CENTER FOR DRUG EVALUATION AND RESEARCH

APPROVAL PACKAGE FOR:

APPLICATION NUMBER
19-962/S-013

Statistical Review(s)

dup

MAY 30 2000

Statistical Review and Evaluation

NDA: 19,962
Applicant: Astra Zeneca LP
Drug Name: Toprol-XL (metoprolol succinate)
Indication: Congestive Heart Failure (CHF)
CDER receiving date: 9/10/1999
Document Reviewed: Volumes 1, 5, 6, 8, 14, 15

This NDA submission is to support the once daily oral administration of metoprolol CR/XL (metoprolol) as adjunctive therapy for congestive heart failure (CHF) when added to optimal standard treatment consisting of other medications. In this NDA submission, Study SH-MET-0024 (the **Merit Study**) is the only pivotal study for the efficacy and safety of metoprolol. Therefore, it is the focus of this review.

1. Outline of the Merit Study (Study SH-MET-0024)

Design

The Merit Study was a multicenter, randomized, placebo-controlled, double-blinded **trial** and involved 3991 patients from several hundred sites in the United States and thirteen European countries. This study had two primary **objectives**. The first primary objective was to determine whether metoprolol CR/XL (metoprolol) reduced total mortality. The second primary objective was to determine whether metoprolol reduced the combined endpoint of all cause mortality and all cause hospitalization. The secondary objectives included assessment of the effect of metoprolol on the combined endpoint of all cause mortality and heart transplantation, death from cardiovascular causes with cause-specific mortality from heart failure and sudden death, the pooled incidence of cardiac death and non-fatal acute myocardial infarction, and number of hospitalizations due to heart failure and other cardiovascular causes. According to the study objectives, in the protocol, time to death from all causes and time to the first event of death or hospitalization from all causes were proposed as two **primary endpoints**. The secondary endpoints were defined as time to the first event corresponding to the mentioned secondary objectives. The change from baseline in NYHA, time to discontinuation, and patient quality of life were also evaluated.

In the Merit Study, after a two-week placebo run-in period, patients with CHF (males and females with age between 40 and 80 years) were randomized to receive either placebo or metoprolol. The starting dose for the patients assigned to the metoprolol treatment was 25mg or a half of that for patients with NYHA III-IV. Then, the dose was titrated up to 200mg in the next few weeks. In this study, the first patient was randomized on Feb 14, 1997, and the study was closed on October 31, 1998. The patient recruitment duration was about 14 months and the double-blind duration was about 20 months.

The proposed **primary analysis** for the primary and secondary efficacy variables was logrank test based on time to event data. Because of the two primary endpoints, the significance level α=0.05 was split by assigning 0.04 to the first primary endpoint and 0.01 to the second primary endpoint to reflect the order of the importance of the two endpoints. Based on the logrank test, the sample size needed to achieve 80% power in detecting 21% reduction in mortality risk associated with metoprolol treatment with a 36 month follow-up was 3200, 1600 each group. The actual total number of patients randomized in this study was 3991 since more than anticipated patients were recruited according to the sponsor.

In this study, formal **interim analyses** were planned at information fractions of 25%, 50%, and 75%. The first interim analysis was carried out around February 28, 1998 when about 25% of the expected total number of deaths occurred. The trial was terminated shortly after the second interim analysis conducted around September 27, 1998 in which a significant mortality difference between the treatment groups was found.

The following table gives the schedule of the planned interim analyses and the stopping boundary for efficacy. The total significance level allocated to all-cause mortality was in fact slightly higher than 0.04 accounting for the correlation between the two primary endpoints.

Table 1.1 Stopping boundary for efficacy

Interim look for analyses	% of total number of expected deaths	Z-value (logrank test)	Cumulative Alpha (one-sided)
1	25%	3.04	0.0012
2	50%	2.98	0.0024
3	75%	2.93	0.0036
Final	100%	2.05	0.0215

Results

The following table gives the summary statistics for the major demographic and baseline characteristics in the Merit Study. The two treatment groups appeared comparable with respect to these factors.

Table 1.2 Demographic and baseline factors

Endpoint	N/total, % or N, mean (placebo)	N/total, % or N, mean (meto)
Gender (male)	1554/2001 (77.7)	1539/1990 (77.3)
Race (white)	1885/2001 (94.2)	1870/1990 (94.0)
NYHA II	825/2001 (41.2)	810/1990 (40.7)
III	1100/2001 (55.0)	1111/1990 (55.8)
IV	76/2001 (3.8)	69/1990 (3.5)
Previous acute MI	1025/2001 (51.2)	1040/2001 (52.3)
Age (year)	2001, 63.7	1990, 63.9
Ejection fraction	2001, 0.28	1990, 0.28
DBP (mmHg)	2001, 78.1	1989, 78.4
SBP (mmHg)	2001, 129.5	1990, 130.0
Heart rate (bpm)	2001, 82.7	1990, 82.7

As mentioned before, the Merit Study was terminated prematurely on 10/31/1998. At the time, small nominal p-values were observed for all primary and the secondary efficacy variables indicating significant differences between the two treatment groups with respect to these variables. The differences were all in the direction in favor of metoprolol. The p-value, adjusted for the interim analysis, for the first primary endpoint, the all-cause mortality, was **0.0062**. The sponsor did not adjust the p-value for the second primary endpoint for the reason that the hypothesis was tested only once at the time of the trial termination.

The outcomes with respect to the primary endpoints and most of the secondary endpoints are summarized in Table 1.3. Two of the sponsors results, in this reviewer's opinion, were misleading. Since only six heart cases of transplantation occurred and five of them in the metoprolol group and one in placebo, a conclusion about the beneficial effect of metoprolol on death and heart transplantation seems improper. The sponsor's analysis of time to discontinuation was invalid because of the uneven censoring from the death in the two treatment groups. The results of the sponsor's analyses of change in NYHA classification and of quality of life data are not presented in the table. A significant difference in change in NYHA classification was found. More patients in the metoprolol group (28.6%) had improved NYHA classification as compared to placebo (25.8%).

From the submitted data, the patient withdrawal from the study and censoring with respect to the efficacy variables appeared comparable between the treatment groups. According to the sponsor, for all patients, death and hospitalization information were obtained up to the closing date of the trial.

Table 1.3 Efficacy outcome / all patients

Endpoint	# of event n=2001 (pla)	# of event n=1990 (meto)	Est. of relative risk, 95% CI	p-value* logrank
All-cause mortality	217 (10.8%)	145 (7.3%)	0.66 (0.53, 0.81)	0.0001**
All cause mortality/interim analysis**, 2/28/98	86/1694 (5.1%)	56/1669 (3.4%)	0.65 (0.46, 0.91)	0.0110
All-cause mort/hosp	767 (38.3%)	641 (32.2%)	0.81 (0.73, 0.90)	0.0001
Death and heart transplantation	218 (10.9%)	150 (7.5%)	0.68 (0.55, 0.84)	0.0002
All-cause mort/ chf hosp	439 (21.9%)	311 (15.6)	0.69 (0.60, 0.80)	<0.0001
Death from card. cause	203 (10.1%)	128 (6.4%)	0.62 (0.50, 0.78)	<0.0001
Death from HF	58 (2.9%)	30 (1.5%)	0.51 (0.33, 0.79)	0.0023
Sudden death	132 (6.6%)	79 (4.0%)	0.59 (0.45, 0.78)	0.0002
Cardiac death and non-fetal MI	225 (11.2%)	139 (7.0%)	0.61 (0.49, 0.75)	<0.0001
Time to discontinuation	310 (15.5%)	279 (14.0%)	0.87 (0.74, 1.02)	0.0795
All-cause mort and CHF hosp and emer. room visit	455 (22.7%)	318 (16.0%)	0.68 (0.59, 0.79)	<0.0001
All cause hosp only**	668 (33.4%)	581 (29.2%)	0.84 (0.76, 0.95)	0.0037

* nominal p-value, ** adjusted p-value for all-cause mortality is 0.0062; ** the reviewer's analysis

To explore any unfavorable outcome in pre-specified risk groups, a number of subgroup analyses were performed by the sponsor. The following table summarizes the results of the sponsor's major subgroup analyses. The outcomes with respect to the listed variables appeared numerically consistent in favor of metoprolol group across the sub-populations.

Table 1.4 Subgroup analyses for the first and second primary endpoints

variable	subgroup	N	event	Relative risk, 95% CI
All-cause mortality (metoprolol vs. placebo)				
gender	male	3093	298	0.61 (0.48, 0.77)
	female	898	64	0.92 (0.56, 1.49)
Age (year)	<=69.4	2663	209	0.62 (0.47, 0.82)
	>69.4	1328	153	0.71 (0.51, 0.97)
Ejection fr	<=0.25	1502	186	0.62 (0.46, 0.83)
	>0.25	2489	176	0.69 (0.51, 0.93)
DBP (mmHg)	<=74	1358	144	0.59 (0.42, 0.83)
	>74	2632	217	0.70 (0.54, 0.92)
NYHA	II	1635	103	0.76 (0.51, 1.12)
	III	2211	232	0.61 (0.47, 0.80)
	IV	145	27	0.67 (0.31, 1.45)
Pre acute MI	No	2065	166	0.74 (0.54, 1.00)
	Yes	1926	196	0.60 (0.45, 0.80)
All-cause mortality and hospitalization (metoprolol vs. placebo)				
gender	male	3093	1107	0.82 (0.73, 0.92)
	female	898	301	0.79 (0.63, 0.99)
Age (year)	<=69.4	2663	849	0.78 (0.68, 0.89)
	>69.4	1328	559	0.87 (0.73, 1.03)
Ejection fr	<=0.25	1502	626	0.68 (0.58, 0.80)
	>0.25	2489	782	0.92 (0.80, 1.06)
DBP (mmHg)	<=74	1358	561	0.87 (0.74, 1.03)
	>74	2632	864	0.79 (0.69, 0.90)
NYHA	II	1635	456	0.78 (0.65, 0.94)
	III	2211	886	0.81 (0.71, 0.93)
	IV	145	66	1.06 (0.66, 1.72)
Pre acute MI	No	2065	658	0.77 (0.66, 0.89)
	Yes	1926	750	0.86 (0.75, 1.00)

2. Reviewer's analyses

A series of analyses were performed by this reviewer to assess the sponsor's efficacy findings and investigate potential problems in the trial.

Using logrank test, this reviewer compared outcomes of the primary endpoints and the secondary time-to-event endpoints in the two treatment groups. The results of the analyses were the same as those of the sponsor (Table 1.3) except a slight numerical difference for the variable of death, CHF hospitalization and emergency room visit. There was a statistically significant difference between the two treatment groups with respect to the primary endpoints. The estimated relative risk for all-cause mortality was 0.66, about a 34% reduction in mortality risk in the metoprolol group as compared to placebo (the nominal p=0.0001). The estimated relative risk for the second primary endpoint, the combined endpoint of mortality and hospitalization was 0.82 with the nominal p-value 0.000062. Several issues related to the findings with respect to these two primary endpoints will be discussed in the following.

Further analyses were performed to confirm that the observed mortality difference between the treatment groups and the differences with respect to the other primary and secondary efficacy variables were due to the treatment effect of metoprolol but not any other causes ,especially potential biases. For this purpose, this reviewer first examined the major demographic and baseline characteristics for the overall population. There was no apparent imbalance with respect to these factors between the two treatment groups.

In reviewing the Merit Study, special attention was paid to the effects of metoprolol in the U.S. patient population. This is because the goal of this NDA submission is to seek approval for marketing the drug to treat the patients with CHF in the United States. Knowing that effects of heart failure drugs can vary across different geographical regions and can be sensitive to differences in clinical practice, health care policy, and other medical and social factors in these regions, one needs to look into the outcomes of the Merit Study in U.S. patients carefully. A negative efficacy outcome for the study drug in the U.S. patients, depending on its nature, may jeopardize the approvability for the drug and limit the scope of its indication.

This reviewer found a strong suggestion of a treatment-by-region (U.S. vs. Europe) interaction with respect to mortality (**p=0.0060**). The test of interaction was based on Cox regression model with the indicators for us and treatment and the product of these two indicators as terms. A non-zero coefficient for the product term indicated a treatment-by-region interaction. In contrast to the mortality outcome in European patient population (relative risk rr=0.55, p=0.0001), the trial did not seem to show a beneficial mortality effect of metoprolol in U.S. patient population. The estimated relative risk (metoprolol vs. placebo) with respect to all cause mortality for U.S. patients was 1.05 (p=0.7961) which was, in fact, slightly in the opposite direction. For other cause-specific deaths much larger relative risks were found for U.S. population as compared to European population. For the endpoints involving mortality only (all-cause or cause-specific), the estimated relative risks for U.S. population generally were in the neighborhood of 1.0 (Table 2.1).

Table 2.1 Efficacy outcome by region

Endpoint	# of event / n (placebo)	# of event / n (metoprolol)	Est. of hazard Ratio, 95% CI	p-value* logrank
U.S. patients				
All-cause mortality	49/539 (9.1%)	51/532 (9.6%)	1.05 (0.71,1.56)	0.7961
All cause mort/ interim**	20/448 (4.5%)	20/443 (4.5%)	1.01 (0.54, 1.88)	0.9698
All-cause mort/hosp	216/539 (40.1%)	184/532 (34.6%)	0.84 (0.69,1.03)	0.0901
Death and heart trans	50/539 (9.3%)	51/532 (9.6%)	1.03 (0.70, 1.52)	0.8812
All-cause mort/ chf hosp	109/539 (20.2%)	91/532 (17.1%)	0.84 (0.64, 1.11)	0.2167
Death from card. Cause	46/539 (8.5%)	44/532 (8.3%)	0.97 (0.64, 1.46)	0.8801
Death from HF	9/539 (1.7%)	11/532 (2.1%)	1.24 (0.51, 2.98)	0.6367
Sudden death	35/539 (6.5%)	28/532 (5.3%)	0.81 (0.49, 1.33)	0.4091
Card death/non fatal MI	55/539 (10.2%)	48/532 (9.0%)	0.88 (0.60, 1.23)	0.5089
Death, CHF hosp, ER visit	109/539 (20.2%)	91/532 (17.1%)	0.84 (0.64, 1.11)	0.2161
All-cause mort/males	39/384 (10.2%)	36/375 (9.6%)	0.95 (0.60, 1.50)	0.8153
All-cause mort/females	10/155 (6.5%)	15/157 (9.6%)	1.45 (0.65,3.24)	0.3571
All-cause mort/white	34/434 (7.8%)	40/422 (9.5%)	1.21 (0.76, 1.90)	0.4225

All-cause mort/black	12/98 (12.2%)	10/103 (9.7%)	0.84 (0.36, 1.96)	0.6921
Death or discontiuation	122/539 (22.6%)	112/532 (21.1%)	0.98 (0.73, 1.21)	0.6255
All-cause hosp only	192/539 (35.6%)	161/532 (30.3%)	0.83 (0.67, 1.02)	0.0819
European patients				
All-cause mortality	168/1462 (11.5%)	94/1458 (6.5%)	0.55 (0.43, 0.70)	0.0001
All cause mort/ interim**	66/1246 (5.3%)	36/1226 (2.9%)	0.54 (0.36, 0.81)	0.0025
All-cause mort/hosp	551/1462 (37.7%)	457/1458 (31.3%)	0.81 (0.71, 0.91)	0.0006
Death and heart trans	168/1462 (11.5%)	99/1458 (6.8%)	0.58 (0.45, 0.74)	0.0001
All-cause mort/ chf hosp	330/1462 (22.6%)	220/1458 (15.1%)	0.65 (0.55, 0.77)	0.0001
Death from card. cause	157/1462 (10.7%)	84/1458 (5.8%)	0.54 (0.40, 0.68)	0.0001
Death from HF	49/1462 (3.4%)	19/1458 (1.3%)	0.38 (0.22, 0.65)	0.0002
Sudden death	97/1462 (6.6%)	51/1458 (3.5%)	0.51 (0.37, 0.72)	0.0001
Card death/non fatal MI	170/1462 (11.6%)	91/1458 (6.2%)	0.53 (0.41,0.68)	0.0001
Death, CHF hosp, ER visit	330/1462 (22.6%)	220/1458 (15.1%)	0.65 (0.55, 0.77)	0.0001
All-cause mort/males	145/1170 (12.4%)	78/1164 (6.7%)	0.53 (0.40, 0.69)	0.0001
All-cause mort/females	23/292 (7.9%)	16/294 (5.4%)	0.68 (0.36, 1.29)	0.2336
All-cause mot/ white	168/1451 (11.6%)	94/1448 (6.5%)	0.55 (0.43, 0.70)	0.0001
Death or discontinuation	324/1462 (22.2%)	241/1458 (16.5%)	0.73 (0.62, 0.87)	0.0002
All-cause hosp only	476/1462 (32.5%)	420/1458 (28.8%)	0.86 (0.75,0.98)	0.0210

* nominal p-value; ** based on information up to 2/28/98

To confirm this finding, an additional test aiming to show that the different mortality outcomes in the two regions were not likely purely due to chance was performed by this reviewer. Let r_1 and r_2 be the relative risks (metoprolol vs. placebo) with respect to all cause mortality for European and U.S. populations, respectively. Let $\beta_1 = \log(r_1)$ and $\beta_2 = \log(r_2)$. Let X_1 and X_2 be the estimates of β_1 and β_2 based on, say, proportional hazards model. Denote the standard errors of X_1 and X_2 by e_1 and e_2, respectively. The following test statistic

$$T = (X_1 - X_2)/(e_1^2 + e_2^2)^{1/2},$$

was used to test the null hypothesis H_0: $\beta_1 = \beta_2$ vs. the alternative hypothesis H_1: $\beta_1 \neq \beta_2$. Under the null hypothesis, the test statistic T has, asymptotically, the standard normal distribution. The p-value based on this test is **0.00297**, indicating β_1 and β_2 or **the relative risks for U.S. and European patient populations are indeed different.** Therefore, in my opinion, it is improper to use the mortality effect of metoprolol in European patients to explain or project the effect of metoprolol in U.S. patients.

Regarding the second primary endpoint, combined mortality/hospitalization, consistent results for U.S and European patient populations were observed. The observed relative risks were in the low 80% in both populations which were close to the observed relative risks in the two regions with respect to hospitalization alone (Table 2.1). It appeared that the outcome of the combined second primary endpoint was mainly driven by the hospitalization component. As compared to mortality, hospitalization is a more subjective endpoint and therefore more vulnerable to operational biases. There is also a concern about the potential inflation of the type I error rate associated with testing the second primary endpoint because of the sequential monitoring based on the first primary endpoint.

To explore possible causes of this inter-region difference in mortality, this reviewer checked the distributions of major demographic and baseline factors in the two treatment groups within each region and compared the distributions of these factors, pulling the treatment groups together. The following table (Table 2.2) summarizes the results of the comparisons. This reviewer noted the numerical differences in the racial and gender composition between U.S. and European patient populations. Several analyses based on the racial and gender subgroups were performed and the results are given in Table 2.1. The treatment-by-region interaction with respect to mortality could be seen in these analyses too. For instance, the estimated relative risks in male patients in the two regions apparently were different (rr=0.95 for U.S. and rr=0.53 for Europe). The mortality rate in the metoprolol group was higher than that in placebo in U.S. Caucasians. (relative risk=1.21). From Table 2.2, it was not surprising to see that the listed demographic and baseline factors in the two treatment groups were apparently balanced in both U.S. and European patient populations because of the by-region stratification in the randomization and the large sample sizes involved.

Table 2.2 baseline and demographic factors

Endpoint	N/total, % or N, mean (placebo)	N/total, % or N, mean (meto)	N/total, % or N, mean (pulled)
U.S. patients			
Gender (male)	384/539 (71.2)	375/532 (70.5)	759/1071 (70.9)
Race (white)	434/539 (80.5)	422/532 (79.3)	856/1071 (79.9)
NYHA II	226/539 (41.9)	207/532 (38.9)	433/1071 (40.4)
III	288/539 (53.4)	300/532 (56.4)	588/1071 (54.9)
IV	25/539 (4.6)	25/532 (4.7)	50/1071 (4.7)
Previous MI	258/539 (47.9)	239/532 (44.9)	497/1071 (46.4)
Age (year)	539, 62.8	532, 63.1	1071, 62.9
Ejection fraction	539, 0.27	532, 0.27	1071, 0.27
DBP (mmHg)	539, 75.0	531, 75.1	1071, 75.0
SBP (mmHg)	539, 125.5	532, 126.1	1071, 125.8
Heart rate (bpm)	539, 80.8	532, 81.1	1071, 81.0
European patients			
Gender (male)	1170/1462 (80.0)	1164/1458 (79.8)	2334/2920 (79.9)
Race (white)	1451/1462 (99.3)	1448 /1458 (99.3)	2899/2920 (99.3)
NYHA II	599/1462 (41.0)	603/1458 (41.4)	1202/2920 (41.2)
III	812/1462 (55.5)	811/1458 (55.6)	1623/2920 (55.6)
IV	51/1462 (3.5)	44/1458 (3.0)	95/2920 (3.3)
Previous MI	718/1462 (49.1)	711/1458 (48.8)	1429/2920 (48.9)
Age (year)	1462, 64.0	1458, 64.1	2920, 64.1
Ejection fraction	1462, 0.28	1458, 0.28	2920, 0.28
DBP (mmHg)	1462, 79.2	1458, 79.6	2920, 79.4
SBP (mmHg)	1462, 131.0	1458, 131.4	2920, 131.2
Heart rate (bpm)	1462, 83.4	1458, 82.9	2920, 83.1

To this reviewer, it seemed unlikely that the difference in mortality outcome between the two regions was due to potential baseline incompatibility. Other potential causes such as potential differences in clinical practices across different regions or biases are of interest.

It is well-known that beta-blockers lower heart rate. This effect of metoprolol as a beta-blocker was mentioned in the sponsor's brochure to the investigators. From the submitted data, a significant difference in reduction in heart rate between the two treatment groups was found. The 95% confidence intervals for mean reduction of heart rate, 28 days or 90 days after use of the study medications, for the two treatment groups were largely separated from each other (Table2.3).

To this reviewer, the effect of metoprolol on heart rate could potentially allow a person to know treatment codes. The lower panel of Table 2.3 shows the probabilities of correct identification of the treatment assignments of patients based on various cutoff points in reduction in HR. For example, based on the reduction in HR at Day 90, using three beat reduction as a cutoff point (reduction>3 for classifying a patient on metoprolol; otherwise on placebo), the chance of correct guess of the treatment assignment of a metoprolol treated patient can be as high as 82 percent. In the calculations, this reviewer assumed that the reductions in HR were normally distributed with the means as the estimated and the population standard deviations as the observed (=11.55 for Day 28 and 12.65 for Day 90).

Table 2.3 Mean change in HR at day 28 and day 90

endpoint	N, Mean (bpm), CI (pla)		N, mean (bpm), CI (meto)		p-value for diff. In mean changes
Day 28	1865, -1.8 (-2.3,-1.3)		1860, -8.1 (-8.6,-7.5)		<0.0001
Day 90	1867, -2.8 (-3.4,-2.2)		1844, -14.9 (-15.4,-14.4)		<0.0001
Reduction in HR (bpm)	Day 28 Spec*(%)	sens ** (%)	Day 90 Spec (%)	sens (%)	
2	50.6	70.6			
3	53.9	67.1	50.8	82.4	
4	57.4	64.0	53.6	81.2	
5	61.0	60.3	57.0	78.3	
6	64.4	56.9	59.8	75.5	
7	67.5	53.8	63.2	73.2	
8	70.5	50.5	65.8	70.9	
9			68.8	67.7	
10			71.8	65.0	
11			74.6	62.1	
12			76.4	59.3	
13			79.2	56.2	

*specificity: probability of classifying a placebo patient as on placebo
*sensitivity: probability of classifying a metoprolol-treated patient as on metoprolol

Knowing this potential unblinding of treatment codes based on heart rate in the Merit Study, one may worry about the potential biases, especially any operational biases, associated with it. However, identifying the biases and making assessment of their impacts, in general, are very difficult tasks. One reason for this is that identifying the biases especially operational biases may require the knowledge, which reviewers usually do not have, of the details of all aspects of the trial including how the trial was actually conducted in individual study sites. Another reason is that individual biases are often difficult to recognize even though the accumulated bias is large because of the mild to moderate contributions of each individual to the overall bias and possible confounding with other factors.

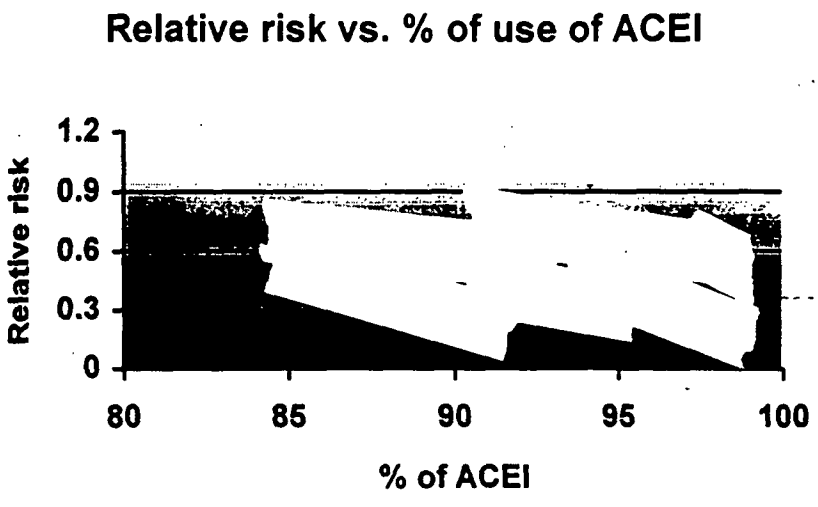
This reviewer did not find any definitive evidence of biases associated with the potential unblinding treatment codes based on reduction in heart rate. However, this reviewer did note the following things. There appeared to be a relationship between patients' hospitalization (all cause) and reduction in heart rate. There seemed to be a numerical trend that a shorter time to hospitalization was associated with a smaller heart rate reduction in placebo patients, meaning that those patients with smaller reduction in heart rate were more aggressively hospitalized. The nominal p-value for this trend (Cox regression with reduction in heart rate at Day 28 as the only covariate) in placebo patients was 0.0645. In this analysis, if a patient died, the time to hospitalization was censored at the time of patient's death. The analysis only focused on placebo patients to avoid any confounding with the potential effect of metoprolol. When the same analysis was performed for the U.S. placebo patients and the European placebo patients separately, the nominal p-values for the mentioned trend were 0.7918 for U.S. patients and 0.0281 for European patients. Based on this analysis, the risk of hospitalization decreased 1.9% for every 10 bpm increase in heart rate (Day 28) in U.S. placebo patients, but increased 9.5% in European placebo patients. These different relationships hinted a difference in medical practice between U.S. and Europe.

This reviewer also checked the background use of ACE inhibitors, a class of drugs known to have a beneficial mortality effect in CHF patients. The following table summarizes the means of the maximum dose and the mean of the cumulative dose for four major ACE inhibitors used in the trial. To calculate the means, first, the maximum dose and the cumulative dose, for an ACE inhibitor, across several visits were calculated for each patient who used the medication. Then the averages were taken. From the table, a numerically larger mean maximum dose and mean cumulative dose for the U.S. patient population as compared to the European patient population could be seen. The mean maximum dose and the mean cumulative dose appeared to be numerically slightly larger for the metoprolol group as compared to placebo.

Table 2.4 Mean maximum dos and mean cumulative dose of ACE inhibitors used

Country	treatment	N	Mean (max dose)	Mean (accu dose)
European patients				
Captopril	Placebo	338	58.5	377.4
	metoprolol	334	61.7	435.4
Enalapril	Placebo	478	15.0	102.5
	metoprolol	466	15.4	105.6
Lisinopril	Placebo	147	13.2	90.4
	metoprolol	138	13.1	90.2
Ramipril	Placebo	113	6.2	40.5
	Metoprolol	104	6.0	40.1
US patients				
Captopril	Placebo	63	101.4	653.4
	metoprolol	66	118.0	786.7
Enalapril	Placebo	111	17.2	113.4
	metoprolol	86	18.7	117.4
Lisinopril	Placebo	166	20.6	122.0
	metoprolol	150	20.7	124.3
Ramipril	Placebo	11	10.6	54.3
	Metoprolol	8	13.4	75.3

The following is a scattered plot of the observed relative risk vs. the percentage of use of ACE inhibitors in placebo group by countries. Three European countries (Iceland, Finland, and Switzland) with a few patients are excluded. The U.S. had lowest percentage of ACE inhibitor use.



The above analyses illustrated some differences in the patterns of patients' hospitalization and use of ACE inhibitors between the U.S. and European patient populations. The existence of such differences between the two regions supports, even though may not explain, the finding of the treatment-by-region interaction with respect to mortality.

3. Reviewer's comment

To this reviewer, the following issues are important in the assessment of the effectiveness of metoprolol in treating patients with CHF.

(i) Strength of evidence

According to the usual requirement of the Agency for approval for marketing a new drug, the drug sponsor needs to demonstrate the efficacy of the new drug in at least two independent well-conducted clinical trials. In case that there is only one pivotal efficacy study, like this NDA submission, the evidence of the drug efficacy needs to be much stronger to be convincing. The common wisdom often asks for an overall significance level of 0.00125 for the strength of the evidence in the "one-trial" case. The Merit Study did not meet this test on the first primary endpoint because of the observed p-value 0.006>0.00125 for the mortality difference. The question is to what extent this 0.00125 rule should be applied.

For the second primary endpoint, the combined endpoint of mortality and hospitalization, the observed nominal p-value was 0.000062. The sponsor did not make any adjustment for this p-value. In the sponsor's p-value calculation, the sample size was treated as if it were a fixed number by ignoring its stochastic nature due to the fact that the stopping time depended on the outcome of the first primary endpoint. As a result, Type I error rate

can potentially be inflated. Simulations conducted by this reviewer suggest that the size of inflation of Type I error rate depends on several factors, such as, the shape of the stopping boundary for the first primary endpoint and the correlation between the two primary endpoints. With a boundary more liberal for early trial termination (e.g. Pocock boundary) or a larger correlation between the two endpoints, the inflation of Type I error rate tends to be large. Considering the conservative stopping boundary and the moderate correlation between the two primary endpoints in the Merit Study, this reviewer thinks that the inflation of the Type I error rate, if any, should be minimal. The simulations seem to suggest that the actual p-value for this composite endpoint should be in the neighborhood of 0.0002. Therefore, to this reviewer, based on the observed small p-value, there was a statistically significant difference between the two treatment groups with respect to the combined endpoint of mortality and hospitalization (the second primary endpoint). However, the effect on this composite endpoint was mainly driven by the hospitalization component in U.S. patients.

(ii) Treatment-by-region interaction

In reviewing the Merit Study, special attention should be paid to the effects of metoprolol in the U.S. patient population, since the goal of this NDA submission after all is to market the drug for treating patients with CHF in this country. With this in mind and with the knowledge that the effects of metoprolol, as a heart failure drug, were potentially sensitive to different medical practices and other factors in different countries, one needs to have a closer look at the efficacy outcome of the trial in U.S. patients.

In the Merit Study, a strong treatment-by-region interaction was observed with respect to mortality. The relative risks in mortality for the European and the U.S. patient populations appeared to be remarkably different (p=0.0030), indicating a low likelihood that the difference was purely due to chance. The estimated relative risk in all cause mortality for European patient population was 0.55 (metoprolol vs. placebo, nominal p-value=0.0001) but was 1.05 (the nominal p-value=0.7961) for U.S. patient population. It should be pointed out that this subgroup analysis has a valid randomization because of the by-country, therefore by-region, stratification in the original randomization, which made the analysis free of potential biases that typical subgroup analyses may have due to incompatibility with respect to baseline and prognostic factors. This reviewer also analyzed the mortality data by countries. For the majority of the European countries (10 out of 13) in the trial, larger reductions in mortality risk in the metoprolol group could be seen as compared to placebo, which is consistent with the findings from by-region analysis. The treatment-by-region interaction was supported by a funnel plot, produced by Dr. Robert Fenichel, the secondary medical reviewer of this NDA, as well as the observations of use of ACE inhibitors and others in Merit Study. Dr. Fenichel's plot illustrated the peculiarity of the mortality outcome in the U.S. patients as compared to that in the European patients (See the Appendix). Knowing this treatment-by-region interaction it is necessary to interpret the mortality outcome for these two regions separately.

In general, the approval of an NDA relies on a significant efficacy outcome based on the overall patient population and totality of the evidence which must be convincing. Any serious deficiency in even one important aspect, for instance, a significant inconsistency in treatment outcome or serious adverse effect in a few patients, may jeopardize the approvability or limit the scope of indication. It should be pointed out that the US patient population analysis is never intended to be used to conclude the overall efficacy of metoprolol which has been established by the pre-specified primary endpoints based on the overall population in this review. It is improper to use a significant subgroup finding to draw conclusion for a clinical trial when there is no significant outcome based on the overall population. However, in this review, the US population analysis is done with an intention to check the internal consistency of the overall trial result. In this review, this reviewer does not require that a significant efficacy finding with respect to mortality be established in US patients, but thinks that it is necessary for US mortality outcome at least not to contradict the overall mortality result.

This reviewer has serious concern about the lack of an indication of the effectiveness of metoprolol on mortality in the U.S. patients. For this NDA, one may wonder whether or not a mortality indication should be granted to metoprolol because of the uncertainty surrounding the mortality outcome in the US population. For future CHF trials one may need to think of requiring to include a substantial number of U.S. patients.

(iii) Potential unblinding of treatment codes and biases

The data of the Merit Study indicated that metoprolol could significantly lower heart rate. The mean reduction in heart rate associated with the metoprolol treatment at Day 90 was 14.9 (bpm) as compared to 2.8 (bpm) for placebo. This obvious heart rate lowering effect could potentially lead to the knowledge of the treatment codes when treating patients. In fact, with a properly chosen rule, the rate of successful guess of patients' treatment assignments could be as high as 80% (Table 2.3). With the knowledge of treatment codes, biases could be introduced into the trial in various ways. For example, the biases could be introduced by differentiating interventions other than the study medication (use of concomitant medications, initiation of hospitalizations, etc.) in favor of metoprolol. Biases could also be introduced during classification of outcome events such as cause-specific clinical events. These biases in general are very difficult to detect during the review because of a lack of detailed information of the actual conduct of the trial, possibly small contribution to the biases from each individual source even though the accumulated impact of all kinds of biases is large, other confounding factors, etc. The large sample size could make the trial more sensitive to biases.

In the limited analyses to explore possible biases, this reviewer noted some numerical patterns in patients' hospitalization associated with the reduction of heart rate and in the use of ACE inhibitors. The findings might give some clues for a possible difference in clinical practice between U.S and European countries. No apparent operational bias was found.

(iv) Uncontrolled use of ACE inhibitors

In the Merit Study, the majority of the patients had used ACE inhibitors, a known class of drugs with beneficial mortality effect in CHF patients. This uncontrolled use of ACE inhibitors may significantly complicate the interpretation of the outcomes of beta-blocker heart failure trials. In general, this may be due to the fact that (a) the use of ACE inhibitors in such a trial is often the consequence or the outcome of the effect of the treatment and in turn it has impact on the final outcome of the trial, (b) the use of ACE inhibitors can be uneven across the treatment groups and can not be taken care of by the original randomization, and (c) the use of ACE inhibitors can be a source of biased intervention in favor of the experimental drug when there is a high probability of knowing or guessing correctly patients' treatment assignments.

The Merit Trial, as intended, only provided the information of the effectiveness of metoprolol used with concomitant medications, especially ACE inhibitors. The trial did not provide the information about the effectiveness of metoprolol used as a single agent in CHF patients. To this reviewer, the impact of the uncontrolled use of ACE inhibitors on the outcome in beta-blocker CHF trials is unclear, and deserves more attention in design of such trials in the future.

4. Conclusion

The Merit Study on its own demonstrated a beneficial effect of metoprolol in treating patients with CHF. However, a treatment-by-region interaction (U.S. vs. Europe) with respect to mortality was found. The benefit of metoprolol treatment in mortality was only limited to the European patients. The metoprolol treatment apparently had no effect on mortality in U.S. patients. The observed relative risk (rr=1.05, metoprolol vs. placebo) for U.S. patient population was significantly different from that (rr=0.55) for the European patient population with a p-value of 0.0030 for this difference. Hence, it is necessary to interpret the mortality outcome for these two regions separately. A funnel plot by Dr. R. Fenichel, the secondary medical reviewer, also supported the view that this difference was unlikely to occur purely by chance. Question is whether or not a mortality indication should be granted to metoprolol because of the uncertainty around the effect of the drug in U.S. Patients.

The Merit Study also demonstrated a statistically significant beneficial effect of metoprolol with respect to the second primary endpoint, the combined mortality/hospitalization endpoint (rr=0.81, the nominal p=0.0001) in the overall patient population. This result appeared consistent across the two regions (rr=0.84 for U.S. and rr=0.81 for Europe). This reviewer noted that the outcome of this combined endpoint was mainly driven by the hospitalization component in the U.S. population.

Here, this reviewer wants to emphasize that the U.S. patient population analysis is never intended to be used to conclude the overall efficacy which has been established by the pre-specified primary endpoints based on the overall population in this review. The U.S. population analysis is done with the intention to check the internal consistency of the

overall trial result. Because of the goal of this NDA submission after all is to market the drug for treating patients with CHF in the United States, the outcome in the U.S. patients is particularly of interest. In this review, this reviewer does not require that a significant efficacy finding with respect to mortality be established in U.S. patients, but thinks that it is at least necessary for U.S. mortality outcome not to contradict the overall mortality result.

The Merit Study was intended to demonstrate the effects of metoprolol as an adjunctive therapy for congestive heart failure (CHF) when added to optimal standard treatment consisting of other medications, especially ACE inhibitors, in treating patients with CHF. The Merit Study provided no information about effects of metoprolol used as a single agent.

/S/

Lu Cui
Ph.D., Mathematical Statistician
5/30/2000

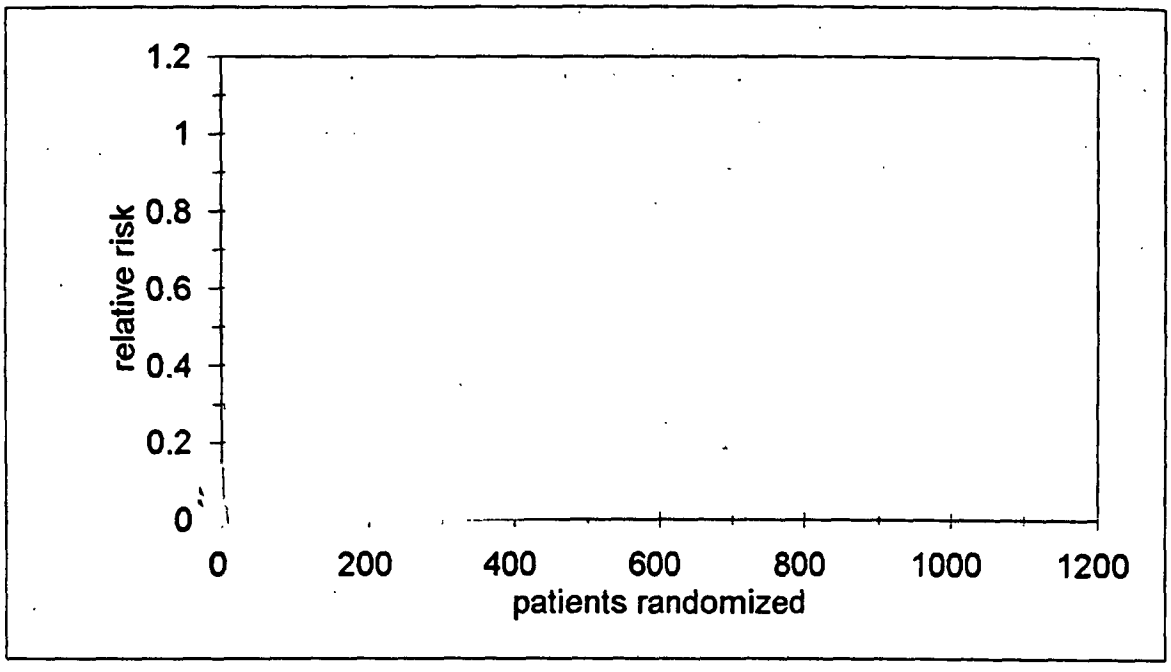
Concur: Dr. H.M. James Hung /S/ 05/30/00

Dr. George Chi /S/

cc:
NDA #19,962, metoprolol succinate
HFD-110
HFD-110 / Dr. Lipicky
HFD-110 / Dr. Fenichel
HFD-110 / Dr. Duarte
HFD-110 / Ms. McDonald
HFD-344 / Dr. Barton
HFD-710 / Dr. Chi
HFD-710 / Dr. Mahjoob
HFD-710 / Dr. Hung
HFD-710 / Dr. Cui
HFD-710 / Chron.

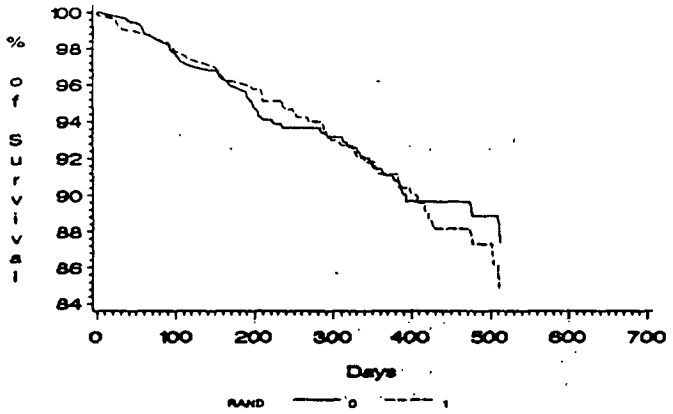
Appendix

The following graph was produced by Dr. Robert Fenichel, the secondary medical reviewer of metoprolol. The graph shows how the estimated relative risks for the European countries converge with increase of sample size. The outcome in US with the largest sample size clearly does not fit into this pattern.

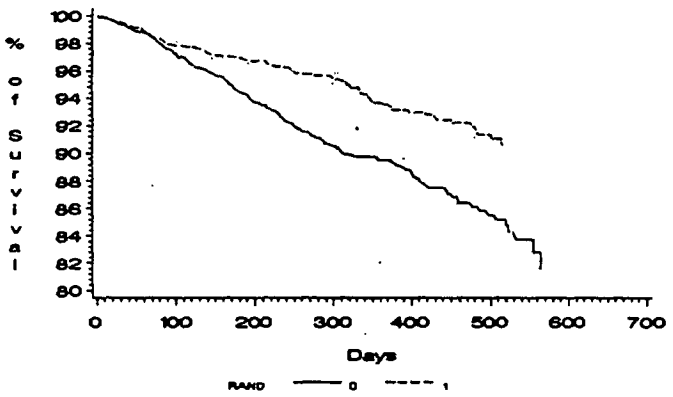


APPEARS THIS WAY
ON ORIGINAL

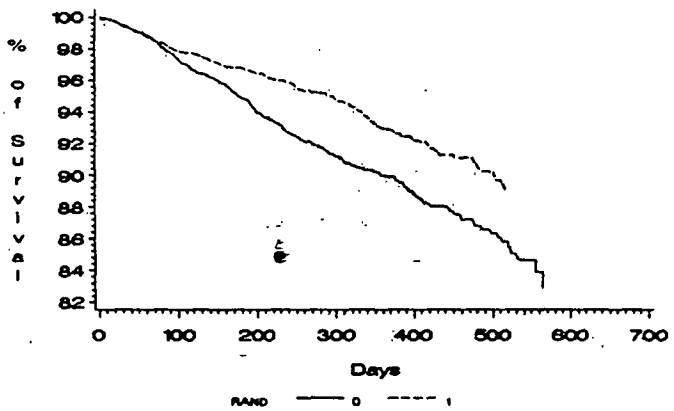
K-M estimate of surv function / mortality / US patients (rand=1 for meto)



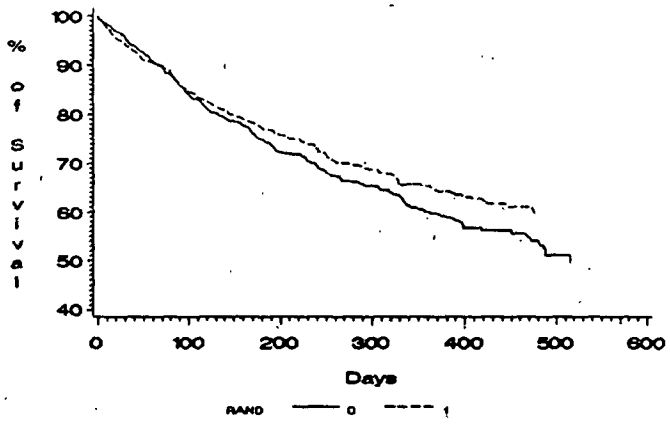
K-M estimate of surv function / mortality / Euro patients (rand=1 for meto)



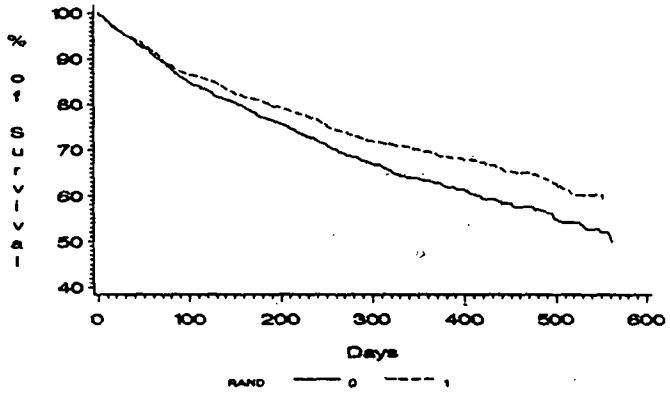
K-M estimate of surv function / mortality / all patients (rand=1 for meto)



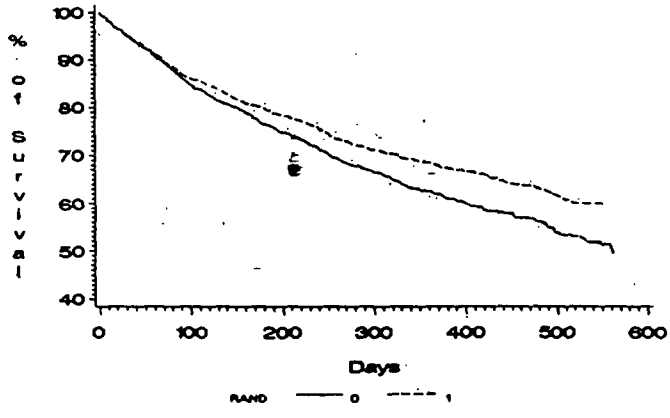
K–M estimate of surv function / mort & hosp / US patients (rand= 1 for meto)



K–M estimate of surv function / mort & hosp / Euro patients (rand= 1 for meto)



K–M estimate of surv function / mort & hosp / all patients (rand= 1 for meto)



Statistical Review and Evaluation
(Addendum)

NDA: 19,962
Applicant: Astra Zeneca LP
Drug Name: Toprol-XL (metoprolol succinate)
Indication: Congestive Heart Failure (CHF)
CDER receiving date: 9/15/1999, 10/17/2000
Document Reviewed: submissions of 9/14/2000 and 10/16/2000

This addendum is to comment on the sponsor's submissions dated 9/14 and 10/16/2000 for NDA 19-962/S-013 (metoprolol) and further clarify some important issues.

(1) Role of protocol and task of reviewer

The protocol in a confirmatory clinical trial sets up rules to follow. However, such rules, though necessary, may not be sufficient to guarantee the quality of the trial. The reviewer's responsibility is to evaluate the quality of the clinical trial and to seek the assurance of the quality of the trial in all aspects. For this purpose, the reviewer needs to look into all spots of the trial which are strategically important for drug approval, even though some of them are not pre-specified in the protocol.

(2) Necessity of US subgroup analysis

Since the primary objective of this NDA submission for metoprolol is to get approval for marketing metoprolol in the United States, there is no question about the necessity for assessment of the treatment effects of metoprolol in the US patient population. The purpose of such assessment is to make sure that the outcomes with respect to the primary variables are consistent with the findings in the overall patient population. Because of possible variation, this reviewer does not expect to see the same magnitude of the treatment effect of metoprolol for the US population as compared to the overall patient population. However, this reviewer does expect to see a mortality outcome in US patients not contradicting the overall finding if the US result is consistent with the overall result. Unfortunately, the US result apparently differs from the overall one. The observed strong beneficial effect of metoprolol on mortality (risk ratio=0.66, metoprolol vs. placebo) for the overall patient population is completely diminished (risk ratio =1.05) for the US patient population. Two statistical tests indicate a significant treatment-by-region interaction (US vs. Europ, see this reviewer's original review). In addition, a new conditional analysis, suggested by Dr. Robert O'Neill, is performed by this reviewer. This analysis indicates a very small chance for observing a mortality effect of metoprolol as extreme as or more extreme than the observed one for US patient population (Appendix). Depending on the assumption on the size of the true effect, this chance can be 0.1% (the observed treatment effect for the European population reflects the true effect), or 1.2% (the overall effect reflects the true one) or 2.8% (the effect for the European population minus 2 times of the standard deviation reflects the true effect). The issues now are (i) the validity of the US subgroup mortality analysis and (ii) the believability of the finding. The second issue has been addressed in this reviewer's original NDA review dated 5/30/2000. The following section is to discuss the scientific basis for the US subgroup analysis.

(3) Validity of US subgroup analysis

Knowing the potential pitfall in interpretation of post-hoc subgroup analyses, this reviewer has paid great attention to assessing the validity of this US subgroup analysis. From a statistical point

of view, in this reviewer's opinion, there are three important factors determining the quality of a subgroup analysis: (i) randomization (the most important one), (ii) multiplicity involved, and (iii) power. For a typical problematic post-hoc subgroup analysis, the original randomization usually is NOT preserved within the subgroup, leading to great potential for biases to be introduced through unevenly distributed prognostic factors between the treatment groups. Since such a subgroup analysis is often selected from several analyses, the chance for type I error rate can be large due to multiplicity involved. Quite often, the power of such a subgroup analysis is low because of the smaller sample size of the subgroup as compared to the overall population.

The US subgroup analysis for Merit trial is scientifically sound in the sense of no major problems in these three aspects, as explained below.

Because of the by-region stratification in Merit trial, the US subgroup has a valid randomization. Contrarily, most of sponsor's pre-specified subgroup analyses do not have a stratified randomization.

As mentioned above, the US subgroup analysis is not a fishing-around practice for getting surprising findings but a well focused analysis to check the primary outcomes in the US subgroup which is directly linked to the objective of this NDA submission. There is much less concern on possible multiplicity involved for the US subgroup analysis. Further, even with the sponsor's very conservative adjustment for multiplicity (Bonferroni adjustment for 13 interaction analyses), which is not usually performed for a test of interaction because of its known low power, the treatment-by-region interaction still deserves attention ($p=0.078<0.10$).

The insignificant mortality result in the US patient population apparently is not due to a reduced power for the subgroup but a significant shift in risk ratio from the overall 0.66 to 1.05 for US patients. In this case, on one hand, there is no basis to talk about the power of the test because the treatment difference seems to be null in US. On the other hand, the substantially large sample size of the US subgroup should be recognized which in fact is larger than the total sample size of all four US studies in the clinical development program for carvidilol, a beta blocker approved for treating heart failure.

This reviewer agrees that the Hill's rule, as pointed out by the sponsor, gives a good guideline for an indirect assessment of the believability of a subgroup finding. However this rule does not mention the direct evaluation of the quality of a subgroup analysis. To this reviewer, the quality of the US subgroup analysis for mortality is unusually good which makes this analysis very different from the typical post-hoc subgroup analyses which are often problematic as exemplified by Peto's ISIS-2 fallacy.

(4) The sponsor's analyses

(i) Analysis of treatment-by-country interaction

Unlike the sponsor's treatment-by-country interaction test, this reviewer focused on the treatment-by-region interaction (US vs. Europe) test for the following reasons.

The goal of the reviewer's analysis is to contrast the US outcome with the non-US outcome. In this reviewer's opinion, the mortality outcomes in the 13 European countries are quite consistent and it seems reasonable to view the mortality outcome for the European patients as a whole. In fact, for ten out of 13 European countries, the risk ratios in mortality are under or about 0.60. The test of homogeneity of the mortality effects of metoprolol in the thirteen European countries yields $p=0.5898$. Such consistency can also be seen from the funnel plot as suggested by Dr. Fenichel (see the original review) which takes the sample size into account. The sponsor believes

that the mortality outcomes are not homogeneous in the European countries due to the different background in epidemiology, social economic characteristics or standard of care. If this is true, to this reviewer, the US mortality outcome needs to be viewed as its own, and such a regional difference is a plausible explanation of the seemingly null mortality effect in US.

(ii) The sponsor's 12 by-subgroup analyses

The sponsor performed twelve treatment-by-subgroup interaction analyses, including treatment-by-diabetic, by-gender, by-blood-pressure subgroup analyses, and etc. The insignificant results of these interaction analyses, in the sponsor's opinion, weaken the believability of the US mortality finding. This reviewer is afraid that the sponsor's analyses provide little information for the assessment of the believability of the US finding. This is because the involved sub-populations, for example, the diabetic patient population, gender group, and hypertensive patient population, are clearly different from the US patient population. It is logically difficult to use the findings based on these subgroups to explain the mortality outcome in the US subgroup.

(5) Summary

The mortality analysis for the US patient population is quite unusual which makes it very different from typically problematic subgroups analyses which are often difficult to interpret. In this reviewer's view, the mortality outcome for the US patient population deserves special attention.

Lu Cui /S/
Ph.D., Mathematical Statistician
11/3/2000

Concur: Dr. H.M. James Hung /S/ 11/20/00

Dr. George Chi /S/ 11/21/00

cc:
NDA #19,962, metoprolol succinate
HFD-110
HFD-100 / Dr. Temple
HFD-110 / Dr. Lipicky
HFD-110 / Dr. Stockbridge
HFD-110 / Dr. Duarte
HFD-110 / Ms. McDonald
HFD-344 / Dr. Barton
HFD-710 / Dr. Chi
HFD-710 / Dr. Mahjoob
HFD-710 / Dr. Hung
HFD-710 / Dr. Cui
HFD-710 / Chron.

Appendix (conditional analysis)

Let the true mortality effect of metoprolol be d. Let p_0 =9.1% and p_1 =9.6% are the estimated mortality rates for placebo and metorolol in US patients, respectively. Let D be the observed mortality effect of metoprolol for the US patient population, D= p_1-p_0 =0.5%. The chance for observing a mortality effect for the US patient population more extreme than D, with the sample size of 535 per treatment group, is

$$P = Pr(Z>(0.005-d)/s),$$

where Z ~N(0,1), d is the assumed true effect, and $s= (p_0(1-p_0)/535+p_1(1-p_1)/535)^{1/2}$. The following table gives the values of P under different assumptions on d.

Table 1. Probability for more extreme mortality outcome than the observed one For US patient population

Assumed true effect (d=metoprolol mortality rate - placebo motality rate)	Prob for the extreme
The observed effect for European population as the true effect (d=-0.050)	0.0010
The observed overall effect as the true effect (d=-0.035)	0.0123
The observed effect for European population –2 sdv as the true one (d=-0.029)	0.0284