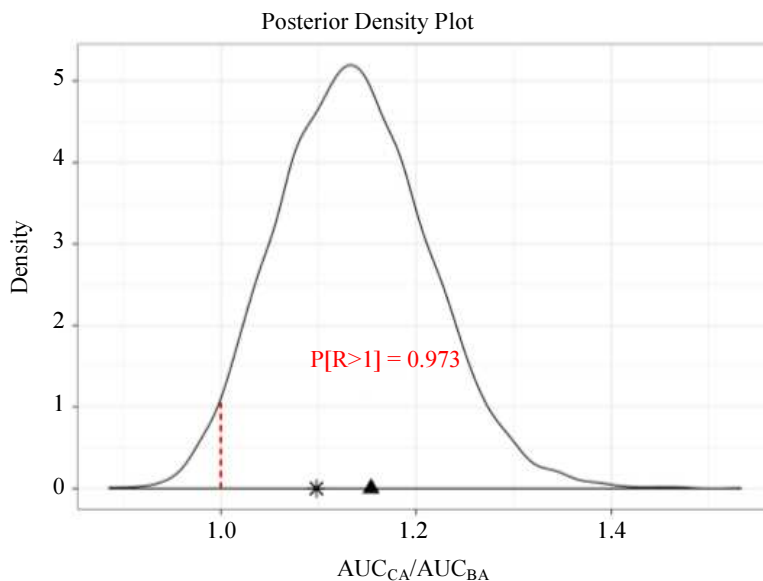

143

Fig.4: Posterior history and density plots for the ratio of AUCs comparing intravitreal aflibercept (C) to Ranibizumab (B). The triangle indicates the empirical ratio of AUCs and the asterisk indicates the true ratio of AUCs

The model given in Equation 31-41 is fitted using SAS Proc MCMC with a burn-in of 15000 iterations, 90000 additional iterations and thinning of 15 resulting in 6000 final iterations. History, autocorrelation and posterior density plots are checked for convergence of each parameter. Figure 4 presents satisfactory convergence diagnostics for R_{CB}. The posterior density plot is centered over the true ratio of AUCs as indicated by an asterisk. The empirical ratio of AUCs is represented by the triangle. The posterior probability that R_{CB} is greater than 1 is 97.3%. Recall that having $R_{CB} > 1$ indicates that the numerator effect (IAI) is more effective then the denominator effect (R) when compared with the laser control.

Discussion

This paper presented a method for making indirect comparisons between treatments when the commonly used normal method for the NMA framework was inappropriate. The method made use of a distribution-free Bayesian MFPT to develop a random effects model when the AUC is the measure of clinical effectiveness. A simulation study is performed to evaluate the performance of the MFPT model using the AUC as a measure of clinical effectiveness. The type I errors for the proposed method is comparable to those found using the binormal model when the data are simulated from a normal distribution. The proposed method performed well when the data were simulated from an extreme value distribution. In both simulations the method was able to control the type I error when testing the indirect comparison of the two treatments.

There is one caveat to the ratio of AUCs. For the case where there is a complete separation between the treatment and control group, the AUC is 1 regardless of the degree of separation between the control and treatment groups. If this is the case for both studies, the ratio of AUCs will not work since the ratio equals 1 and a conclusion of no significant difference is made when it is possible that there might be a difference between the two groups. Since this case is so extreme, one suspects that a simple visual inspection of the respective histograms or density functions will keep one from making this error.

The proposed method was demonstrated using two clinical trials for treatment of Diabetic Macular Edema (DME). The two trials, RESTORE and VIVID, separately investigated the safety and efficacy of Ranibizumab and Aflibercept as compared to a laser control. The example illustrated the capabilities of the proposed method of making an indirect comparison between two treatments that have been compared separately to common control. Since, we used the summary statistics from the two studies to generate simulated data we were not overly concerned with the outcome of this exercise. It should be noted, that the real value of the procedure presented in this paper might be in the in-house investigations that pharmas undertake when trying to access theirs (and other's) drug pipeline.

In conclusion, the MFPT model provides a reasonable alternative for an indirect comparison within a network meta-analysis when the normality assumption for the measure of clinical effectiveness might not hold. The use of the ratio of AUCs has the added benefit of providing a stochastic interpretation, $AUC_{AB} = \Pr[Y_B > Y_A]$, instead of using the difference in location parameters for Y in the two treatment groups. An additional advantage of using the AUCs is the ability to include covariates, X, $AUC_{AB}(X) =$

$\Pr[Y_B > Y_A \mid X]$. We were not able to demonstrate the use of covariates with the DME examples since we did not have access to the subject-wise data. It should be noted that modeling the AUC as a function of a continuous covariates X is somewhat problematic. A potentially better approach would be to model the ROC as a function of covariates X (Alonzo and Pepe (2002); Cai, (2004); Rodriguez-Alvarez *et al*. (2011) and Stanley and Tubbs, (2018). Stanley and Tubbs, (2018) utilized the relationship that the $ROC_{AB|X}(t) = \Pr[PV_{AB} \le t \mid X]$ where PV_{AB} is the conditional placement value when an observation from *B* is placed into the survival curve for group A. One can perform the indirect comparison by testing the hypothesis that the CDFs defined by the two ROCs are equal using standard test of fit procedures. This work will appear in a separate manuscript.

Conclusion

The MFPT model provides a reasonable alternative for an indirect comparison within a network meta-analysis when the normality assumption for the measure of clinical effectiveness might not hold. The use of the ratio of AUCs has the added benefit of providing a stochastic interpretation, $AUC_{AB} = \Pr[Y_B > Y_A]$, instead of using the difference in location parameters for *Y* in the two treatment groups.

Acknowledgment

The authors wish to thank Professor James Stamey for providing meaningful insight into our proposed method of using the ratio of AUCs as a means of accessing clinical measure of effectiveness.

Author's Contributions

Johanna S van Zyl: Participated in all steps in conducting research, computer programming, analyzing the real data example, reporting and drafted the manuscript.

Jack D. Tubbs: Contributed to oversight of all stages of the research, to the writing of the manuscript and approving the final manuscript.

Ethics

The authors declare that there is no conflict interest regarding the publication of this manuscript.

References

Alonzo, T. A. and M.S. Pepe, 2002. Distribution-free ROC analysis using binary regression techniques. Biostatistics, 3: 421-432.

Bayer and Regeneron Pharmaceuticals, 2016. Intravitreal aflibercept injection in vision impairment due to dme (vivid-dme)-study results. Tech. Rep. Bethesda (MD): National Library of Medicine.

Branscum, A.J., W.O. Johnson, T.E. and A.T. Baron, 2015. Flexible regression models for ROC and risk analysis, with or without a gold standard. Statistics Med., 34: 3997-4015.

Branscum, A.J., W.O. Johnson, T.E. Hanson and I.A. Gardner, 2008. Bayesian semiparametric ROC curve estimation and disease diagnosis. Statistics Med., 27: 2474-2496.

Cai, T., 2004. Semi-parametric ROC regression analysis with placement values. Biostatistics, 5: 45-60.

Christensen, R., T. Hanson and A. Jara, 2008. Parametric nonparametric statistics: An Introduction to mixtures of finite polya trees. Am. Statistician, 62: 296- 306.

Der Simonian, R. and N. Laird, 2015. Meta-analysis in clinical trials revisited. Contemporary Clinical Trials, 45: 139-145.

Dorfman, D.D. and E. Alf, 1968. Maximum likelihood estimation of parameters of signal detection theory-A direct solution. Psychometrika, 33: 117-124.

Faraggi, D. and B. Reiser, 2002. Estimation of the area under the ROC curve. Statistics Med., 21: 3093-3106.

Hanson, T.E., 2006. Inference for mixtures of finite polya tree models. J. Am. Statistical Assoc., 101: 1548-1565.

Li, J., M. Sasane, J. Zhao and V. Horton, 2018. Comparative efficacy of treatments for previously treated advanced of metastatic non-small cell lung cancer: A Network Meta-Analysis. Adv Ther, pp: 35.

Novartis, 2011. Safety and efficacy of Ranibizumab in diabetic macular edema with center involvement (resolve) - study results. Tech. Rep., Bethesda (MD):

Pepe, M., G. Longton and H. Janes, 2009. Estimation and comparison of receiver operating characteristic curves. Stata. J., 9: 1.

Rodriguez-Alvarez, M.X., P.G. Tahoces, C. Cadarso-Suarez M.J. Lado, 2011. Comparative study of roc regression techniques-applications for the computer-aided diagnostic system in breast cancer detection. Computational Statistics Data Analysis, 55: 888-902.

Stangl, D. and D.A. Berry, 2000. Meta-Analysis in Medicine and Health Policy. CRC Press, Google-Books-ID: Zj7PoheDEtsC.

Stanley, S. and J. Tubbs, 2018. Beta regression for modeling a covariate adjusted ROC. J. Applied Math. Stat., 6: 110-118.

Sutton, A.J. and K.R. Abrams, 2001. Bayesian methods in meta-analysis and evidence synthesis. Stat. Methods Med. Res., 10: 277-303. PMID: 11491414.

Sutton, A.J., K.R. Abrams, D.R. Jones, T.A. Sheldon and F. Song, 2000. Methods for meta-analysis in medical research. John Wiley and Fujian Song, ISBN: 978-0-471-49066-1, pp: 348.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the meta for package. J. Stat. Software, 36: 1-48.