



U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Pharmacoepidemiology and Statistical Science
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/Serial Number: 21-686
Drug Name: Exanta (ximelagatran) 36 mg bid oral formulation
Indication(s): reduction of risk of stroke or systemic embolic events in patients with atrial fibrillation
Applicant: AstraZeneca
Date(s): December 23, 2003
Review Priority: Standard

Biometrics Division: Biometrics I (HFD-710)
Statistical Reviewer: John Lawrence
Concurring Reviewers: Jim Hung, Kooros Mahjoob

Medical Division: Cardiorenal (HFD-110)
Clinical Team: Mehul Desai, Thomas Marciniak

Project Manager: Alice Kacuba

Keywords: Active control/ non-inferiority, meta-analysis

Table of Contents

1. EXECUTIVE SUMMARY	3
1.1 CONCLUSIONS AND RECOMMENDATIONS	3
1.2 BRIEF OVERVIEW OF CLINICAL STUDIES	3
1.3 STATISTICAL ISSUES AND FINDINGS	3
2. INTRODUCTION	4
2.1 OVERVIEW.....	4
2.2 DATA SOURCES	5
3. STATISTICAL EVALUATION	6
3.1 EVALUATION OF EFFICACY	6
3.2 EVALUATION OF SAFETY	14
4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS	15
4.1 GENDER, RACE AND AGE	15
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS	16
5. SUMMARY AND CONCLUSIONS	17
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE	17
5.2 CONCLUSIONS AND RECOMMENDATIONS	17
APPENDIX I.....	18
APPENDIX II	19
APPENDIX III.....	20

1. EXECUTIVE SUMMARY

1.1 Conclusions and Recommendations

Based on one double blind study of exanta versus the active control warfarin, there is very little evidence that exanta is effective at reducing the risk of the combined incidence of stroke or systemic embolic events. The most easily interpretable scenario would be if the effect of warfarin versus placebo was known to be large and was estimated precisely and if exanta had beaten or nearly beaten warfarin in this study. Here, we have a scenario where the magnitude of the effect of warfarin versus placebo is not precisely known for this patient population. Moreover, warfarin was numerically better than exanta (using the point estimate) in the double-blind study and the difference was nearly statistically significant.

1.2 Brief Overview of Clinical Studies

In the submission, there are results from two efficacy studies (one open-label, one double-blind) for the treatment of patients with atrial fibrillation. In both studies, patients with chronic nonvalvular atrial fibrillation and at least one risk factor for stroke were randomized to the test drug exanta or the active control warfarin. The dose of exanta was 36 mg bid and warfarin was titrated to achieve an international normalized ratio (INR) between 2 and 3. The primary endpoint was the proportion of patients who experienced the combined endpoint of systemic embolic event, ischemic stroke, or hemorrhagic stroke. SPORTIF III (the open label study) enrolled 3407 patients, while SPORTIF V (the double blind study) enrolled 3922 patients. This review will mainly discuss the results of the double-blind study.

1.3 Statistical Issues and Findings

There are some technical statistical issues related to the way that the meta-analysis of the historical studies of warfarin relative to placebo was done. This review presents some discussion of these issues and an alternative meta-analysis. This is an important component of the interpretation of SPORTIF III and SPORTIF V because the non-inferiority margin is in part derived from this meta-analysis (in addition to clinical judgement). On the primary endpoint, both studies failed to show a difference between exanta and warfarin. The point estimates of the event rates in SPORTIF III were $40/2446=1.64\%$ (exanta) and $56/2440=2.30\%$ (warfarin) for an estimated difference of -0.66% [95% CI for risk difference = (-1.4%, 0.13%), $p = 0.10$] or a decreased risk of 29% [95% CI for risk ratio = (0.48, 1.06), $p = 0.10$]. The point estimates of the event rates in SPORTIF V were $51/3160=1.61\%$ (exanta) and $37/3186=1.16\%$ (warfarin) for an estimated difference of 0.45% or an increased risk of 39%. Since SPORTIF V was

double-blind, it could be considered as the pivotal efficacy study and the other study (SPORTIF III) serves as a supportive study that provides additional safety information. In the only double-blind study of exanta versus warfarin, the difference in the rate of the primary endpoint has a point estimate of 0.45% (in favor of warfarin) with a 95% confidence interval of (-0.13%, 1.03%). The lower limit of the confidence interval (the best case scenario for exanta) would give a miniscule benefit to exanta over warfarin. The upper limit is below the noninferiority margin of 2% that was pre-specified by the sponsor, but reflects a potential loss of about 1% of the effect of warfarin. The noninferiority margin of 2% may be too liberal and earlier letters from the FDA to the sponsor conveyed this. Using the risk ratio, the point estimate is $1.61/1.16=1.39$, i.e. a 39% increase in risk for patients using exanta [95% CI = (0.91, 2.12), $p = 0.12$]. Some alternate methods of defining the margin are described in section 3.1 of this review and these alternate methods give smaller margins than that proposed by the sponsor. There is some uncertainty about the magnitude of the effect of warfarin relative to placebo because of the variability between the six historical trials in terms of their design and the observed results. Consequently, there is a great deal of uncertainty about whether exanta retains a significant portion of the benefit of warfarin, and even if exanta is better than placebo.

2. INTRODUCTION

2.1 Overview

In the submission, there are results from two efficacy studies (one open-label, one double-blind) for the treatment of patients with atrial fibrillation. In both studies, patients with chronic nonvalvular atrial fibrillation and at least one risk factor for stroke were randomized to the test drug exanta or the active control warfarin. The dose of exanta was 36 mg bid and warfarin was titrated to achieve an international normalized ratio (INR) between 2 and 3. The primary endpoint was the proportion of patients who experienced the combined endpoint of systemic embolic event, ischemic stroke, or hemorrhagic stroke. SPORTIF III (the open label study) enrolled 3407 patients, while SPORTIF V (the double blind study) enrolled 3922 patients.

This review will briefly discuss the six trials comparing warfarin to placebo. Then, there is a discussion about the appropriateness of the meta-analysis and the choice of the non-inferiority margin. Finally, the double-blind study, SPORTIF V, is reviewed.

2.2 Data Sources

All electronic documents were obtained from the CDER document room in location \\CDSesub1\n21686\N_000\2003-12-23

The electronic study report, statistical analysis plan, protocol and amendments for SPORTIF III and SPORTIF V contained in \\Cdsesub1\n21686\N_000\2003-12-23\clinstat\af\controlled\sh-tpa-0003 and \\Cdsesub1\n21686\N_000\2003-12-23\clinstat\af\controlled\sh-tpa-0005

The electronic SAS transport data sets for SPORTIF V, \\Cdsesub1\n21686\N_000\2003-12-23\crt\datasets\SH-TPA-0005\STROKE.xpt

The following journal articles provided electronically by the sponsor in the study report appendix:

Stroke Prevention in Atrial Fibrillation Investigators. Stroke prevention in Atrial Fibrillation study: Final Results. *Circulation* 1991;84:527-539

The Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators. The effect of low-dose Warfarin on the risk of stroke in patients with nonrheumatic Atrial Fibrillation. *New Engl J Med* 1990;323:1505-1511.

Ezekowitz MD, Bridgers SL, James KE, Carliner NH, et al. Warfarin in the prevention of stroke associated with nonrheumatic Atrial Fibrillation. *N Engl J Med* 1992;327:1406-1412.

EAFT (European Atrial Fibrillation Trial) Study Group. Secondary prevention in non-rheumatic Atrial Fibrillation after Transient Ischemic Attack or minor stroke. *Lancet* 1993;342:1255-1262.

Conolly SJ, Laupacis A, Gent M, Roberts RS, Cairns JA, Joyner C. Canadian Atrial Fibrillation Anticoagulation (CAFA) study. *J Am Coll Cardiol* 1991;18:349-355.

Petersen P, Boysen G, Godtfredsen J, Andersen E.D, Andersen B. Placebo-controlled, randomised trial of Warfarin and Aspirin for Prevention of thromboembolic complications in chronic Atrial Fibrillation. *The Lancet* 1989;175:179

The following journal articles referenced in this review:

Farrington CP, Manning G. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference and non-unity relative risk. *Statistics in Medicine* 1990; 9:1447-1454

Grambsch P, Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; 81:515-526

Hung HMJ, Wang SJ, Tsong Y, Lawrence J, O'Neil RT. Some fundamental issues with non-inferiority testing in active controlled trials. *Statistics in Medicine* 2003; 22:213-225

Rothmann M, Li N, Chen G, Chi GYH, Tsou HH. Non-inferiority methods for mortality trials. *Proceedings of the Biopharmaceutical Section of the American Statistical Association* 2001.

Holmgren EB. Establishing equivalence by showing that a pre-specified percentage of the effect of the active control over placebo is maintained. *Journal of Biopharmaceutical Statistics* 1999;9 (4):651 –659.

3. STATISTICAL EVALUATION

3.1 Evaluation of Efficacy

There are six studies of warfarin versus placebo. Table 1 shows the event rates, estimated risk difference, risk ratio, and confidence intervals for the six historical studies of warfarin versus placebo. The numbers are the same as in Table 1 of the study report after the amendment (the study report did not include risk differences or confidence intervals).

Table 1 Summary of the six historical studies of warfarin versus placebo.

Study	Summary	Events/patient years		Risk difference (95% CI)*	Risk ratio (95% CI)*
		Warfarin	Placebo		
AFASAK	open label. 1.2 yr follow-up	9/413 = 2.18%	21/398 = 5.28%	-3.10% (-5.71, -0.49)	0.41 (0.19, 0.89)
BAATAF	open label. 2.2 yr follow-up	3/487 = 0.62%	13/435 = 2.99%	-2.37% (-4.12, -0.63)	0.21 (0.06, 0.72)
EAFT	open label. 2.3 yr followup. patients with recent TIA	21/507 = 4.14%	54/405 = 13.3%	-9.19% (-12.9, -5.45)	0.31 (0.19, 0.51)
CAFA	double blind. 1.3 yr followup	7/237 = 2.95%	11/241 = 4.56%	-1.61% (-5.02, 1.79)	0.65 (0.26, 1.64)
SPAF I	open label. 1.3 yr followup	8/260 = 3.08%	20/244 = 8.20%	-5.12% (-9.15, -1.09)	0.38 (0.17, 0.84)
SPINAF	double blind. 1.7 yr followup	9/489 = 1.84%	24/483 = 4.97%	-3.13% (-5.40, -0.85)	0.37 (0.17, 0.79)

*Estimates and Wald-type confidence intervals calculated using SISA software at <http://home.clara.net/sisa/index.htm>

A margin of 2% was used in these studies to show non-inferiority of ximelagatran. In the statistical analysis plan, an estimated overall risk reduction of 0.64 (warfarin versus placebo) with a confidence interval of (0.52, 0.73) is given. This comes from combining the data from all six studies in a fixed effects meta-analysis. This meta-analysis also makes the assumption that the hazard is constant across time in all treatment groups in all studies (i.e. event times are exponential) to estimate the standard errors within each study. The Statistical Analysis Plan comments that this gives a more efficient estimate of the event rates. This is true if the event times are truly exponential and there is non-informative censoring. However, if the event times are not exponential, then it may not be true. One would need the original data sets from all the studies to check all of these assumptions. The justification for the margin of 2% in the study report and statistical analysis plan comes from the point estimate of 0.64 (for the relative risk of warfarin vs. placebo) and an expected event rate in this study of roughly 3.1%. Taking all of this for granted, it is still unclear how the sponsor derives a noninferiority margin of 2%. The actual observed event rates were much smaller (1.2% and 1.6%). Hence, however the margin of 2% was derived, it is not valid if it depends on the assumption of an event rate close to 3.1%. It's particularly confusing because the meta-analysis combines the studies to obtain a global estimate of a risk ratio, but the margin is defined as a risk difference.

As can be seen in the table, most of these studies were open label studies. Open label studies can provide some evidence of the effectiveness of a drug, especially when there are a large number of such studies with an objective endpoint. But, the estimates of the effect and associated standard errors from such studies may not be as reliable as those from a double-blind study (due to conscious or unconscious factors that could bias the results). Also, the magnitude of the effect and the actual event rates in the EAFT study suggest that this study should not be combined together with the others in a meta-analysis. Furthermore, the patient population studied in the EAFT study seems to be different from the patient population studied in the remaining five historical studies because only patients with a recent TIA or stroke were enrolled in that study. Most importantly, it may also be different than the population studied in SPORTIF III and SPORTIF V.

A random effects model may be more appropriate than a fixed effects model for the purpose needed here because the random effects formulation allows for small differences in the treatment effect between studies. I will use the method described by DerSimonian and Laird (1986) to fit the random effects model. The distribution of the parameter estimates are generated by a resampling method discussed in Appendix I.

Using only the double-blind studies of warfarin vs. placebo (CAFA and SPINAF), the point estimate of the global treatment effect for the risk difference is -2.66% and the 95% confidence interval is (-6.81%, 1.49%). Since this confidence interval includes 0, there is not enough evidence from the two double-blind studies combined to conclude that warfarin is different from placebo. If one wanted to rely on these two studies alone to define a non-inferiority margin, then one could argue that an interpretable result would occur only if ximelagatran is shown to be superior to warfarin.

Using all of the studies in the table except EAFT, the point estimate of the treatment effect (risk difference) is -2.81% and the 95% confidence interval is $(-4.26\%, -1.36\%)$. One proposal for selecting the margin is to take half of the magnitude of the worst limit of this confidence interval (the 95-95 method). This would give a margin of 0.68% , far from the actual margin used by the sponsor for this study (2%). The ideas described in several articles (e.g. Holmgren (1999), Rothmann et al (2003), Hung et al (2003)) have given rise to a method of defining a margin that is always larger than the margin from the 95-95 method (I will call it the Holmgren method). In this case, the margin from this method is 1.24% (see Appendix I for the details).

Finally, if all six studies are combined together in a random effects meta-analysis, the point estimate (risk difference) is -3.75% and the 95% confidence interval is $(-5.98\%, -1.52\%)$. The margins from the 95-95 and Holmgren methods would be 0.76% and 1.35% respectively.

It can be argued that the risk ratio may be a more appropriate measurement for across trial comparisons on the assumption or empirical observation that the risk ratio is usually more stable than the risk difference across trials. We can do the same type of calculations as above to obtain a margin on the risk ratio scale and test for non-inferiority using the risk ratio. Table 2 contains the margins for each of the methods of testing using different subsets of the historical studies.

Table 2 Margin using different methodologies.

Studies included	Method of defining margin	Margin
Risk Difference		
CAFA + SPINAF	NA [*]	0%
All except EAFT	95-95	0.68%
	Holmgren ^{**}	1.24% ^{**}
All	95-95	0.76%
	Holmgren ^{**}	1.35% ^{**}
Risk Ratio		
CAFA + SPINAF	NA [*]	1.00
All except EAFT	95-95	1.23
	Holmgren ^{**}	1.56 ^{**}
All	95-95	1.38
	Holmgren ^{**}	1.65 ^{**}

^{*}Since there is not enough evidence from these studies to prove that warfarin is better than placebo, it does not make sense to allow exanta to be worse than warfarin by any amount regardless of the method.

^{**}The margin calculated this way depends on the sample size of the active control study and other nuisance parameters in addition to the constancy assumption (see Appendix I for further explanation and comments about this method).

It can be seen from Table 2 that the margin used by the sponsor (2% risk difference) is larger than that found by any of these methods. The margin on the risk ratio scale,

depending on which studies are used and the method, could be anywhere from 1.00 to 1.65.

SPORTIF V enrolled and randomized 3922 patients. The first patient entered the study on 24 July 2000 and the last patient completed their final study contact on 19 June 2003. Patients were enrolled from 61 centers in Canada and 361 centers in the USA. Of these patients, roughly 2/3 of them in both groups remained in the study on their randomized treatment for between 12 and 24 months. About 4% in both groups withdrew consent before the end of the study. About 6% in both groups died during the study. Patients who died during the study (without having the primary endpoint) were censored for the primary endpoint at the time of death. The primary endpoint was the proportion of patients who experienced the combined endpoint of systemic embolic event, ischemic stroke, or hemorrhagic stroke. The demographic summaries of the patients appear in Table 3. No significant differences are seen between the two groups at baseline.

Table 3 Baseline demographic characteristics for SPORITF V.

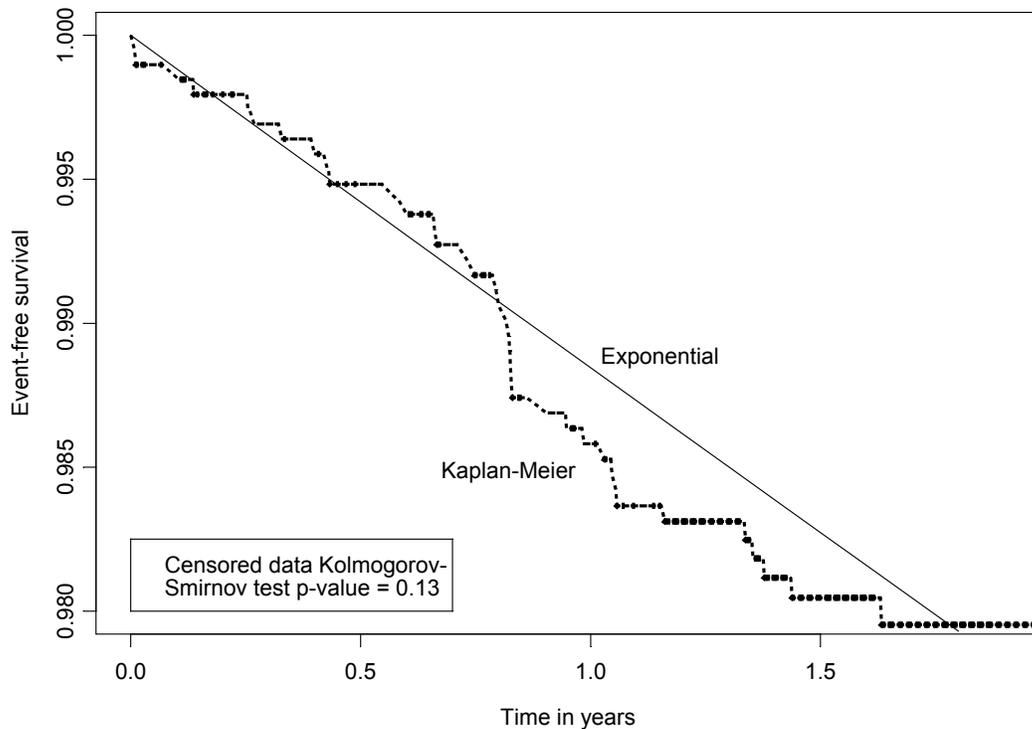
Characteristic		Ximelagatran N=1960	Warfarin N=1962	Total N=3922
Gender	Male	1365 (69.9)	1353 (69.0)	2718 (69.3)
	Female	595 (30.4)	609 (31.0)	1204 (30.7)
Race	Caucasian	1875 (95.7)	1888 (96.2)	3763 (95.9)
	Black	67 (3.4)	58 (3.0)	125 (3.2)
	Asian	15 (0.8)	10 (0.5)	25 (0.6)
	Other	3 (0.2)	6 (0.3)	9 (0.2)
Age	<65	383 (19.5)	401 (20.4)	784 (20.0)
	65 to75	739 (37.7)	741 (37.8)	1480 (37.7)
	75+	838 (42.8)	820 (41.8)	1658 (42.3)
ASA use	No	1398 (71.3)	1393 (71.0)	2791 (71.2)
	Yes	542 (27.7)	544 (27.7)	1086 (27.7)
	Missing	20 (1.0)	25 (1.3)	45 (1.1)
Number of unique stroke risk factors in addition to AF	0	3 (0.2)	4 (0.2)	7 (0.2)
	1	490 (25)	509 (25.9)	999 (25.5)
	2	600 (30.6)	597 (30.4)	1197 (30.5)
	3	472 (24.1)	459 (23.4)	931 (23.7)
	4 or more	395 (20.1)	393 (20.1)	788 (20.1)

Source: Table 29 of study report.

SPORTIV V had several interim analyses for safety with no possibility of stopping early for efficacy. The primary analysis assumed exponential event times with the maximum likelihood estimates of the event rates and standard errors derived from this parametric model. I found some evidence that the event times in the warfarin group do not follow an exponential distribution. Figure 1 shows the Kaplan-Meier curve for the event times in the warfarin group and the corresponding exponential curve using the maximum likelihood estimate. These curves appear to be different and the analog of the

Kolmogorov-Smirnov test for censored data confirms that there is a difference between them [p=0.13]. This test uses the maximum difference between the Kaplan-Meier curve and the best fitting curve in the exponential family. I calculated the p-value by simulation under the null distribution of this test statistic assuming independent exponential event times and censoring times obtained from the Kaplan-Meier estimate [see Appendix II]. Although a p-value of 0.13 can hardly be called convincing evidence against the null hypothesis for a post hoc test, I believe the burden of proof should go in the opposite direction (i.e. I don't need to prove that the distribution is not exponential, rather the data should prove to me that the distribution is exponential).

Figure 1 Kaplan-Meier curve and best fitting exponential curve for warfarin group.

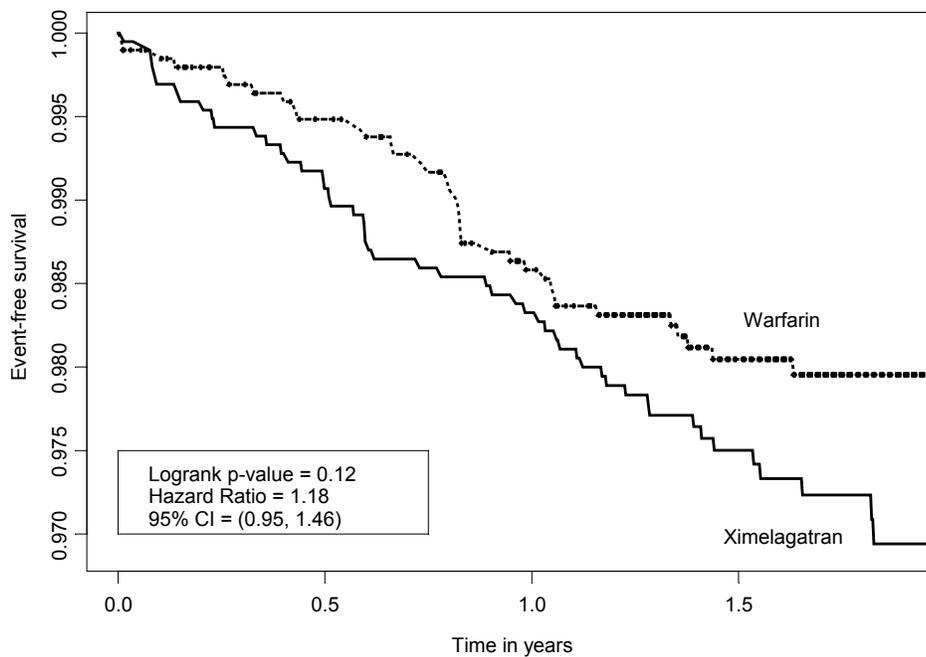


The point estimates of the event rates in SPORTIF V were 51/3160=1.61% (exanta) and 37/3186=1.16% (warfarin). The difference in the rate of the primary endpoint has a point estimate of 0.45% (in favor of warfarin) with a 95% confidence interval of (-0.13%, 1.03%) using the assumption of exponential event times as specified in the protocol. The lower limit of the confidence interval (the best case scenario for exanta) would give a minuscule benefit to exanta over warfarin. The upper limit is below the noninferiority margin of 2% that was pre-specified by the sponsor, but indicates a potential loss of about 1% of the effect of warfarin. Using the risk ratio, the point

estimate is $1.61/1.16=1.39$, i.e. a 39% increase in risk for patients using exanta [95% CI assuming exponential event times= (0.91, 2.12), $p = 0.12$ - see Appendix III]. This point estimate and confidence interval appears also in Table 44 of the sponsor's study report.

As a supportive analysis, I calculated the Kaplan-Meier curves for the two groups without the assumption of exponential event times and I calculated the risk ratio, confidence interval, and p-value semi-parametrically (Cox proportional hazards model or logrank statistic not using exponential event times assumption). The point estimate as well as the upper limit of the 95% confidence interval for the hazard ratio is smaller in this analysis than in the parametric analysis. The curves and these estimates appear in Figure 2. Thus, this analysis makes exanta and warfarin appear to be closer to each other than the parametric analysis. Nonetheless, the curves appear to be clearly different with the warfarin curve appearing superior. I found no evidence that the proportional hazard functions assumption is violated using the test of Grambsch and Therneau ($p=0.982$).

Figure 2 Kaplan-Meier curves and semi-parametric hazard ratio estimates



Returning to the question of non-inferiority, Table 4 shows the results for the tests of noninferiority using the various margins that were described earlier in this section and summarized in Table 2.

Table 4 Results for non-inferiority using various margins defined by various methods

Studies included	Method of defining margin	Margin	Result Fail/Succeed to infer noninferiority
Risk Difference			
SPORTIF V Point estimate and 95% CI Exanta vs Warfarin: 0.45%, (-0.13%, 1.03%)			
CAFA + SPINAF	NA [*]	0%	Fail
All except EAFT	95-95	0.68%	Fail
	Holmgren ^{**}	1.24% ^{**}	Succeed
All	95-95	0.76%	Fail
	Holmgren ^{**}	1.35% ^{**}	Succeed
Risk Ratio			
SPORTIF V Point estimate and 95% CI Exanta vs Warfarin: 1.39, (0.91, 2.12)			
CAFA + SPINAF	NA [*]	1.00	Fail
All except EAFT	95-95	1.23	Fail
	Holmgren ^{**}	1.56 ^{**}	Fail
All	95-95	1.38	Fail
	Holmgren ^{**}	1.65 ^{**}	Fail

^{*}Since there is not enough evidence from these studies to prove that warfarin is better than placebo, it does not make sense to allow exanta to be worse than warfarin by any amount regardless of the method.

^{**}The margin calculated this way depends on the sample size of the active control study and other nuisance parameters in addition to the constancy assumption (see Appendix I for further explanation and comments about this method).

The putative placebo analysis in Section 7.2.1.1 of the study report indicates that the point estimate for exanta vs. placebo would be 0.5 with a confidence interval of (0.3, 0.83). The report mentions that this analysis rests on the constancy assumption for the effect of warfarin vs. placebo across all studies. The report makes an argument for why this assumption should be believed in this case. However, it cannot be proven. One possible rebuttal to this argument is that the predicted event rates for the SPORITF V study was very different from the actual observed event rates. So, this constancy assumption is one potential problem with the analysis among others. First, a random effects model is a more realistic way of combining data across studies, but the reported analysis uses a fixed effects model. Second, this analysis does not require that exanta retain any particular fraction of the effect of warfarin. Third, there is the problem with all meta-analyses in that patients are not randomized to the different studies and therefore the basis for statistical inference is unsound.

In summary, the method that the sponsor used to define the hypothesis for noninferiority is not valid because it was based on an assumed event rate that was very different from what was actually observed in the trials. A more reliable way of defining the hypotheses would be based on the risk ratio. Using the same distributional assumption that the sponsor used (exponential event times), the confidence interval for the risk ratio is (0.91, 2.12). Therefore, the SPORTIF V trial does not rule out a two-fold risk in the

exanta group compared to the warfarin group. None of the analytic methods for defining the margin on the risk ratio scale produce a margin as high as 2. The Kaplan-Meier curves appear to be very different visually with the warfarin group appearing to have less risk over time. Hence, there is a pretty good case from this data that warfarin is actually superior to exanta (two-sided p-value for superiority = 0.12) and no evidence that exanta is noninferior to warfarin unless one uses a very large margin that is not supported by the historical studies of warfarin compared to placebo.

3.2 Evaluation of Safety

According to the SPORTIF V study report, both study drugs were generally well tolerated, with only 354 (18.1%) ximelagatran-treated patients and 300 (15.4%) warfarin-treated patients discontinuing study drug due to AEs. In the safety population, 239 patients had an AE with a fatal outcome (116 [5.9%] ximelagatran; 123 [6.3%] warfarin); 74 of the fatalities (33 [1.7%] ximelagatran; 41 [2.1%] warfarin) occurred during treatment. Bleeding events in the On-Treatment analysis set were significantly ($p < 0.0001$) less frequent in the ximelagatran group (event rate 37%/year) than in the warfarin group (event rate 47%/year). Major bleeding events were numerically less frequent in the ximelagatran group: 63 patients (event rate 2.4%/year) in the ximelagatran group versus 84 patients (event rate 3.1%/year) in the warfarin group. Elevations of ALAT (Alanine aminotransferase) to greater than 3 times the upper limit of normal were noted at a higher incidence during the treatment period in the ximelagatran group (117 patients; 6.0%) than in the warfarin group (15 patients; 0.8%; $p < 0.0001$) and 372 patients experienced liver-related SAEs during the treatment period (245 ximelagatran; 127 warfarin).

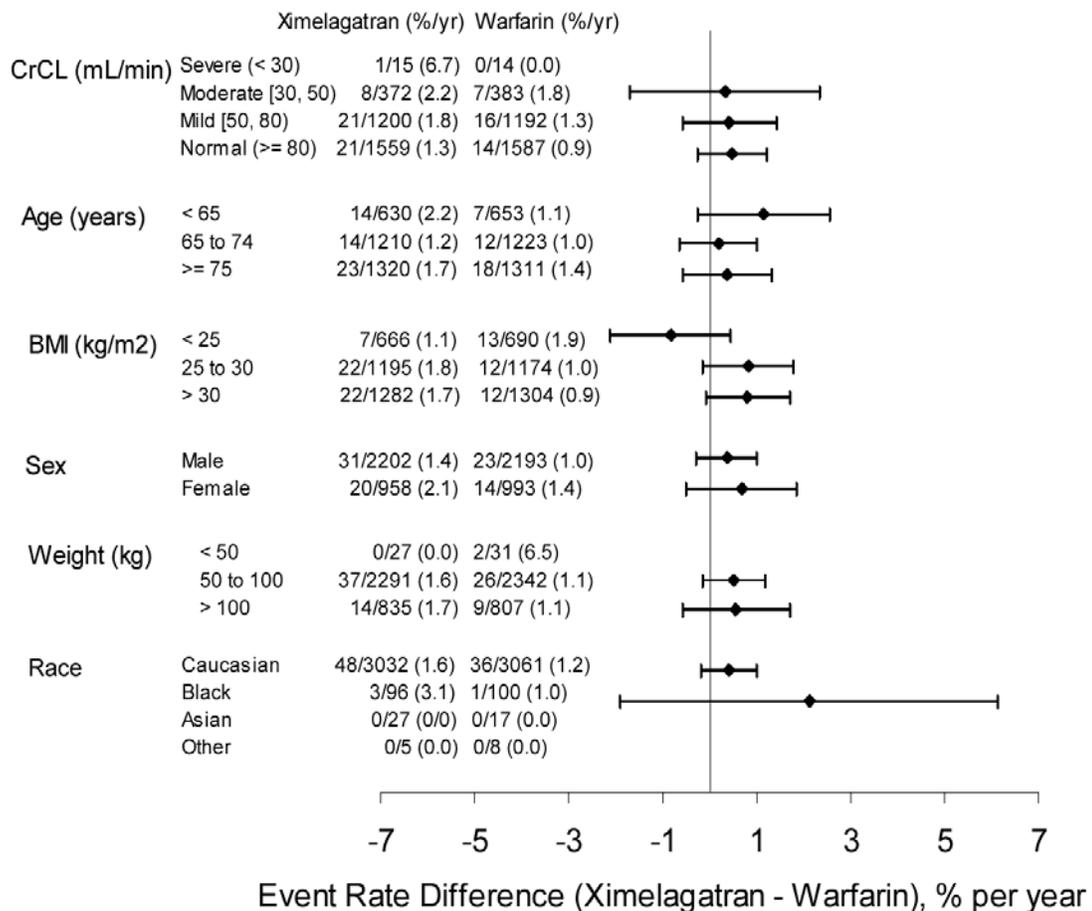
According to the SPORTIF III study report, a total of 185 (11%) ximelagatran-treated patients and 100 (6%) warfarin-treated patients experienced AEs leading to discontinuation of study drug. In the safety population, 145 patients had an AE with a fatal outcome (75 ximelagatran; 70 warfarin) of which 90 (48 ximelagatran; 42 warfarin) occurred on treatment. Haemorrhagic stroke occurred in 4 patients in the ximelagatran group and in 9 patients in the warfarin group; corresponding rates were 0.16% per year and 0.37% per year, respectively. Major bleeds were reported for 29 (1.7%) patients in the ximelagatran group and 41 (2.4%) patients in the warfarin group. Major or minor bleeding events in the On-Treatment analysis set were statistically significantly ($p = 0.007$) less frequent in the ximelagatran group (25.8% per year) than in the warfarin group (29.8% per year). Elevations of ALAT (Alanine aminotransferase) to greater than 3 times the upper limit of normal were noted at a higher incidence in the ximelagatran group (107 patients; 6.3%) than in the warfarin group (14 patients; 0.8%) ($p < 0.0001$). Thirty-three patients experienced liver-related SAEs (21 ximelagatran; 12 warfarin).

4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Gender, Race and Age

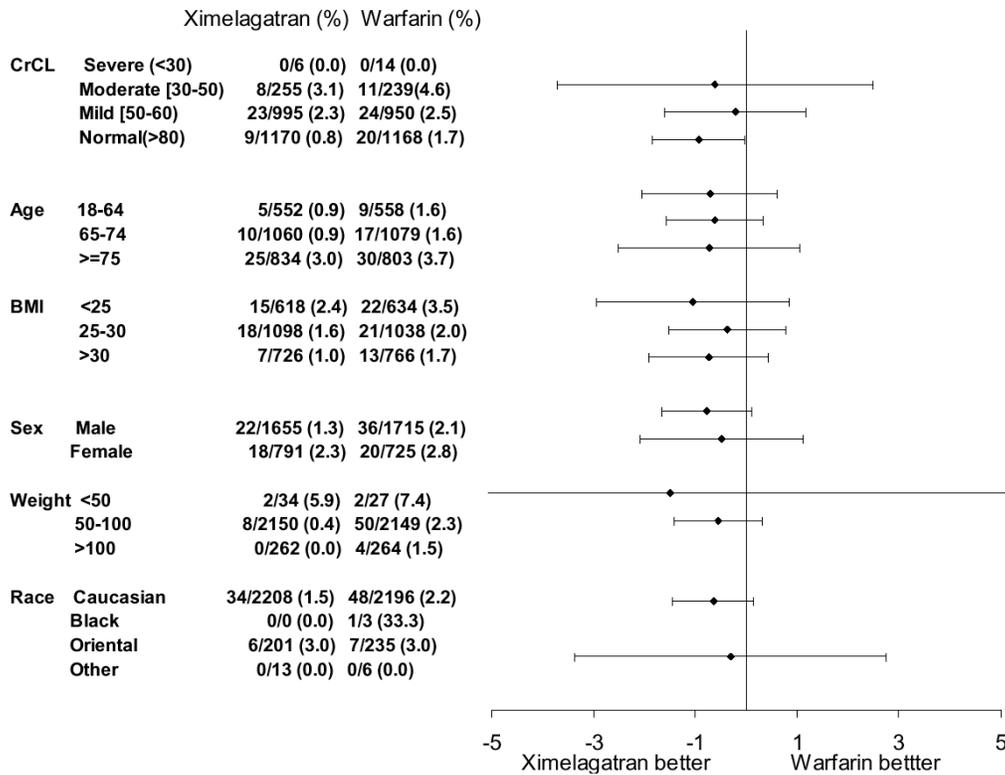
The difference in the rate of stroke or systemic embolic events for the SPORTIF III study in different subgroups are shown in Figure 3. Except for the subgroup of patients with Body Mass Index less than 25 kg/m², there is no other indication of inconsistency across these subgroups. In the remaining subgroups, warfarin appears to be consistently better numerically than exanta.

Figure 3 Difference in event rate (stroke or systemic embolic event) within subgroups-SPORTIF V (Source Figure 16 of study report).



The difference in the rate of stroke or systemic embolic events for the SPORTIF III study in different subgroups are shown in Figure 4. There is no indication of inconsistency across these subgroups. Exanta appears to be numerically better than warfarin across all the listed subgroups.

Figure 4 Difference in event rate (stroke or systemic embolic event) within subgroups-SPORTIF III (Source Figure 13 of study report).



4.2 Other Special/Subgroup Populations

See section 4.1.

5. SUMMARY AND CONCLUSIONS

5.1 Statistical Issues and Collective Evidence

Exanta was not shown to be superior to warfarin in either the open label study (SPORTIF III) or the double blind study (SPORTIF V). The margin for concluding noninferiority (a risk difference of 2%) was too large and was calculated based on an assumed event rate that was much larger than was observed in SPORTIF V. The efficacy results of the two studies were quite different. The point estimates of the event rates in SPORTIF III were 40/2446=1.64% (exanta) and 56/2440=2.30% (warfarin) for an estimated difference of -0.66% [95% CI for risk difference = (-1.4%, 0.13%), p = 0.10] or a decreased risk of 29% [95% CI for risk ratio = (0.48, 1.06), p = 0.10]. The point estimates of the event rates in SPORTIF V were 51/3160=1.61% (exanta) and 37/3186=1.16% (warfarin) for an estimated difference of 0.45% [95% CI = (-0.13%, 1.03%)] or an increased risk of 39% [95% CI = (-9%, +112%)]. There is no obvious reason for the difference in the efficacy results between the two studies based on patient demographics. The only obvious difference between the two studies is that one was open label and the other double blind. In general, the results from double blind studies are more reliable for many reasons.

The safety results were more consistent in the two studies. There were more patients in both studies that discontinued for adverse events related to study drug (18% vs 15% in SPORTIF V and 11% vs 6% in SPORTIF III). There were more bleeding events in the warfarin group compared to the exanta group in both studies. On the other hand, there were more patients with elevations of ALAT (Alanine aminotransferase) to greater than 3 times the upper limit of normal in the exanta group compared to the warfarin group in both studies.

5.2 Conclusions and Recommendations

Since exanta was not clearly safer than warfarin, it should be considered as an option to warfarin only if it has been proven to retain a significant fraction of the effect of warfarin. This fraction, or alternately the noninferiority margin, should be based on the historical data of studies comparing warfarin to placebo and clinical judgement. Based solely on analytical methods and the hypothesis of preservation of 50% of the effect of warfarin, this review shows some values for the margin (Table 2). The actual margin used for these studies (2%) is much larger than any of the margins in that table. Unless the clinical judgement is that a loss of 2% of the effect of warfarin is clinically acceptable, in my opinion exanta has not been demonstrated to be noninferior to warfarin.

APPENDIX I

The goal of the Holmgren method to show that at least 50% of the active control effect is preserved is to test the hypothesis

H_0 : True difference of exanta – warfarin = $\frac{1}{2}$ (True difference of placebo – warfarin)

versus the alternative hypothesis

H_1 : True difference of exanta – warfarin $> \frac{1}{2}$ (True difference of placebo – warfarin)

with a false positive rate of 0.05 (two-sided) under the assumption that both of the parameters (the true differences) can be estimated by statistics that would be observed by repeating both the historical studies and the current study over and over again. Let Δ_{W-P} denote the estimated global treatment difference (warfarin vs. placebo) and Q_{W-P} denote its estimated variability between trials under the assumed random effects model and let Δ_{E-W} denote the estimated treatment difference (exanta vs. warfarin). The test statistic we will use is

$$T = (\Delta_{E-W} + \frac{1}{2} \Delta_{W-P}) / \{ \{ \text{Estimated SE}(\Delta_{E-W}) \}^2 + \frac{1}{4} \{ \text{Estimated SE}(\Delta_{W-P}) \}^2 \}^{1/2}$$

One can resample data from the historical studies from binomial distributions with treatment difference in each trial drawn from a normal distribution with mean Δ_{W-P} and variance Q_{W-P} and group rates defined by the restricted maximum likelihood estimates derived in Farrington and Manning (1990). Also, resample data from the current study from binomial distributions with treatment difference of $-\frac{1}{2} \Delta_{W-P}$ and group rates defined by the restricted maximum likelihood estimates with this difference. Let F denote the distribution of T from these resampled data sets. Under appropriate conditions, F would be close to the distribution function of a standard normal random variable. However, when there are a small number of trials or the event is rare, this approximation may be poor. For this reason, we will resample a large number, say 100 thousand, of data sets to estimate F and test the hypothesis by comparing the observed value of T to the critical value $F^{-1}(0.025)$. Using a little algebra, one can see that this is operationally equivalent to comparing the upper limit of a 95% confidence interval for the true treatment difference (exanta – warfarin) to the non-inferiority margin defined by

$$-\frac{1}{2} \Delta_{W-P} + F^{-1}(0.025) * \{ \{ \text{Est. SE}(\Delta_{E-W}) \}^2 + \frac{1}{4} \{ \text{Est. SE}(\Delta_{W-P}) \}^2 \}^{1/2} + 1.96 \text{ Est. SE}(\Delta_{E-W})$$

This is what I will call the margin using the Holmgren method. Note that this margin depends on the sample size and nuisance parameters estimated from the current study (the average rate). These are needed in calculating $\text{Est. SE}(\Delta_{E-W})$ and F . Hence, the use of the margin can be problematic in terms of interpretation and designing the study (see Hung et al 2003).

APPENDIX II

I estimated the censoring distribution using the Kaplan-Meier estimate where time for each patient is the time for the primary endpoint and the censoring variable is 1 if censored (for the primary endpoint) and 0 otherwise; i.e. the time for censoring is observed if censored for the primary endpoint and censored if the primary endpoint was observed. Then, I simulated 10,000 new data sets under the null distribution with 1962 patients in each data set. For each patient, I sampled a time for primary endpoint and a censoring time independently. I then calculated the Kolmogorov-Smirnov statistic from the simulated data sets by finding the maximum difference between the Kaplan-Meier estimate of the survival curve and the best-fitting exponential curve for that data set. Finally, the conditional p-value is calculated as the proportion of the test statistics that are at least as large as the observed value. My S-plus program follows.

```
# this gives the Kaplan-Meier estimate of the survival
# distribution for the warfarin group
surv1<-survfit(Surv(fdaq2$DAYSP[fdaq2$RXGRP=="Warfarin"],
fdaq2$PRIM[fdaq2$RXGRP=="Warfarin"]=="Yes") ~ 1)

# Maximum-likelihood estimate of the exponential parameter
# for warfarin group
rw<- sum(fdaq2$PRIM[fdaq2$RXGRP=="Warfarin"]=="Yes")/
sum(fdaq2$DAYSP[fdaq2$RXGRP=="Warfarin"])

# observed value of one sample Kolmogorv-Smirnov test statistic
obsks<-max(abs(surv1$surv-exp(-surv1$time*rw)))

# this gives the Kaplan-Meier estimate of the censoring distribution
# for the warfarin group
surv2<-survfit(Surv(fdaq2$DAYSP[fdaq2$RXGRP=="Warfarin"],
fdaq2$PRIM[fdaq2$RXGRP=="Warfarin"]=="No") ~ 1)

# the start of the Kaplan-Meier curve defined to be 1
surv2$surv[1]<-1

# this vector holds the censoring times or the values
# of 0 or 1 to indicate censoring
cens<-rep(0,1962)
# this holds the event times
event<- rep(0,1962)
# this holds the smaller of censoring time or event time
time<- rep(0,1962)
# vector of simulated kolmogorov-smirnov statistics
# under the null hypothesis
kssim<-rep(0,10000)

for (i in 1:10000) {

  # this loop creates a simulated data set by generating a censoring
  # time and survival time for each patient
  for (j in 1:1962) {
    cens[j]<-max(surv2$time[surv2$surv>runif(1)])
    event[j]<-(-log(runif(1)))/rw
```

```

if (cens[j]<event[j]) {
  time[j]<-cens[j]
  cens[j]<-0}
else {
  time[j]<-event[j]
  cens[j]<-1}}

# calculate Kaplan-Meier curve for simulated data
surv<- survfit(Surv(time, cens) ~ 1)
# calculate the Kolmogorov-Smirnov statistic
kssim[i]<-max(abs(surv$surv-exp(-surv$time*sum(cens)/sum(time))))}

#conditional p-value
mean(kssim>=obsks)

```

APPENDIX III

Suppose X_1, \dots, X_n are observed time to events or time to censoring with censoring indicators Y_1, \dots, Y_n . Censoring times are independent of event times and event times are exponential with cumulative distribution function $1 - e^{-e^r t}$. The likelihood function is

$$\prod_{\substack{i \text{ such that} \\ Y_i=1}} e^r e^{-e^r \times X_i} \times \prod_{\substack{i \text{ such that} \\ Y_i=0}} e^{-e^r \times X_i}$$

and the log-likelihood is

$$\{\# \text{ of } Y_i \text{ such that } Y_i = 1\} \times r - e^r \sum_i X_i$$

The derivative of the log-likelihood function with respect to r is

$$\{\# \text{ of } Y_i \text{ such that } Y_i = 1\} - e^r \sum_i X_i$$

This derivative will be 0 when $r = \hat{r} = \log \left\{ \frac{\# \text{ of } Y_i \text{ such that } Y_i = 1}{\sum_i X_i} \right\}$. This is the maximum

likelihood estimate of r since the second derivative evaluated is $-e^r \sum_i X_i$ and this is

guaranteed to be negative. Since the maximum likelihood estimator is asymptotically efficient and normally distributed, we can estimate its variance using the inverse of the

estimated Fisher information, i.e. $\frac{e^{-\hat{r}}}{\sum_i X_i} = \frac{1}{\# \text{ of } Y_i \text{ such that } Y_i = 1}$ and can make a confidence

interval for the difference between the parameters in the two groups followed by exponentiating the limits to obtain a 95% confidence interval for the risk ratio. Let $d_E = 51, E_E = 3160, d_W = 37, E_W = 3186$ be the number of events and number of patient years of exposure in the exanta and warfarin arms in the SPORTIF V trial. A formula for the 95% confidence interval for the risk ratio is $\left(\frac{d_E E_W}{E_E d_W} e^{-1.96 \sqrt{d_E^{-1} + d_W^{-1}}}, \frac{d_E E_W}{E_E d_W} e^{1.96 \sqrt{d_E^{-1} + d_W^{-1}}} \right)$.