# 4

# A Bayesian Meta-analysis of Randomized Mega-trials for the Choice of Thrombolytic Agents in Acute Myocardial Infarction

**James Brophy**
*University of Montreal, Montreal, Quebec, Canada*

**Lawrence Joseph**
*McGill University, Montreal, Quebec, Canada*

## Summary

Thrombolytic agents have been shown by several randomized trials to substantially reduce mortality following an acute myocardial infarction. Nevertheless, the choice of thrombolytic agents remains unclear. Three large randomized trials have reported results which directly compare two agents, streptokinase (SK) and tissue-plasminogen activator (t-PA). In this chapter, we evaluate the evidence from these trials via increasingly complex Bayesian meta-analytic models, including bias adjustments for possible differences between the trials. Our analyses suggest that the clinical superiority of t-PA over SK remains uncertain.

## I. INTRODUCTION

Several randomized clinical trials (Gruppo Italiano per lo Studio della Streptochinase Nell'Infarto Miocardico (GISSI) (1), Wilcox et al. (2), and Fibrinolytic Therapy Trialists' (FTT) Collaborative Group (3)) have shown that thrombolytic agents improve 30-day survival by approximately 30 percentage points following an acute myocardial infarction (AMI). Thrombolytic agents are plasminogen activators which convert plasminogen, a proenzyme, to plasmin, an enzyme capable of cleaving fibrin and producing clot lysis. Streptokinase, SK, is the oldest identified plasminogen activator and was the first commercially available thrombolytic agent. SK not only acts on clots but also on circulating fibrinogen, giving rise to systemic fibrinogenolysis. Tissue plasminogen activator (t-PA) is a direct plasminogen activator which is produced endogenously by endothelial cells. Commercial production is available by means of recombinant DNA technology, which makes it approximately eight times more expensive than SK. This agent produces less systemic fibrinolysis than SK since it converts plasminogen to plasmin more efficiently in the presence of clot-bound fibrin.

While there is universal agreement about the clinical usefulness of thrombolytic agents, controversy remains about the choice of SK or t-PA, especially given the differences in the costs of these agents. Three large randomized clinical trials have directly compared SK to t-PA in AMI patients. The GISSI-2 trial (4) compared t-PA to SK both with and without subcutaneous heparin beginning 12 hours after the start of therapy. The ISIS-3 trial (5) compared t-PA and SK both with and without subcutaneous heparin in a similar factorial design, but began heparin 4 hours after the start of therapy. The 35-day mortality and stroke rates for these trials are shown in Table 1, where the different heparin groups have been combined since the survival rates in these groups were indistinguishably close. These two trials showed little difference in mortality rates between t-PA and SK, although the stroke rate was consistently elevated in the t-PA arm.

The next comparative trial was GUSTO-1 (6), which randomized 41,021 patients to four different thrombolytic strategies following an acute myocardial infarction. The treatment groups included two SK arms, one with intravenous and the other with subcutaneous administration of heparin, a t-PA arm, and a group given a combination of the two agents. This multi-center trial recruited patients from the United States



**Table 1** Mortality, Stroke and Combined Endpoint for Three Mega-trials Comparing SK to t-PA

| Trial | Agent | No. patients | Death | Stroke | Stroke or death |
|---|---|---|---|---|---|
| GISSI-2 | SK | 10396 | 958 (9.2%) | 98 (0.9%) | 1014 (9.8%) |
| | t-PA | 10372 | 993 (9.6%) | 136 (1.3%) | 1067 (10.3%) |
| ISIS-3 | SK | 13780 | 1455 (10.6%) | 141 (1.0%) | 1530 (11.1%) |
| | t-PA | 13746 | 1418 (10.3%) | 188 (1.4%) | 1513 (11.0%) |
| GUSTO-1 | SK | 21251 | 1475 (7.4%) | 261 (1.5%) | 1636 (8.1%) |
| | t-PA | 10396 | 653 (6.3%) | 161 (1.6%) | 746 (7.2%) |
| | t-PA + SK | 10374 | 723 (7.0%) | 170 (1.6%) | 817 (7.9%) |

(17,796), Canada (2,898) and 13 other countries. Compared with the combined SK branches, the strategy of "front-loaded" or "accelerated" t-PA used throughout the trial showed a statistically significant lowered mortality (6.3% vs 7.3%, respectively; $p = 0.001$) and combined end point of 30-day mortality or disabling stroke (6.9% vs 7.8%, respectively; $p = 0.006$). See Table 1, where again the two SK groups have been combined, as in the original analysis of the GUSTO investigators. While GISSI-2 and ISIS-3 report 35-day mortality rates and GUSTO-1 only 30-day rates, the probability of a death occurring between 31 and 35 days is very low, in line with a 2% probability of death uniformly spread between 31 days and one year.

The GUSTO-1 trial was well executed, and the sample size was designed to have at least 80% power to detect a 15% reduction in mortality, equivalent to an absolute mortality difference of 1% between experimental groups. This value has been (somewhat arbitrarily) defined by the GUSTO-1 investigators as the minimal clinically important difference between the two agents. Here we will continue to accept a 1% decrease as a clinically meaningful difference, although pharmaco-economic analyses may be required to further investigate the cost-effectiveness ratio. Most clinicians would accept the frequentist analysis ($p = 0.001$) of this study as being conclusive (or almost conclusive) proof of the superiority of t-PA, that is, that the mortality rate for t-PA is less than that for SK, but in this chapter we will discuss whether this is an adequate summary of the available evidence. GUSTO-1 was a Herculean effort that was carefully carried out, but this does not exempt it from an equally careful examination of the conclusions, particularly in light of the cost differential and the contradictory data from the other trials. While many critiques of the GUSTO-1 trial have been published (7,8,9,10), these have mostly centered on design issues and the interpretation of the clinical relevance of the observed mortality differences. This chapter will raise further questions while performing increasingly complex meta-analyses of the data from the three major clinical trials comparing the two drugs.

It is important to note that, although all the trials were randomized with uniform entry criteria and drug dosages, reservations have been expressed about the relevance of any comparisons between these studies. The major sources of controversy are as follows:



- The t-PA used in ISIS-3 was of a slightly different form, although the clinical difference is not believed to be materially important.
- Adjunctive therapy accompanying t-PA in GUSTO-1 included more aggressive use of intravenous heparin.
- In GUSTO-1, t-PA was administered in an accelerated fashion.
- The t-PA arm in the GUSTO-1 trial experienced more revascularizations than the SK arm. Since the trial was not blinded and revascularization may improve survival, GUSTO-1 may have overestimated the benefits of t-PA compared to SK due to confounding with the revascularization rate.
- Subgroup analysis of the GUSTO-1 data have revealed that there was a larger observed t-PA-to-SK difference in mortality in US patients compared to other patients. Therefore, the degree of benefit of t-PA over SK may in part be tied to an interaction with a particular health-care system.

Therefore, while there is an abundance of information comparing these two agents, there are varying opinions as to the extent to which the data from the trials could be combined. As we will show, conclusions about the advantages of t-PA can dissipate if data from the GISSI-2 and ISIS-3 trials are considered along with the GUSTO-1 data. Clinicians may vary in their weighting of the importance of the similarities and differences between the trials, leading to controversy about whether t-PA has been shown to be superior to SK.

The outline of this chapter is as follows. We begin in Section II with a very simple Bayesian analysis which considers that the three studies are sufficiently similar for a fixed-effects meta-analytic model to be appropriate, possibly with some downweighting of the GISSI-2 and ISIS-3 results compared to those from GUSTO-1. We also consider the conclusions that could be drawn from considering the GUSTO-1 results alone. Recognizing the differences in the settings of the three trials, we construct a hierarchical (random-effects) model for the difference between the death rates with t-PA and SK in Section III. Finally, the above discussion suggests that slightly different parameters may be operating in the studies, because of biases that may occur when studies use slightly different preparations, become unblinded, or use different rates of other treatments such as increased use of revascularization. Therefore, in Sec. IV, we attempt to adjust our hierarchical model to take into account some of these differences, using evidence in the literature concerning the possible

magnitudes of these biases. Since the stroke rate was higher in the t-PA groups across all three studies, we will focus on mortality, where conclusions are less clear.

## II. SIMPLE BAYESIAN META-ANALYSES

Perhaps the simplest form of meta-analysis is the pooling of data across trials as if all data arose from a single trial. From a Bayesian perspective, this is equivalent to using the data from previous trials to construct a prior distribution for the current trial, and updating via Bayes' theorem. However, as many authors have recently pointed out (11,12), no single prior distribution is likely to be sufficient to represent the diversity of clinical opinions that exists before a trial is carried out. Indeed, this diversity is usually a prerequisite for ethical randomization. Therefore, trial results should usually be reported starting from a range of prior distributions. The corresponding set of posterior distributions then summarizes the range of post-trial beliefs. If this latter set of distributions includes only a sufficiently narrow range of possible effects, conclusions could be drawn with which most clinicians should agree regardless of their initial opinions. Otherwise, the debate continues and further research is indicated.

Since our trial results are given in terms of binomial outcomes, a convenient conjugate family of distributions is the beta. A random variable $\theta$ follows a beta distribution with parameters $\alpha$ and $\beta$ if

$$f(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{(\alpha-1)}(1-\theta)^{(\beta-1)} \quad \text{for } 0 < \theta < 1,$$

where $\theta$ represents the probability of a "success" (which in our case paradoxically is a death), and where $B(\alpha, \beta)$ is the Beta function which represents the required constant coefficient in a beta density. The mean of a beta$(\alpha, \beta)$ density is given by $\alpha/\alpha+\beta$, and the variance is $\alpha\beta/(\alpha+\beta)^2(\alpha+\beta+1)$. The parameters $\alpha$ and $\beta$ may be interpreted as the prior numbers of successes and failures, respectively. For example, according to Table 1, there were a total of $\alpha = 958 + 1455 = 2413$ deaths and $\beta = 9438 + 12325 = 21763$ survivors under SK in the GISSI-2 and ISIS-3 trials, so that a prior density for the probability of death from SK in the GUSTO-1 trial might be a beta(2413, 21763) density.

This prior distribution considers that the information on each patient in the GISSI-2 and ISIS-3 trials is as important and relevant as



in the GUSTO-1 trial. Alternatively, a researcher who believes that the difference in trial protocols cannot be ignored might elect to only partially consider the earlier results. For example, one could arbitrarily treat the value of each observation in the previous trials as worth only 50% or even 10% of each observation in the GUSTO-1 data, so that the above prior distribution would change to a beta(2413/2, 21763/2) or a beta(2413/10, 21763/10) density, respectively. A more extreme position would be that the trials are too dissimilar to be combined, and that consequently all previous research should be ignored. This approach assumes that nothing is known about the mortality rate, implying that a diffuse prior distribution which spreads mass equally over the entire feasible region from 0 to 1 might be appropriate. Therefore, a uniform prior distribution, equivalent to a beta(1, 1) density, could be used as the prior density for each rate. Other prior distributions are also possible and are not necessarily derived by a weighting of previous data. Most of these would fall in between the above extremes. For example, an expert panel could be convened to elicit a prior distribution for the GUSTO-1 trial taking into account the similarities and differences between the three trials, and all other relevant information.

Similar prior densities can be constructed for the t-PA arm of the trial. We are interested in the difference in the mortality rates for patients given t-PA and SK, that is, the difference between two beta densities. Since no convenient closed-form solution exists to express this density, one can either use Monte Carlo simulation or normal approximations. With our large sample sizes, the normal approximation is virtually indistinguishable from the true posterior density, so this was used to create Figs 1 and 2, discussed below. We chose to use Monte Carlo simulations via a Gibbs sampler (13) for the analyses presented in Secs 3 and 4, which are more complex.

Figure 1 shows three prior probability densities for the difference in mortality between t-PA and SK derived from using 100%, 50%, and 10% weighting of the data from GISSI-2 and ISIS-3. The prior mean of these curves is close to zero (0.0002), suggesting that no important difference exists between the two agents. Fully accepting the results of these two trials would suggest almost no *a priori* possibility of t-PA being clinically superior to SK, since a decrease in the mortality rate with t-PA of greater than or equal to 1% is represented by the area under the curve to the left of −0.01, and this area is essentially zero in the case of using 100% of the prior data. This leads to a prior distribution representing skepticism as to the superiority of t-PA. Using weights of 50% or 10% of

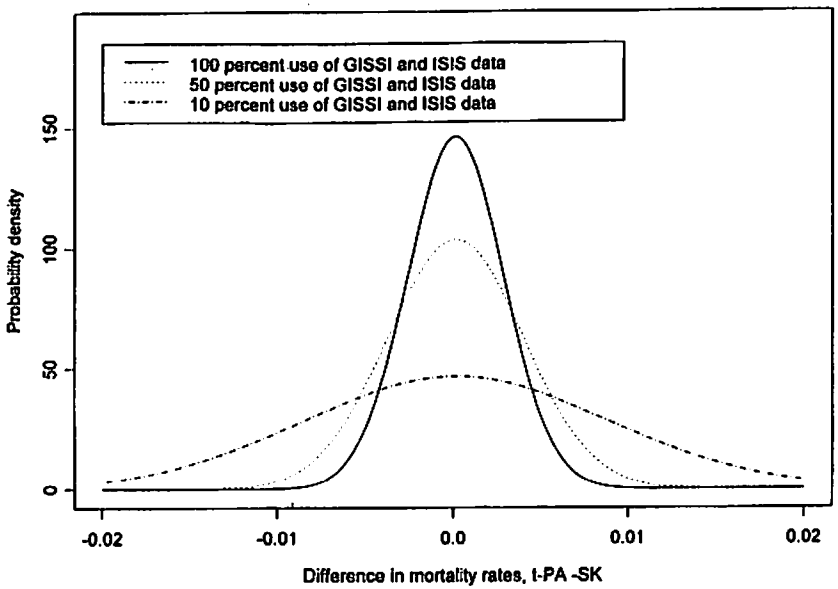


**Figure 1** Plot of the prior distributions for the difference in mortality rates between tissue-type plasminogen activator (t-PA) and streptokinase (SK) using weights of 100 and ISIS-3 data, representing a range of prior beliefs in the relevance of these trials to the GUSTO-1 trial.

course widens the prior variance of the mortality difference, increasing the prior probability of clinical superiority. Using a 10% prior weight leads to a 12% prior probability of clinical superiority of t-PA, and this number increases to approximately 49% under uniform prior densities for the mortality rates of the two agents.

Combining a beta($\alpha, \beta$) prior density with binomial data of $x$ successes in $n$ trials, via Bayes theorem, leads to a beta($\alpha + x, \beta + n - x$) posterior density, so that the posterior density for the difference in death rates in using SK and t-PA is again a difference between two beta densities. Figure 2 displays the range of posterior densities from our four different prior distributions. If one assigns equal weight to each observation from GISSI-2, ISIS-3, and GUSTO-1, then the posterior mean difference in mortality between t-PA and SK is 0.13% (0.0013 in favor of SK), and the final (posterior) probability of t-PAs being superior to SK is only about 31% (area under the curve to the left of 0). The



probability that t-PA is clinically superior to SK is almost zero, as indicated by the area under the curve to the left of −0.01. Similarly, if one considers observations from the previous randomized clinical trials to have 50% the value of each observation in GUSTO-1, then Fig. 2 shows that the probability that t-PA is superior to SK for mortality is about 53%. Further, accepting that a difference of 1% mortality is the minimum clinically significant value, the probability that t-PA is clinically superior remains negligible. If all prior data from GISSI-2 and ISIS-3 are considered irrelevant and are ignored, then t-PA is virtually certain to have a lower death rate than SK (99.96%), but the probability that t-PA exceeds the defined clinical superiority point is only 50.3%.

This preliminary analysis assumes a fixed-effects model, where the mortality difference between t-PA and SK is assumed constant between trials. Given the different settings and treatment regimens between the trials, this asumption may not be reasonable. Furthermore, neither the $p$-values reported by the GUSTO investigators nor the Bayesian analysis

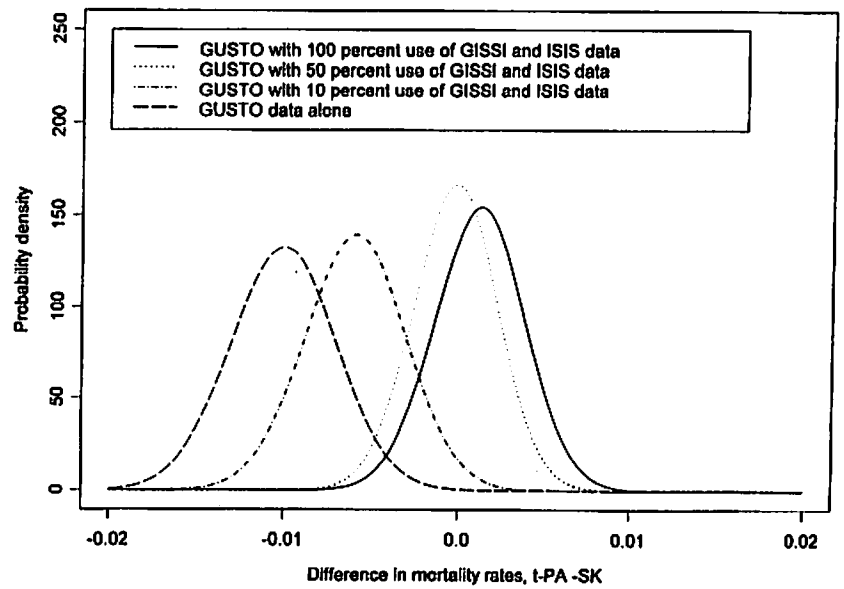


**Figure 2** Plot of the posterior distributions for the difference in mortality rates between tissue-type plasminogen activator (t-PA) and streptokinase (SK), using data from the GUSTO-1 trial and a range of prior distributions.

presented above considers potential biases or confounding. The GUSTO-1 trial like GISSI-2 was unblinded, which may lead to some degree of confounding. For example, while not reported in the original article, it appears that 9.5% of the t-PA group underwent coronary-artery bypass surgery compared with 8.5% in the SK group. This difference may have contributed to the observed mortality differences between the trials. A Bayesian approach to adjustments for such biases is presented in Sec. IV. In the next section, we calculate a hierarchical model that predicts the difference in mortality for the next randomly selected study that is similar to the three trials for which we currently have data.

## III. BAYESIAN HIERARCHICAL MODEL

Much of the controversy surrounding the interpretation of the three trials under discussion here arises from the heterogeneity of their observed mortality differences. It may therefore be reasonable to consider a Bayesian hierarchical random-effects model, wherein it is assumed that there is a distribution of possible "true effects" across different settings, and where each trial is assumed to be exchangeable (14). From this model, we can calculate a posterior density for the overall mean effect (difference in mortality rates between t-PA and SK) as well as predict what the effect might be for the "next" study.

Bayesian hierarchical models for meta-analysis of clinical trials with dichotomous outcomes are discussed by Carlin (15). Following this general approach, our model can be described in a hierarchy consisting of three levels. These levels describe the statistical relationships between data and parameters at the individual, study, and population strata.

Level I. At the individual level, the number of deaths in each of the control and treatment groups follows a binomial distribution, so that

$$s_{ti} \sim \text{binomial}(p_{ti}, n_{ti}) \quad \text{and} \quad s_{ci} \sim \text{binomial}(p_{ci}, n_{ci}),$$

where $s_{ti}$, $p_{ti}$, and $n_{ti}$ respectively denotes the number of successes, the probability of success, and the total sample size in the treatment group of the $i$th trial, for $i = 1, 2, 3$, and similarly $s_{ci}, p_{ci}$, and $n_{ci}$ for the control group.

Level II. At the study level, the logit of the control rate in study $i$, for $i = 1, 2, 3$, is defined as



$$\log\left(\frac{p_{ci}}{1-p_{ci}}\right) = \mu_{ci}.$$

The difference between the logits of the control and treatment probabilities for the $i$th trial is denoted by $\delta_i$. Therefore, we have

$$\delta_i = \log\left[\frac{p_{ti}}{1-p_{ti}}\right] - \log\left[\frac{p_{ci}}{1-p_{ci}}\right],$$

so that $\delta$ represents the log odds ratio. We assumed

$$\delta_i \sim \text{N}(\mu_\delta, \sigma_\delta^2).$$

This normal distribution represents the main hierarchical component of the model.

Level III. The third stage sets the remaining prior distributions and prior parameters. We used:

$$\mu_{ci} \sim \text{N}(0, 1000^2) \quad \text{and} \quad \sigma_{ci}^2 \sim \text{gamma}(0.001, 0.001), \quad (i = 1, 2, 3),$$
$$\mu_\delta \sim \text{N}(0, 1000^2), \qquad \sigma_\delta^2 \sim \text{gamma}(0.001, 0.001).$$

Diffuse prior distributions were appropriate since, before these three trials were performed, little was known concerning the mortality differences between the two treatment agents. In addition, with the large numbers of subjects from the trials, unless they were very strong, prior densities would have little influence.

Since no closed-form solution is available, we followed Carlin (15) in using the Gibbs sampler (13) to approximate the marginal posterior densities. See Larose (Chap. 8), Dominici and Parmigianni (Chap. 5), and DuMouchel and Normand (Chap. 6) for other examples where the Gibbs sampler has been used in meta-analysis. Using the Raftery and Lewis Gibbsit algorithm (16), we decided that 10,000 iterations of the Gibbs sampler after a burn-in of 1000 iterations was more than sufficient for highly accurate estimation of all of our parameters.

The results are displayed in Fig. 3. Looking at the results from the individual trials, the mean t-PA–SK mortality difference in the GISSI-2 study is 0.002 (in favour of SK), with a 95% credible interval of (−0.005, 0.010). Similarly, the mean mortality difference from the ISIS-3 study was also close to zero at −0.003, with 95% credible interval of (−0.010 to 0.004), and the mean from GUSTO-1 was −0.009 (in favour of t-PA) with 95% credible interval (−0.015, −0.002). Comparing the values given by the hierarchical model with those from an analysis of the results from

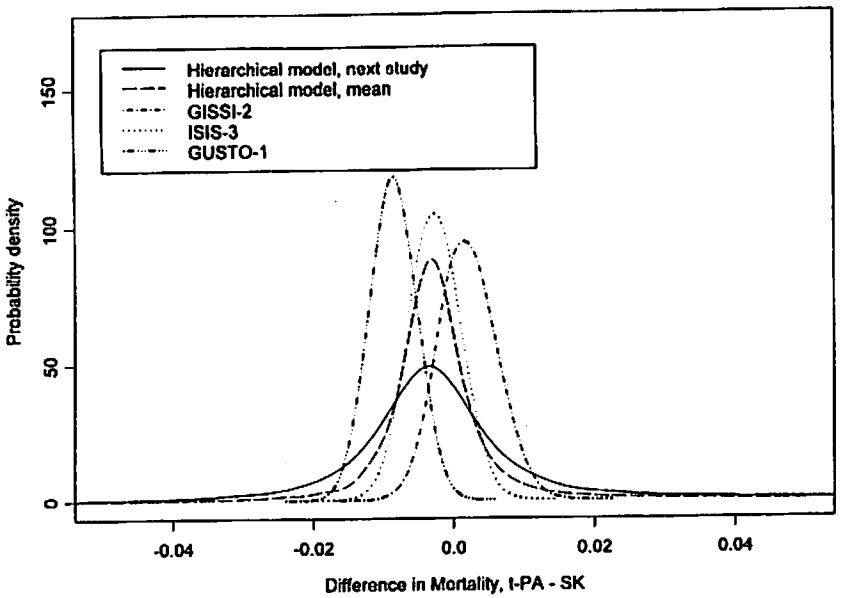


**Figure 3** Plot of the posterior distributions for the difference in mortality rates between tissue-type plasminogen activator (t-PA) and streptokinase (SK) from a hierarchical model without bias adjustments.

each trial alone (no hierarchy) showed that the model had little influence on the individual trial estimates, with all differences of the order of 0.001.

Inference about the overall mean difference on the logit scale showed that there was a 0.74 probability that $\delta < 0$, meaning that there was a 74% posterior probability that the true mean across studies was less than 0 (less mortality in the t-PA group). This is calculated from the area to the left of 0 under the posterior density curve for $\delta$ (curve not shown). However, using the predictive distribution for the "next $\delta$" gave a probability of only 66%, reflecting that the trial-to-trial variability still gives a 34% chance that the next trial would be in favor of SK. This analysis assumes exchangeability among the studies and among patients in the same study, but does not assume exchangeability of patients in different studies.

It is of course more satisfactory to draw conclusions about mean differences in mortality directly on the probability scale, rather than on the difference-in-logits scale. However, to do so requires setting a baseline



rate for the logit of the probability of a death in the control group, sinc the hierarchical model for $\delta$ provides differences from this baseline rate We compared the inferences obtained from using the mean of the poster ior distributions from each of the three studies, and fortunately all wer quite similar, with differences in means and 95% interval limits of at mos 0.01 across all comparisons. Therefore, below we report the transforma tions from the logit scale back to the difference-in-probabilities scal using the mean of the posterior density from the SK group in th GUSTO study. This value on the logit scale was −2.55, which corre sponds to a probability of death in the SK group of 0.072.

The posterior density for the mean difference and the predictiv density for the mortality difference for the "next study" are given i Fig. 3. Inference about the overall mean across all studies gives a valu very close to zero at −0.002, with 95% credible interval of (−0.017 0.016). Hence, overall, while the mean from this analysis is quite clos to zero, a 1% advantage in either direction is not ruled out. Of course, th mean predicted difference in rates for the "next study" is also −0.003, bu with a very wide credible interval of (−0.030, 0.037). This wide interva results from the relatively wide variability in mortality difference observed among these trials, as illustrated in Fig. 3, and from the fac that only three trials are considered.

While the Bayesian hierarchical model presented here accounts fo the trial-to-trial variability, it does not attempt to explain these differ ences. In the next section, we will examine whether protocol difference and possible biases in these trials can be explicitly modelled.

## IV. BIAS MODELING

Some controversy has arisen because treatment assignment in GUSTO-I was not blinded. The investigators have countered this criticism by point ing out that not all other thrombolytic trials have been blinded, tha mortality trials do not need blinding since death is a "hard outcome" and multivariate logistic regression may account for any unbalancing introduced by the unblinded nature of the trial (17,18). However, blind ing remains a key necessity even in a randomized controlled trial with hard outcomes, as unblinded randomized trials have been associated with exaggerated treatment effects (19).

The most important reason for blinding is to avoid intentional and unintentional bias in assessing outcomes across treatment arms. In this case, however, a death is a death, so does lack of blinding really matter? Despite better outcomes with accelerated t-PA (increased early reperfusion rates and decreased mortality), this strategy was associated with 1% more early revascularizations, an unbalancing well beyond what is expected by the play of chance. While reductions in mortality by appropriately employed revascularizations in stable angina may take a long time to be realized (20), this is not necessarily the case in the setting of acute ischemic syndromes. North American physicians are increasingly performing routine early post infarctus angiography and revascularization (21) with the widespread—albeit possibly unsupported—belief (22,23) that such interventions may reduce morbidity and mortality. The GUSTO-1 investigators (17) included a revascularization term in their logistic model and maintain this did not negate the "significant" mortality advantage for t-PA. Nevertheless, with a sample size of over 40,000, a mortality difference that had shrunk from 1% to perhaps 0.2% or 0.3% could remain statistically significant. Furthermore, mathematical modelling such as logistic regression can only adjust for confounding variables that are measured and included in the model. The difference in revascularization rates may not represent simple confounding (risk-factor control) but also suggests other unmeasured or unmeasurable bias in the selection of future treatments, which is of special concern in an unblinded trial.

In this light, the recent publication (24) of variations in patient management and outcomes between patients in the US and other countries participating in GUSTO is interesting. After controlling for baseline characteristics, overall prognosis was statistically improved in patients randomized in the United States. Randomization in the US may be a marker for increased revascularizations, since these procedures were three times more common in US patients. Further, there was a statistically insignificant trend for an interaction term between treatment arm/country ($p = 0.07$) and this reached statistical significance for the case of t-PA versus combination therapy ($p = 0.02$). Therefore, differences in patient management may have affected survival, although perhaps to a limited extent. While such post hoc subgroup analyses must be viewed very cautiously, the large sample size, the prospective planning, and consistency with previous European thrombolytic trials do suggest that part of the 30-day mortality advantage attributed to accelerated t-PA may be the



result of an interaction with the significantly different healthcare system in the US.

While any bias corrections are necessarily speculative, we reasoned as follows. Of the several potential biases that may make GUSTO different from the other studies, the two most important are the differences in revascularization rate between the SK and t-PA groups, and the combination of accelerated t-PA and more aggresive heparinization given in the GUSTO-1 trial but not in the others. We therefore decided to first create a prior distribution for the increase in the probability of death that might have been due to decreased revascularization in the SK arm of the GUSTO-1 trial. Using this prior distribution to impute the "true" probability of death that might have occurred had the revascularization rate been balanced, we re-created the hierarchical analyses of Sec. III. This allows the examination of the impact this adjustment could have on the between-trial variability, and on predictions for the "next" trial. Second, we created prior densities for the change in the probabilities of death in the t-PA arms of the ISIS-3 and GISSI-2 trials, to impute the "true" rates had the possibly more effective accelerated t-PA protocol been used in these trials. Combining both of these biases, we again re-estimated the hierarchical model of Sec. III. The results of these analyses are presented below.

## A. Adjusting for Possible Biases Arising from Differences in Revascularization Rates in GUSTO-1

The t-PA group benefited from a slightly higher rate of revascularization by approximately 1%, although the exact figures do not seem to be published. Overall, the 41021 GUSTO-1 patients received a total of 31% (22% angioplasty and 9% by-pass surgery) revascularization procedures (24). Since undoubtedly some patients received both angioplasty and by-pass surgery, we will assume that 25% received one or both procedures. Suppose that the rates are such that 24% of the SK cohort and 25% of the t-PA cohort received a revascularization procedure. It seems reasonable to assume that the more severe cases were sent for early revascularizations. The most severe patients across all treatment groups included 2200 who experienced cardiogenic shock. Of these, 406 received a revascularization, with mortality rate of 38%, and 1794 did not receive the procedure, with a mortality rate of 62% (25). If 1% of the SK cohort (200 patients) were at high risk and did not receive revascularization, it is

then conceivable that this may have lead to approximately 50 additional deaths ($200 \times (0.62 - 0.38)$), with a range of perhaps 0 to 100 deaths. This may have introduced a bias of 2.5 deaths per 1000 patients treated, with a range of 0 to 5 deaths per 1000 patients treated.

Using this range, we re-estimated the posterior density of the mortality rate difference in the GUSTO-1 study, and updated the hierarchical model discussed in Sec. 3. The resulting range of posterior predictive densities for the "next" study are displayed in Fig. 4. While the probability of mortality in the SK group decreased from 7.2% (95% CI = 6.9%, 7.6%) to 7.0% (95% CI = 6.5%, 7.5%) to 6.7% (95% CI = 6.4%, 7.1%) for the models without bias adjustments and with adjustments of 2.5 and 5 additional deaths per 1000 treated, respectively, Fig. 4 shows that this bias adjustment had a visible impact on the predictive distribution for the "next" study. Nevertheless, the overall message

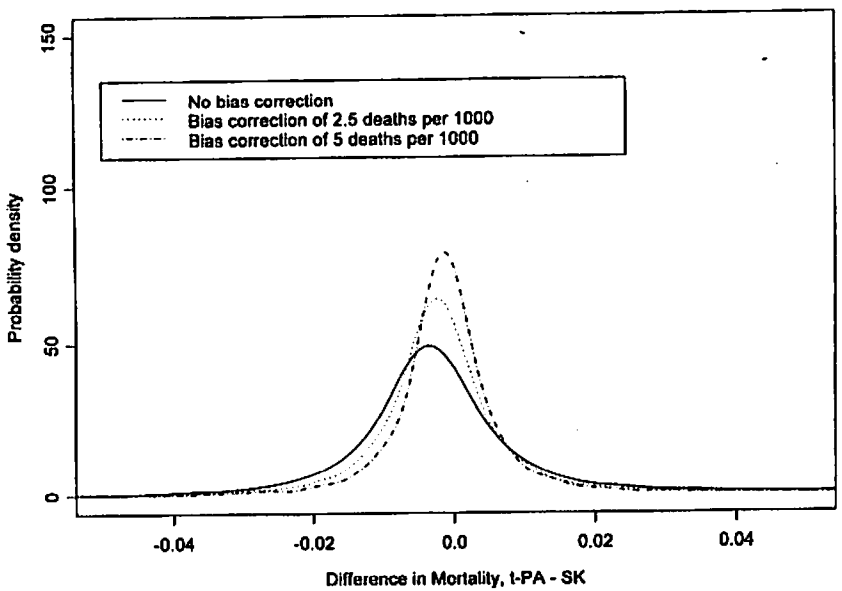


**Figure 4** Plot of the posterior distributions for the difference in mortality rates between tissue-type plasminogen activator (t-PA) and streptokinase (SK) from the hierarchical model without bias adjustments (same curve as in Fig. 3), and with bias adjustments for revascularization of 2.5 and 5 lives saved per 1000 treated in the SK arm of the GUSTO-1 study.



remains the same. There is not much evidence of a mortality-rate difference between the trials, and although some narrowing of the posterior densities occurs due to the decreased between-study variability, the predictive intervals remain wide.

## B. Adjusting for Possible Biases Arising from Differences in t-PA Administration in GISSI-2 and ISIS-3 Compared to GUSTO-1

The prevalent paradigm is that increased coronary patency as measured by the TIMI flow score is responsible for different survival rates between thrombolytic agents. In particular, the lower mortality with the accelerated t-PA strategy (drug administration over 90 minutes) employed in GUSTO-1 has been attributed to improved early 90-minute patency. There are nevertheless some observations that are difficult to reconcile with this theory. First, virtually all thrombolytic agents give similar patency rates at longer times such as 180 minutes (26). Also, the newer strategies for administering t-PA, for example as a double bolus injection, have shown marked improvement in TIMI flow but no improved clinical outcomes when compared to accelerated t-PA (27). Further, while a standard 3-hour infusion of t-PA has improved TIMI grade 3 flow compared to SK (50.2% vs 31.5%) (28), earlier comparative trials did not find any mortality reductions.

At 90 minutes, the rates of TIMI grade 3 flow for accelerated t-PA are improved by 13% (63.2% vs. 50.2%) compared with standard-dose t-PA (29). Accelerated t-PA had 54% TIMI 3 with a mortality rate of 4%, and 46% TIMI 0–2 with a mortality rate of 8% for an overall predicted mortality of approximately 5.9%. Standard t-PA might have had only 41% TIMI 3, with a mortality rate of 4%, and 59% TIMI 0–2 with a mortality rate of 8%, leading to an overall predicted mortality of 6.4%. The difference in rates (6.4% – 5.9%) suggests that accelerated t-PA could be responsible for 5 lives saves per 1000 treated. A reasonable range for the difference in the numbers of deaths might then be 0 (no advantage at all for accelerated t-PA) to 10 lives per 1000 treated (completely explaining the SK to t-PA difference seen in GUSTO).

Using this range to decrease the numbers of deaths in the t-PA arms of the GISSI-2 and ISIS-3 trials, both alone and combined with the revascularization adjustment of Section IV.A produced the range of curves in Fig. 5. As expected, assuming that the accelerated t-PA protocol

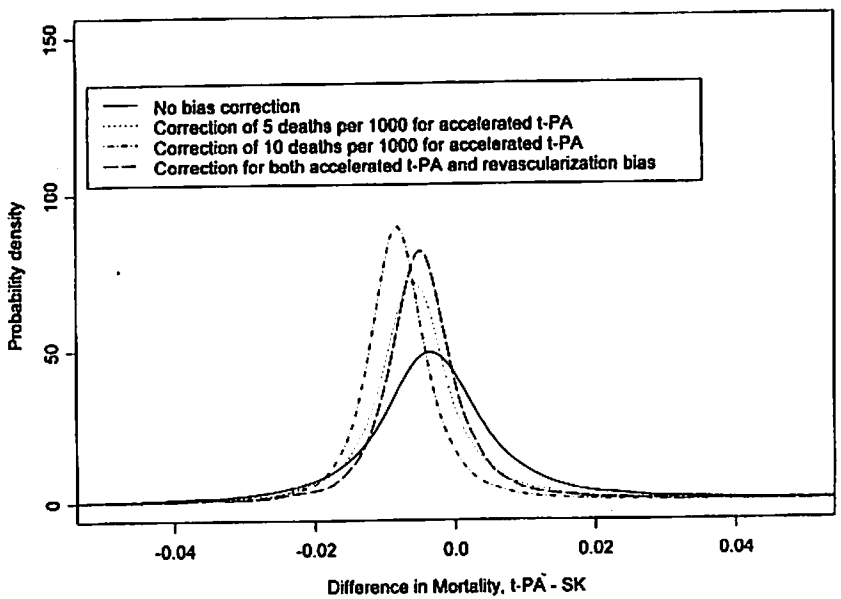


**Figure 5** Plot of the posterior distributions for the difference in mortality rates between tissue-type plasminogen activator (t-PA) and streptokinase (SK) from the hierarchical model without bias adjustments (same curve as in Fig. 3), and with bias adjustments for accelerated t-PA in the GISSI-2 and ISIS-3 study, and for a combination of both revascularization and accelerated t-PA bias (see text).

saves 5 or 10 lives per 1000 patients treated shifts the predictions for the "next" study in favor of t-PA. Correcting for both revascularization (2.5 lives saved per 1000 treated in the SK arm of GUSTO-1) and accelerated t-PA (5 lives saved per 1000 treated in the t-PA arms in GISSI-2 and ISIS-3) does not substantially change the mean prediction for the "next" study compared to a model with no adjustments, although the prediction interval narrows somewhat.

Even in the scenario most favorable to t-PA, there is only a 32.7% chance of t-PA being clinically superior (mortality difference of at least 1%), compared to a 19.6% chance without bias adjustments, and a 15.9% chance with both adjustments included. We conclude that, while the biases discussed here may explain some of the differences in the results of the three trials, adjusting for them does not substantially change the conclusions of the simpler analyses presented in Secs II–III.



## V. DISCUSSION

In this chapter, we have extended the analyses originally presented by Brophy and Joseph (30) to include random effects and bias adjustments. The range of different Bayesian analyses presented herein suggests that restraint in accepting t-PA into routine clinical practice would be appropriate, agreeing with our previous analyses. The same conclusion was reached by Diamond et al. (31), who used a Bayesian point null-hypothesis test. This is especially true when one remembers that the stroke rate was higher in the t-PA group across all three studies, as well as the cost differences.

In assessing the evidence relating to choosing a thrombolytic agent, the reporting of $p$-values from a single trial is a poor tool for formulating policy, even when there is a considerable amount of data from a well-designed randomized clinical trial. This is due to the shortcomings of standard significance tests in addressing clinically relevant questions—and to the problems in their interpretation, especially across different sample sizes. Furthermore, classical analysis of clinical trials does not easily permit the synthesis of trial results with the range of clinicians' prior beliefs, nor the adjustment for possible biases. This makes it difficult to evaluate the coherence of the conclusions and what clinical impact the conclusions should have. Following Eddy et al. (32), we have illustrated a method that can explicitly model possible biases, helping to explore the magnitude of the effects of various differences among a set of clinical trials with different protocols. Bayesian analyses along the lines presented herein may help to overcome these problems, thereby raising the level of debate following publication of a clinical trial, or the synthesis of a set of trials.

## References

1. Gruppo Italiano per lo Studio della Streptochinase Nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Lancet, number 8478:397–402, 1986.
2. RG Wilcox, G Von der Lippe, CG Olsson, et al. Trial of tissue plasminogen activator for mortality reduction in acute myocardial infarction. Anglo-Scandinavian Study of Early Thrombolysis (ASSET). Lancet (1988, number 8610):525–530.

3. Fibrinolytic Therapy Trialists' (FTT) Collaborative Group. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. Lancet 343:311–322, 1994.
4. The International Study Group. In-hospital mortality and clinical course of 20,891 patients with suspected acute myocardial infarction randomised between alteplase and streptokinase with or without heparin. Lancet 336:71–75, 1990.
5. ISIS-3 (Third International Study of Infarct Survival) Collaborative Group. ISIS-3: a randomised comparison of streptokinase vs tissue plasminogen activator vs anistreplase and of aspirin plus heparin vs aspirin alone among 4,299 cases of suspected acute myocardial infarction. Lancet 339:753–770, 1993.
6. The GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. New England Journal of Medicine 329:673–682, 1993.
7. PM Ridker, C O'Donnell, VJ Marder, CH Hennekens. Large-scale trials of thrombolytic therapy for acute myocardial infarction: GISSI-2, ISIS-3, and GUSTO. Annals of Internal Medicine 119:530-532, 1993.
8. PM Ridker, C O'Donnell, VJ Marder, CH Hennekens. A response to 'holding GUSTO up to the light.' Annals of Internal Medicine 120:882–884, 1994.
9. P Sleight. Thrombolysis after GUSTO: a European perspective. J Myocard Ischemia. 5:25–30, 1993.
10. E Rapaport. GUSTO: assessment of the preliminary results. J Myocard Ischemia 5:15–24, 1993.
11. M Hughes. Reporting Bayesian analyses of clinical trials. Statistics in Medicine 12:1651–1663, 1993.
12. DJ Spiegelhalter, LS Freedman, MKB Parmar. Bayesian approaches to randomised trials. Journal of the Royal Statistical Society, Series A 157:357–416, 1994. Reprinted in: D. Berry and D. Stangl, eds. Bayesian Biostatistics. Marcel Dekker, 1996.
13. WR Gilks, S Richardson, DJ Spiegelhalter. Markov Chain Monte Carlo in Practice. New York: Chapman and Hall, 1996.
14. A Gelman, JB Carlin, H Stern, DB Rubin. Bayesian Data Analysis. New York: Chapman and Hall, 1995.
15. JB Carlin. Meta-analysis for 2 × 2 tables: a Bayesian approach. Statistics in Medicine 11(2):141–158, 1992.
16. A Raftery, S Lewis. How many iteration in the Gibbs sampler? In: J Bernardo et al. Bayesian Statistics 4. Oxford: University Press, 1992, pp 763–773.
17. KJ Lee, RM Califf, J Simes, F Van der Werf, EJ Topol. Holding GUSTO up to the light. Annals of Internal Medicine 120:882–885, 1994.



18. CD Naylor, PW Armstrong. From thrombolytic megatrials to clinical policy-making: Facing the facts. Canadian Journal of Cardiology 11:472–476, 1995.
19. K Schulz, I Chalmers, RJ Hayes, DG Altman. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273:408–412, 1995.
20. S Yusuf, D Zucker, P Peduzzi, et al. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the coronary artery bypass graft surgery trialists collaboration [see comments] [published erratum appears in Lancet 344 (no. 8934):1446, 1994 Nov 19]. Lancet 344:563–570, 1994.
21. DB Mark, CD Naylor, MA Hlatky, et al. Use of medical resources and quality of life after acute myocardial infarction in Canada and the United states. New England Journal of Medicine 331:1130-1135, 1994.
22. SWIFT (Should We Intervene Following Thrombolysis?) Trial Study Group. SWIFT trial of delayed elective intervention v conservative treatment after thrombolysis with anistreplase in acute myocardial infarction. BMJ 302:555–560, 1991.
23. The TIMI Study Group. Comparison of invasive and conservative strategies after treatment with intravenous tissue plasminogen activator in acute myocardial infarction. Results of the thrombolysis in myocardial infarction (TIMI) phase II trial. New England Journal of Medicine 320:618–627, 1989.
24. F Van de Werf, EJ Topol, KL Lee, et al. Variations in patient management and outcomes for acute myocardial infarction in the United States and other countries. Results from the GUSTO trial. Global utilization of streptokinase and tissue plasminogen activator for occluded coronary arteries. JAMA 273:1586–1591, 1995.
25. PB Berger, DR Holmes Jr, AL Stebbins, ER Bates, RM Califf, EJ Topol. Impact of an aggressive invasive catheterization and revascularization strategy on mortality in patients with cardiogenic shock in the global utilization of streptokinase and tissue plasminogen activator for occluded coronary arteries (GUSTO-1) trial. An observational study. Circulation 96:122–127, 1997.
26. The GUSTO Angiographic Investigators. The effects of tissue plasminogen activator, streptokinase, or both on coronary-artery patency, ventricular function, and survival after acute myocardial infarction. New England Journal of Medicine 329:1615–1622, 1993.
27. Anonymous. A clinical trial comparing primary coronary angioplasty with tissue plasminogen activator for acute myocardial infarction. The global use of strategies to open occluded coronary arteries in acute coronary syndromes (GUSTO-iib) angioplasty substudy investigators. New England Journal of Medicine 336:1621–1628, 1997.

28. RJ Simes, EJ Topol, DR Holmes Jr, et al. Link between the angiographic substudy and mortality outcomes in a large randomized trial of myocardial reperfusion. Importance of early and complete infarct artery reperfusion. Circulation 91:1923–1928, 1995.
29. NA Barbagelata, CB Granger, E Oqueli, LD Suarez, M Borruel, EJ Topol, RM Califf. TIMI grade 3 flow and reocclusion after intravenous thrombolytic therapy: a pooled analysis. American Heart Journal 133(3):273–282, 1997.
30. JM Brophy, L Joseph. Placing trials in context using Bayesian analysis: GUSTO revisited by Reverend Bayes. Journal of the American Medical Association 273(11):871–875, 1995.
31. GA Diamond, TA Denton, JS Forrester, PK Shah. Is tissue plasminogen really superior to streptokinase? Circulation 88:1-452 (Abstract), 1993.
32. DM Eddy, V Hasselblad, R Shachter. Meta-analysis by the Confidence Profile Method. New York: Academic Press, 1992.

# 5

# Combining Studies with Continuous and Dichotomous Responses: A Latent-Variables Approach

**Francesca Dominici**
*Johns Hopkins University, Baltimore, Maryland*

**Giovanni Parmigiani**
*Duke University, Durham, North Carolina*

## Abstract

A challenging problem in meta-analysis is combining studies in which similar medical outcomes are captured in some studies as continuous variables and in others as binary variables. A common approach is to dichotomize the continuous responses and proceed as in the simpler binary case. This approach is practical, but it has limitations. One is that there may be arbitrariness in the choice of the cutoff point. Another is that there is a loss of information. In this paper, we propose a strategy that overcomes both of these difficulties. It is based on assuming that the binary responses are the result of dichotomizing some underlying unobserved continuous variable. Bayesian reconstruction of the unobserved continuous variable preserves the full information from the studies reporting continuous variables and does not require the choice of arbitrary cutoff points.

We develop our method in the context of a simulation example. We then apply it to a meta-analysis of efficacy of calcium-blockers for preventing migraine headaches. In the headache example, we illustrate our approach in the context of hierarchical Bayesian