

# Clinical trials in acute myocardial infarction: Should we adjust for baseline characteristics?

Ewout W. Steyerberg, PhD,<sup>a</sup> Patrick M.M. Bossuyt, PhD,<sup>b</sup> and Kerry L. Lee, PhD<sup>c</sup> Rotterdam and Amsterdam, The Netherlands, and Durham, NC

**Background** Clinical trials concerning acute myocardial infarction often evaluate short-term death. Several baseline characteristics are predictors of death, most notably age. Adjustment for one or more predictors in a multivariable analysis may be considered to correct the estimate of the treatment effect for any imbalance that by chance may have occurred between the randomized groups. Moreover, adjustment results in a stratified estimate of the effect of treatment.

**Methods and Results** The effects of adjustment (correction for imbalance and stratification) were studied with logistic regression analysis in the Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO)-I trial. The primary end point was 30-day death, which occurred in 6.3% of 10,348 patients randomly assigned to tissue plasminogen activator and 7.3% of 20,162 patients randomly assigned to streptokinase thrombolytic therapy. This is equivalent to an unadjusted odds ratio of 0.853. No significant imbalance had occurred for any of 17 baseline characteristics considered, including well-known demographic, presenting, and history characteristics. Adjusted for age, the odds ratio was 0.829, which is an 18% increase in estimated effect on the logistic scale. When adjusted for 17 characteristics, the odds ratio was 0.820, an increase of 25%. The increase in effect estimate was largely explained by the stratification effect and only partly by imbalance of predictors.

**Conclusions** Adjustment for predictive baseline characteristics, even when largely balanced, may lead to clearly different estimates of the treatment effect on mortality rates. Adjustment for important predictors such as age is recommended in clinical trials studying patients with acute myocardial infarction. (*Am Heart J* 2000;139:745-51.)

---

## See related Editorial on page 761.

---

Randomized clinical trials provide the most reliable evidence of effectiveness of treatments in acute myocardial infarction (MI). Trials have established the beneficial effects and relative safety of several thrombolytic agents<sup>1</sup> (streptokinase,<sup>2,3</sup> tissue plasminogen activator [TPA]<sup>4</sup>) and adjunctive medical therapy<sup>5</sup> ( $\beta$ -adrenergic antagonists,<sup>6</sup> angiotensin-converting enzyme inhibitors).<sup>7-9</sup> Through similar studies, other adjunctive therapies have been shown to be of no more than limited value (addition of heparin to aspirin, nitrates, calcium-channel blockers, some antiarrhythmic drugs, magnesium).<sup>5,8</sup>

Short-term death (within 28 to 42 days from randomization) is the main end point in most larger trials.<sup>2,4,8-12</sup> Sometimes a composite measure is used, for example, consisting of the combination of death and severe left ventricular dysfunction,<sup>7,13</sup> or death and reinfarction.<sup>14</sup>

The main value of randomization is that treatment groups are on average comparable in terms of known and unknown patient characteristics. Selection on indication is excluded, a feature that cannot be guaranteed in nonrandomized comparative studies. However, baseline differences in prognosis between treatment groups may well occur, even in relatively large trials. Such differences between treatment groups are from pure chance when treatment allocation was truly random. In 5% of the cases, baseline characteristics will show "significant" imbalance between randomized groups ( $P < .05$ ).  $P$  values of statistical tests of imbalance cannot be interpreted as indicating whether randomization worked<sup>15-17</sup> and may only serve a descriptive purpose. Furthermore, absence of significant differences does not guarantee comparability of prognosis because the impact of imbalance also depends on the prognostic strength of the characteristics considered; a small difference in a very powerful predictor can make treatment groups clearly different in prognosis. Hence the similarity between randomized groups should be judged by an appraisal of the prognostic strength of the baseline characteristics and the magnitude of any imbalance.<sup>18</sup>

---

From <sup>a</sup>the Center for Clinical Decision Sciences, Department of Public Health, Erasmus University, Rotterdam; <sup>b</sup>the Department of Clinical Epidemiology and Biostatistics, Academic Medical Center, University of Amsterdam; and <sup>c</sup>the Department of Community and Family Medicine, Duke University Medical Center, Durham.

Supported in part by a grant from the Netherlands Organization for Scientific Research (NWO, S96-156).

Guest Editor for this manuscript was David L. DeMets, PhD, University of Wisconsin Medical School, Madison.

Submitted December 28, 1998; accepted September 20, 1999.

Reprint requests: Ewout W. Steyerberg, PhD, Center for Clinical Decision Sciences, Ee2091, Department of Public Health, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands.

E-mail: [steyerberg@mgz.fgg.eur.nl](mailto:steyerberg@mgz.fgg.eur.nl)

Copyright © 2000 by Mosby, Inc.

0002-8703/2000/\$12.00 + 0 4/1/104240

doi:10.1067/mhj.2000.104240

**Table 1.** Hypothetical example of stratification in a randomized clinical trial in which treatment A is compared with B and sex is exactly balanced

Treatment	Total: OR = 0.70			Men: OR = 0.50		Women: OR = 0.50	
	Dead	Survived	Total	Dead	Survived	Dead	Survived
A	82	98	180	10	80	72	18
B	98 (50%)	82	180	18 (16%)	72	(84%)	10
Total	180 (50%)			28 (16%)		152 (84%)	

An estimate of the effectiveness of treatment can statistically be adjusted for prognostic baseline characteristics with multivariable regression techniques. When predictors are fully balanced between treatment groups, one might expect that unadjusted and adjusted analyses would give identical results. Somewhat counterintuitively, this is not the case when an odds ratio is calculated for a dichotomous outcome such as death.<sup>19,20</sup> As an illustration, we consider a hypothetical randomized clinical trial with 180 patients, 90 allocated to treatment A and 90 to treatment B (Table 1). The sex distribution is exactly balanced (50% men in treatment group A and 50% men in B), but sex is a strong predictor (mortality rate in men 16%, in women 84%). The inclusion of men and women in the trial implies that the patient population is heterogeneous. The mortality rate odds ratio (OR) is 0.50 in the strata formed by sex, meaning that treatment exactly halves the odds of death both in men and women. Correspondingly, the OR is estimated as 0.50 when the trial is analyzed with adjustment for sex ( $P = .02$ ). In contrast, the OR is estimated as 0.70 ( $P = .09$ ) in an unadjusted analysis that considers the total table and ignores the heterogeneity according to sex. This example illustrates that an unadjusted OR may substantially differ from the adjusted OR, leading to an inappropriate judgment on the effectiveness of a therapy.

Multivariable adjustment is infrequent in the analysis of treatment effectiveness in randomized clinical trials. Trials in acute MI, published since 1985 and including  $>10,000$  patients, were mostly reported with the unadjusted treatment effect as the main result.<sup>2-4,6-10,12-14</sup> The only exception is the Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO)-III trial, in which 30-day death was analyzed with adjustment for 4 prespecified baseline characteristics in a logistic regression model.<sup>11</sup>

In this report we investigate the quantitative effects of adjustment for baseline characteristics in the comparison between TPA and streptokinase thrombolytic treatment in the GUSTO-I trial.<sup>4</sup> We further assess the effects of adjustment for age in the GUSTO-IIb<sup>14</sup> and GUSTO-III trials.<sup>11</sup> On the basis of the results in the GUSTO studies and previous statistical research, we will argue that adjustment for a limited number of prespecified baseline characteristics is the preferred mode

of analysis for large-scale randomized clinical trials in acute MI.

## Methods

The GUSTO-I trial enrolled patients with acute MI from 1082 hospitals in 14 countries.<sup>4</sup> Death within 30 days after random assignment was the primary end point. We compared 30-day death between 10,348 patients randomly assigned to TPA and 20,162 patients randomly assigned to streptokinase in combination with subcutaneous or intravenous heparin. The average mortality rate was 7.0% (2128 of 30,510 patients). Seventeen baseline characteristics were considered for their relation with 30-day death on the basis of previous analyses.<sup>21-26</sup>

The GUSTO-IIb trial enrolled 12,142 patients with acute coronary syndromes.<sup>14</sup> ST-segment elevation was present on the baseline electrocardiogram in 4131 patients and absent in 8011. The effectiveness of intravenous heparin or hirudin was assessed with regard to the composite end point of death or nonfatal MI or reinfarction at 30 days.

The GUSTO-III trial enrolled 15,059 patients who were seen within 6 hours after the onset of symptoms with ST-segment elevation or bundle-branch block.<sup>11</sup> Thirty-day death was compared between 10,138 patients randomly assigned to double-bolus reteplase and 4921 randomly assigned to accelerated alteplase.

Logistic regression was used to compare the treatment effects on the primary end point in each trial. Detailed analyses were performed for the GUSTO-I trial. The relative prognostic strength of 17 baseline characteristics was evaluated by their univariable  $\chi^2$  model, which was calculated as the difference in  $-2$  log-likelihood between a univariable logistic regression model with and without the characteristic. We further calculated an  $R^2$  measure on the log-likelihood scale, which approximately indicated the percentage of variance explained.<sup>27</sup> For illustrative purposes, differences between randomized groups were tested with the use of the nonparametric Mann-Whitney tests for continuous variables and Pearson  $\chi^2$  tests for categorical variables.

We considered 3 logistic regression models in GUSTO-I. The first model included treatment as the single covariable: log odds (death) =  $\alpha + \beta \cdot \text{treatment}$ , where  $\alpha$  indicates the intercept and  $\beta$  the regression coefficient for treatment, which was coded as 0 for streptokinase and 1 for TPA. A second model adjusted the treatment effect for age (in years): log odds (death) =  $\alpha + \beta \cdot \text{treatment} + \beta_1 \cdot \text{age}$ . A third model adjusted for all 17 predictors considered in this study (including age): log odds (death) =  $\alpha + \beta \cdot \text{treatment} + \beta_1 \cdot \text{age} + \beta_2 \cdot \text{predictor 2} + \dots + \beta_{17} \cdot \text{predictor 17}$ . The first model represented an unadjusted analysis; the second and third models represented analyses adjusted for baseline characteristics.

**Table II.** Baseline characteristics of 30,510 patients with acute MI randomly assigned to TPA or streptokinase in the GUSTO-I Trial

Baseline characteristics	TPA (n = 10,348)	Streptokinase (n = 20,162)	Imbalance P value	$\chi^2$	R <sup>2</sup>
<b>Demographics</b>					
Age (y)	61.03 (52;70)	60.86 (52;70)	.29	1492*	12.0%
Female sex	25.3	25.3	.86	248	2.0%
Weight (kg)	79.6 (70;89)	79.4 (69;88)	.20	315*	2.6%
Height (cm)	171.1 (165;178)	171.0 (165;178)	.51	355*	2.9%
<b>Risk factors</b>					
Hypertension	38.2	38.1	.81	82	0.7%
Diabetes	14.5	15.1	.20	115	0.9%
Smoking			.92	370*	3.0%
Current	42.9	42.9			
Ex-smoker	27.3	27.5			
Never	29.8	29.6			
Hypercholesterolemia	34.6	34.3	.59	54	0.4%
<b>Other history</b>					
Family history of MI	42.0	42.8	.18	104	0.9%
Previous MI	16.9	16.5	.46	189	1.6%
Previous angina	37.7	36.9	.15	61	0.5%
Previous revascularization	8.1	7.6	.12	7	0.1%
<b>Presenting characteristics</b>					
Hypotension (blood pressure < 100 mm Hg)	8.0	8.3	.33	388	3.2%
Tachycardia (pulse >80 beats/min)	32.5	32.7	.76	219	1.8%
Anterior infarct location	38.9	38.9	.95	253	2.1%
Killip class			.44	1014*	8.2%
I	85.0	85.4			
II	12.8	12.5			
III	1.4	1.4			
IV	0.8	0.7			
ST elevation on electrocardiography (>4 leads)	37.3	37.8	.41	202	1.7%

Values in the TPA and streptokinase columns: median (25th, 75th percentiles) or the percentage of patients with the characteristic. Imbalance P values, Mann-Whitney test (continuous variables) or Pearson  $\chi^2$  test (discrete variables);  $\chi^2$  is univariable  $\chi^2$  indicating prognostic strength; R<sup>2</sup>, Nagelkerke R<sup>2</sup> index, indicating percent variance explained.

\*Treated as linear terms in logistic regression analysis.

The difference between the unadjusted and the adjusted regression coefficient for the treatment variable was attributed to imbalance and stratification. We note that synonyms may be used for adjustment (eg, "controlling"), for imbalance (eg, "confounding"), and for stratification (eg, "conditioning"). Although "stratification" usually refers to conditioning on categorical subgroups (eg, on sex, Table I), we also use this term when continuous variables are involved, for example, age. We approximated the difference in treatment effect that was attributable to imbalance by multiplying the difference in mean value of the baseline characteristic(s) with the regression coefficient(s) in the adjusted analysis. For example, the TPA group was 0.17 years older than the streptokinase group, and the regression coefficient for age in the adjusted analysis was 0.082. Imbalance hence was considered responsible for a difference of  $0.17 \cdot 0.082 = 0.014$  in the logistic regression coefficient that represented the adjusted treatment effect. The remaining part of the difference between unadjusted and adjusted treatment effect was attributed to stratified estimation.

The statistical power of the adjusted analyses was compared with that of the unadjusted analysis. We hereto calculated the Wald statistic (coefficient divided by its standard error) for the unadjusted and adjusted treatment effect for a fully balanced distribution of baseline characteristics. The coefficient of the adjusted treatment effect was corrected for the observed imbalance

in the GUSTO-I trial in an attempt to include only the stratification effect. Because the standard error is inversely related to n, the sample sizes could be calculated that would give identical Wald statistics and hence identical power for the adjusted and unadjusted analyses.

In the GUSTO-IIb and GUSTO-III trials, we evaluated two logistic regression models: one unadjusted and one age-adjusted model. Age was included as a simple linear term in the latter model to illustrate the effect of adjustment for a single, relatively strong baseline characteristic.

## Results

The distribution of 17 baseline characteristics is shown in Table II for the patients randomly assigned to TPA or streptokinase in the GUSTO-I trial. No statistically significant imbalance had occurred for any of these characteristics. The strongest prognostic factor was age (R<sup>2</sup> = 12%), followed by Killip class and hypotension.

In the TPA group, 30-day mortality rate was 6.3% (653 of 10,348 patients), in contrast to 7.3% (1475 of 20,162 patients) in the patients randomly assigned to streptokinase. This corresponded to an OR of 0.853 as the estimate for the unadjusted treatment effect or a

**Table III.** Results of unadjusted and adjusted analyses of the effect of TPA versus streptokinase on 30-day mortality rates in the GUSTO-I trial (n = 30,510)

	OR	Coefficient	SE	$\Delta$ Coef	$\Delta$ SE	Imbalance	Stratification
Unadjusted	0.853	-0.1586	0.09	—	—	—	—
Adjusted for age	0.829	-0.1878	0.050	18%	3%	9%	9%
Adjusted for 17 predictors	0.82	-0.1982	0.053	25%	8%	4%	21%

$\Delta$ Coef and  $\Delta$ SE, Difference between adjusted and unadjusted logistic regression coefficient and standard error.

logistic regression coefficient of  $-0.159$  (Table III). The standard error of this coefficient was  $0.049$ . For comparison with Table II, we note that the chi-square value of the treatment effect was  $10.8$  ( $P = .001$ ) and the  $R^2$  value was  $0.1\%$ . The prognostic strength of treatment hence was small compared with patient characteristics such as age (chi-square =  $1492$ ).

The OR of the treatment effect was  $0.829$  when adjusted for age and  $0.820$  when adjusted for 17 baseline characteristics (Table II). The adjusted logistic regression coefficients were  $-0.188$  and  $-0.198$ , respectively, with corresponding standard errors of  $0.050$  and  $0.053$ . The treatment effect hence increased by  $18\%$  or  $25\%$  on the logistic scale, respectively. Part of this difference could be attributed to imbalance. For example, patients in the TPA group were slightly older, and age was a very strong predictor of short-term death. The slight difference in age accounted for a difference of approximately  $9\%$  between unadjusted and age-adjusted treatment effect. The TPA group also had a slightly less favorable prognosis when all 17 characteristics were considered. This imbalance accounted for a difference of approximately  $4\%$  in the treatment effect. Considerable differences in treatment effect ( $9\%$  and  $21\%$ ) were attributable to stratification on predictors.

Adjustment increased the standard errors by  $3\%$  and  $8\%$ , respectively. This increase was less than the increase in treatment effect caused by stratification. This observation illustrates that the power of a study increases by adjustment for predictive baseline characteristics. The power of the unadjusted analysis with  $30,510$  patients could be achieved with  $26,900$  or  $24,850$  patients when adjusting for age or 17 characteristics, which equals a reduction in sample size by  $12\%$  or  $19\%$ , respectively.

In the GUSTO-I trial, both imbalance and stratification led to more extreme estimates of the treatment effect. This was not the case in the GUSTO-IIb trial. The unadjusted treatment effect for hirudin was an OR of  $0.89$  (logistic regression coefficient  $-0.120$ ). However, this effect was explained partly by a slightly younger average age in the hirudin group compared with the heparin group ( $63.71$  vs  $63.86$  years,  $P > .20$ ). The age-adjusted coefficient was  $-0.118$ , which represents a decrease in effect of  $-1.6\%$ . This effect is explained by the combination of a  $-5.0\%$  reduction because of imbalance in

age and a  $+3.4\%$  increase because of stratification. The stratification effect was smaller than that observed in GUSTO-I, which is related to the smaller prognostic effect of age (OR  $1.5$  per 10 years compared with  $2.3$  in GUSTO-I).

In GUSTO-III, a slightly higher 30-day mortality rate was observed in the reteplase group compared with the alteplase group ( $7.47\%$  vs  $7.24\%$ ). The unadjusted OR was  $1.03$ , or a logistic regression coefficient of  $0.033$ . The age-adjusted coefficient was  $0.042$ . The difference between these effect estimates could fully be explained by the slightly younger age in the reteplase group ( $62.16$  vs  $62.30$  years,  $P > .20$ ).

## Discussion

In this era of evidence-based medicine, randomized clinical trials are essential instruments to quantify benefits and risks of treatment. The results reported in this study illustrate that adjustment for baseline characteristics has 2 effects on the estimation of the OR in a clinical trial. First, it is well known that adjustment may correct for imbalance that occurred between the randomized groups. This imbalance is from pure chance if a proper randomization procedure has been used, but correction for imbalance may facilitate the interpretation of the trial result.<sup>15,16</sup> In the GUSTO-I, GUSTO-IIb, and GUSTO-III trials, the imbalances in age were approximately  $0.15$  years. These small differences were statistically nonsignificant but explained a difference of up  $9\%$  in the adjusted estimate of the treatment effect (TPA vs streptokinase in GUSTO-I). Larger imbalances may well occur, especially in smaller trials. Second, adjustment leads to an estimate of the treatment effect that is stratified on baseline characteristics. From the biostatistical literature it is known that a stratified odds ratio is more extreme than the unadjusted odds ratio in a heterogeneous population.<sup>19,20</sup> Stratification accounted for a difference up to  $21\%$  between unadjusted and adjusted treatment effect in the GUSTO-I trial, when 17 predictors were considered. This difference is specific to nonlinear statistical analyses of dichotomous end points such as short- or long-term death, in which logistic or Cox regression may be applied<sup>28-31</sup>; it does not occur with linear regression analysis of continuous end points.<sup>19,20</sup>

Adjusted and unadjusted analyses of mortality rates result in identical estimates of the treatment effect in homogeneous populations, in which no predictor effects are known.<sup>19</sup> This theoretical expectation was confirmed by the relatively small stratification effect in GUSTO-IIb, in which the relation between age and the composite end point of death or nonfatal MI or reinfarction was less strong than between age and 30-day death in GUSTO-I or GUSTO-III. In GUSTO-III, no stratification effect could be identified because the treatment effect was virtually absent (OR near 1).

It is well known that patients with acute MI constitute a heterogeneous population with respect to 30-day risk of death. For example, mortality rate may be <2% in those <50 years of age and exceed 25% in those >80 years of age.<sup>26</sup> An unadjusted estimate of the treatment effect may be interpreted as relating to "a patient" with an acute MI. Adjustment for age results in an estimate for "a patient of a certain age." Adjusting for all known predictive baseline characteristics results in an estimate for "a patient with a certain risk profile." Adjustment thus results in more individualized estimates of the treatment effect (more "subject-specific"), in contrast to the unadjusted estimate ("population-averaged").<sup>32</sup> However, we must realize that we will never be able to truly adjust for all relevant predictors because many predictors are yet unknown or unmeasurable.<sup>33</sup> Adjustment for known predictors leads to the closest empiric estimates of individual treatment benefit.

It has been argued that the net benefit of treatment for the individual cardiologic patient depends on the effect of treatment combined with his or her absolute risk.<sup>34-37</sup> The latter can be obtained from multivariable risk equations, which must be well calibrated. The treatment effect can be derived from a randomized clinical trial, in which adjusted analyses provide more individualized estimates than unadjusted estimates. An important assumption in this approach is that the relative risk reduction is constant for all patients, that is, independent from their absolute risk. Treatment effects in specific subgroups might be assessed to check this assumption, although multiple testing makes such analyses often exploratory in nature. We note that unadjusted analyses make an even stronger assumption by implying that the trial enrolled a homogeneous patient population in which a single treatment effect is applicable to all patients.<sup>17</sup>

In practice, it may be difficult to select an appropriate set of baseline characteristics for adjustment.<sup>17,38</sup> Literature review and clinical knowledge may indicate important predictors. In the case of short-term death after acute MI, a large number of studies have analyzed risk factors.<sup>21-26</sup> Important predictors include especially age, followed by presenting characteristics, such as Killip class, systolic blood pressure, heart rate and infarct location, and, to a lesser extent, history characteristics

such as a previous infarction. Combinations of these 6 predictors provided most of the prognostic information in GUSTO-I.<sup>26</sup> In the GUSTO-III trial, 4 of these characteristics (age, systolic blood pressure, heart rate, and infarct location) were prespecified for adjustment of the effect of reteplase versus alteplase.<sup>11</sup> The trial protocol specified the adjusted analysis as the primary statistical analysis, and the selection of these characteristics did not depend on statistical testing of the predictive effect nor imbalance of the characteristics.<sup>43</sup> This set of baseline characteristics constitutes a small number of covariables in the regression analysis relative to the number of events (deaths). This limits any problems of statistical overfitting, which might especially occur in small trials with few events.<sup>39-42</sup>

We illustrated that adjusted analyses have a higher statistical power to detect effects of treatment.<sup>19,20</sup> In GUSTO-I, the power of the unadjusted analysis could have been achieved with approximately 15% fewer patients. We note that this sample size calculation was a post hoc exercise; an adjusted sample size calculation before the start of the trial might make more sense. Statistically, however, such a calculation is complicated and requires further study. Moreover, the prognostic impact of the baseline characteristics would have to be estimated in addition to making assumptions about the size of the treatment effect and the average incidence of the outcome. Therefore we propose that the required sample size of a trial is based on the unadjusted analysis, as was done for GUSTO-III.<sup>43</sup> The estimated power then will be a conservative estimate of the power of the actual analysis.

Since adjusted analyses have advantages in making individualized treatment recommendations and in statistical power,<sup>20,32</sup> we may speculate why adjusted analyses are not more common. Many investigators may not be aware of the stratification effect, as illustrated in this study (Table I). Also, multivariable methods are more complex and may be distrusted.<sup>41</sup> One might argue that adjusted analyses consider the trial data in a more derived manner than an unadjusted analysis, whereas the difference in estimated treatment effect is probably small. Especially, trials may cause concern when adjusted analyses show statistically significant results although unadjusted analyses fail to do so.<sup>44</sup>

First, assigning an adjusted analysis as the primary tool for statistical inference from a randomized clinical trial does not preclude the reporting of a cross-table with an unstratified OR as well (the GUSTO-III report).<sup>11</sup> Next, valid procedures should be followed to prevent the evaluation of multiple models until a significant treatment effect is found. One procedure is that the trial protocol prespecifies the adjusted analysis, for example, the baseline characteristics to be included in the model and preferably also their coding.<sup>43</sup> Prespecification of the exact statistical analy-

sis is also recommended in regulatory guidelines.<sup>45</sup> A risk of this procedure is that the adjustment is suboptimal, for example, because important predictors are omitted or because of a poor fit of the model to the data. The latter may be caused by incorrect modeling of nonlinear relation or by omission of interaction terms between predictors.<sup>42</sup> We note that the relation between age and 30-day death was surprisingly linear in GUSTO-I.<sup>26</sup> Another approach toward adjustment was recently proposed that circumvents the limitations of prespecification of the model.<sup>46</sup> The adjustment procedure is described in the protocol, but the exact model is not. To guarantee validity, randomization tests should be applied instead of model-based tests (as used in our analyses).<sup>38</sup> To test the null hypothesis of no treatment effect, a model should be selected that is blinded to the actual treatment assignment, for example, in the pooled treatment arms. For example, the selection could be based on the prognostic strength of the baseline characteristics and/or the magnitude of the imbalance. If the null hypothesis is rejected, the treatment effect can be estimated with a second model in which the actual treatment assignment is taken into account.<sup>46</sup>

The effects of new treatments for acute MI are expected to be small in comparison with currently available therapies, and randomized clinical trials will need to include very large numbers of patients to provide sufficient statistical power. Adjustment for well-established predictive baseline characteristics increases the power of these trials, such that beneficial effects of treatments on clinically relevant end points will more clearly be detected. Analyses of trials in acute MI should at least adjust for age, coded as a continuous, linear covariable. The large size of trials in acute MI makes substantial differences in prognosis between treatment groups unlikely. The examples in this study, however, illustrate that the effect of (statistically nonsignificant) imbalance is not negligible. We therefore conclude that trials designed to test the effectiveness of treatments in the reduction or equivalence<sup>47</sup> of mortality rates in acute MI should be analyzed with adjustment for important baseline characteristics regardless the amount of imbalance between treatment groups.

*We thank Amanda L. Stebbins, MPH, Duke Clinical Research Institute, Duke University, Durham NC, for discussions and practical support and the GUSTO investigators for the use of their data. Two anonymous reviewers and several colleagues from Erasmus University, Rotterdam, The Netherlands, provided helpful comments on previous versions of this article.*

## References

1. Fibrinolytic Therapy Trialists 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* 1994;343:311-22.
2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988;2:349-60.
3. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986;1:397-402.
4. GUSTO Investigators. An international randomized trial comparing 4 thrombolytic strategies for acute myocardial infarction. *N Engl J Med* 1993;329:673-82.
5. Hennekens CH, Albert CM, Godfried SL, et al. Adjunctive drug therapy of acute myocardial infarction: evidence from clinical trials. *N Engl J Med* 1996;335:1660-7.
6. ISIS-1 (First International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous atenolol among 16,027 cases of suspected acute myocardial infarction: ISIS-1. *Lancet* 1986;2:57-66.
7. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. *Lancet* 1994;343:1115-22.
8. ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet* 1995;345:669-85.
9. Oral captopril versus placebo among 13,634 patients with suspected acute myocardial infarction: interim report from the Chinese Cardiac Study (CCS-1). *Lancet* 1995;345:686-7.
10. 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 41,299 cases of suspected acute myocardial infarction. *Lancet* 1992;339:753-70.
11. The Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO III) Investigators. A comparison of reteplase with alteplase for acute myocardial infarction. *N Engl J Med* 1997;337:1118-23.
12. The Continuous Infusion versus Double-Bolus Administration of Alteplase (COBALT) Investigators. A comparison of continuous infusion of alteplase with double-bolus administration for acute myocardial infarction: *N Engl J Med* 1997;337:1124-30.
13. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. GISSI-2: a factorial randomised trial of alteplase versus streptokinase and heparin versus no heparin among 12,490 patients with acute myocardial infarction. *Lancet* 1990;336:65-71.
14. Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO) IIb Investigators. A comparison of recombinant hirudin with heparin for the treatment of 4 thrombolytic strategies for acute myocardial infarction. *N Engl J Med* 1996;335:775-82.
15. Altman DG, Dore CJ. Randomisation and baseline comparison in clinical trials. *Lancet* 1990;335:149-53.
16. Begg CB. Significance tests of covariate imbalance in clinical trials. *Control Clin Trials* 1990;11:223-5.

17. Senn S. Testing for baseline balance in clinical trials. *Stat Med* 1994;13:1715-26.
18. Altman DG. Comparability of randomised groups. *Statistician* 1985;34:125-36.
19. Gail MH, Wieand S, Piantadosi S. Biased estimates of treatment effect in randomised experiments with nonlinear regressions and omitted covariates. *Biometrika* 1984;71:431-44.
20. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *Int Statist Rev* 1991;58:227-40.
21. Dubois C, Pierard LA, Albert A, et al. Short-term risk stratification at admission based on simple clinical data in acute myocardial infarction. *Am J Cardiol* 1988;61:216-9.
22. Mueller HS, Cohen LS, Braunwald E, et al. Predictors of early morbidity and mortality after thrombolytic therapy of acute myocardial infarction: analyses of patient subgroups in the Thrombolysis in Myocardial Infarction (TIMI) trial, phase II. *Circulation* 1992;85:1254-64.
23. Maggioni AP, Maseri A, Fresco C, et al, on behalf of the Investigators of the Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico (GISSI-2). Age-related increase in mortality among patients with first myocardial infarctions treated with thrombolysis. *N Engl J Med* 1993;329:1442-8.
24. Maynard C, Weaver WD, Litwin PE, et al, for the MITI Project Investigators. Hospital mortality in acute myocardial infarction in the era of reperfusion therapy (the Myocardial Infarction Triage and Intervention project). *Am J Cardiol* 1993;72:877-82.
25. White HD, Barbash GI, Califf RM, et al. Age and outcome with contemporary thrombolytic therapy: results from the GUSTO-I trial. *Circulation* 1996;94:1826-33.
26. Lee KL, Woodlief LH, Topol EJ, et al, for the GUSTO-I Investigators. Predictors of 30-day mortality in the era of reperfusion for acute myocardial infarction: results from an international trial of 41 021 patients. *Circulation* 1995;91:1659-68.
27. Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika* 1991;78:691-2.
28. Cox DR. Regression models and life tables (with discussion). *J Roy Stat Soc B* 1972;34:187-220.
29. Chastang C, Byar D, Piantadosi S. A quantitative study of the bias in estimating the treatment effect caused by omitting a balanced covariate in survival models. *Stat Med* 1988;7:1243-55.
30. Ford I, Norrie J, Ahmadi S. Model inconsistency, illustrated by the Cox proportional hazards model. *Stat Med* 1995;14:735-46.
31. Schmoor C, Schumacher M. Effects of covariate omission and categorization when analysing randomized trials with the Cox model. *Stat Med* 1997;16:225-37.
32. Hauck WW, Anderson S, Marcus SM. Should we adjust for covariates in nonlinear regression analyses of randomized trials? *Control Clin Trials* 1998;19:249-56.
33. Chatfield C. Model uncertainty, data mining and statistical inference. *J Roy Stat Soc A* 1995;158:419-66.
34. Glasziou PP, Irwig LM. An evidence based approach to individualizing treatment. *BMJ* 1995;311:1356-9.
35. Califf RM, Woodlief LH, Harrell FE Jr, et al, for the GUSTO-I Investigators. Selection of thrombolytic therapy for individual patients: development of a clinical model. *Am Heart J* 1997;133:630-9.
36. Boersma E, Steyerberg EW, van der Vlugt MJ, et al. Reperfusion therapy for acute myocardial infarction: which strategy for which patient? *Drugs* 1998;56:31-48.
37. Simon A, Megnien JL, Levenson J. Coronary risk estimation and treatment of hypercholesterolemia. *Circulation* 1997;96:2449-52.
38. Gail MH, Tan WY, Piantadosi S. Tests for no treatment effect in randomised clinical trials. *Biometrika* 1988;75:57-64.
39. Copas JB. Regression, prediction and shrinkage (with discussion). *J Roy Stat Soc B* 1983;45:311-54.
40. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. *Stat Med* 1990;9:1303-25.
41. Concato JC, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201-10.
42. Harrell F, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
43. The GUSTO Organization. A randomized trial of reteplase (r-PA) versus accelerated alteplase (t-PA) for the treatment of acute myocardial infarction: the Global Use of Strategies To Open Occluded Coronary Arteries (GUSTO-III) project. Protocol August 8, 1995.
44. Cornfield JA. Recent methodological contributions to clinical trials. *Am J Epidemiol* 1976;104:408-21.
45. CPMP Working Party on Efficacy of Medicinal Products. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products. *Stat Med* 1995;14:1659-82.
46. Edwards D. On model pre-specification in confirmatory randomized studies. *Stat Med* 1999;18:771-86.
47. Ware JH, Antman EM. Equivalence trials. *N Engl J Med* 1997;337:1159-61.