any IV antifibrinolytic agent during CABG surgery or because they received more than one antifibrinolytic agent during CABG surgery so that drug exposure was no longer clearly identifiable or because they received tranexamic acid.

Of these 78,000 patients, 33,000 received aprotinin, 45,000 aminocaproic acid. These patients were the basis of the primary analysis.

[Slide.]

Unadjusted analyses and analyses adjusted for 41 presurgery patient in-hospital characteristics showed an increased risk for renal failure requiring dialysis and in-hospital death.

[Slide.]

Aprotinin shows a 60 percent increase in the risk of renal failure and 64 percent increase in the risk of death.

[Slide.]

The effect increased slightly when the analysis was limited to the first seven days after surgery to study the immediate outcomes of aprotinin use during CABG surgery.

[Slide.]

We separated aprotinin exposures into very very low, low and high cumulative doses and compared them to low dose aminocaproic acid the most commonly given amount. This does not represent a formal dose-response analysis since dose depends on body weight and duration of surgery, two quantities that could not be determined from these data.

In the very low-dose group, by our definition, only one vial was charged but it is not clear whether that vial was actually used. The data were analyzed in many ways to explore alternative explanations. No matter what analysis, the increased risk of renal failure and death persisted.

[Slide.]

Limiting the analysis to the first seven days slightly increased the associations.

[Slide.]

Adjusting standard errors for clustering of patients within hospital slightly widened the

confidence intervals.

[Slide.]

Further adjusting for the performing surgeon, and, therefore, the treatment variation associated with surgeons, in a conditional logistic regression analysis increased the associations.

[Slide.]

The logistic-regression analyses that were used showed very good model fit as measured by the C-statistic. In fact, the C-statistics of the study, 0.79 and 0.83 respectively, were as good or better compared with widely used risk scores for CABG surgery which are considered gold standards for risk prediction in CABG surgery.

These numbers are directly comparable because the present study makes only claims about its 78,000 subjects.

[Slide.]

This very good model prediction was achieved by including a large number of presurgery patient hospital baseline characteristics into the regression analysis. The adjusted covariates

included the social demographic factors like smoking, markers of severity in risk prognosis including renal, CABG surgery, the number of grafted vessels, et cetera, patient comorbidities and the treatment derived from discharge diagnosis and medications or procedures used prior to surgery including the history of chronic kidney disease, hospital characteristics including CABG volume as well as surgeon CABG volume.

Because of the very large sample size, it is expected that very small imbalances of patient characteristics between treatment groups; for example, a 1 percentage point difference may result in statistically significant p-value, although this will have little impact on confounding.

[Slide.]

The study further identified a data-dense cohort that consisted of patients who had at least two hospital days before the CABG surgery for better confounder assessment who were treated by high-volume surgeons--that is, 50 or more surgeries--and were treated by adequate doses.

That means patients receiving very low doses were excluded from this analysis. This resulted in 13,000 patients or 17 percent of the primary study population.

[Slide.]

The data-dense cohort was analyzed after propensity-score matching to adjust for measured covariates. Propensity scores can be seen as a technique to condense a large number of covariates, and, in our case, there were 41 prespecified covariates, into a single score.

If the distribution of the propensity score are perfectly overlapping, then treatment choice is uncorrelated with any of the measured predictors and therefore free of confounding by those factors.

In this study, before matching, the propensity-score distributions were largely overlapping but not perfectly overlapping, as you can see in this figure. Such a pattern is to be expected for medications with similar indications averaged across many institutions.

[Slide.]

After matching each aprotinin recipient to an aminocaproic acid with a similar propensity-score value, using the Greedy matching algorithm, the propensity score distributions were almost perfectly overlapping. The resulting relative-risk estimates were 1.39 for renal failure and 1.32 for death.

Both estimates are now reduced by comparison to the adjusted primary analysis but still statistically significantly increased. This may be partially due to better control for confounding or to more selective nature of the subgroup presented here.

[Slide.]

In order to address unmeasured confounding, we wanted to substitute the actual study exposure with treatment preference that is unrelated to patient characteristics. This way, and if certain assumptions hold, your observed associations will be unlinked from possible measured and unmeasured confoundings.

We will all agree that patients do not select the surgeons by whether they use aprotinin or aminocaproic acid. At the same time, study data show that a number of high-volume surgeons exclusively use aprotinin for all their patients and other high-volume surgeons exclusively use aminocaproic acid.

Now, for patients treated by such surgeons, disease severity should play a minor role in treatment choice whether such characteristics were measured or not because these surgeons have decided, a priori, already, which of the two medications to use. In some cases, this might be due to restrictive hospital formularies.

It is anticipated that there will still be differences between surgeons. Therefore, we adjusted in a two-stage regression model for the measured 41 covariates in this analysis. This type of analysis is called instrumental variable analysis and has been shown to produce less biased results in situations of strong confounding by unmeasured covariates.

[Slide.]

This instrumental variable estimation, again, resulted in positive associations between aprotinin use and renal failure requiring dialysis and in-hospital death.

[Slide.]

The association persisted but was slightly weaker when the instrument definition was relaxed.

[Slide.]

Lastly, sensitivity analyses were performed to estimate how strong a confounded would have go be in order to fully explain the study's findings.

[Slide.]

The conclusion of that analysis was that a single unmeasured confounder has to be five times more likely to be present among aprotinin users and also be seven times more likely to cause renal failure to fully explain the observed associations. However, individual confounding factors can add up to a larger net confounding or they can cancel each other out.

[Slide.]

Such a confounding factor could be incomplete adjustment for pre-existing renal failure. In the amended analysis, such patients were excluded from the analysis.

[Slide.]

The primary study results did not change meaningfully. However, the matched propensity-score analysis deserves more discussion.

[Slide.]

After exclusion of patients with pre-existing dialysis, the propensity-score-matched data-dense population shows no overall increased risk of renal failure but a 20 percent risk increased for the analysis limited to the first seven days after CABG surgery. This is consistent with residual confounding for analyses of renal failure late in hospitalization.

Therefore, the analysis limited to the first seven days after surgery appears more robust and, as you have seen, such seven-day analysis resulted in slightly stronger associations than the

prespecified primary analysis.

Members of the advisory committee, Mr. Chairman, based on these data that you have available in much greater detail, the following conclusions can be drawn.

The is the largest cohort study on the safety of aprotinin today using both simple and complex variance of design and analysis to address existing limitations of administrative in-hospital data.

It needs to be acknowledged that, to answer this research question, nonrandomized studies are likely confounded by patient predictors of unintended outcomes. This study found an association between aprotinin use and renal failure requiring dialysis as well as in-hospital mortality compared with aminocaproic acid that persisted through multiple analytic approaches.

It is possible that these associations are fully explained by confounding, but this is not probable. The September 28th report on the trial abstraction will address this issue quantitatively.

A secondary analysis excluding pre-existing renal failure suggests that an analysis of short-term outcomes is more valid than the prespecified primary analysis of longer-term outcomes. As you have seen, such seven-day analysis resulted in slightly stronger associations than the prespecified primary analysis.

Thank you very much. I would like to hand over to Dr. Pam Cyrus.

Trasylol (aprotinin injection)

Review of Clinical Data with Focus on Specific Safety Events

DR. CYRUS: Good morning.

[Slide.]

Thank you for the opportunity to present Bayer's clinical-trial data and put it in perspective today of the observational studies that we have seen.

[Slide.]

What I would like to do for you today is review the revised indication, then the efficacy and then focus on safety with emphasis on

mortality, stroke and myocardial infarction, renal failure and, finally, hypersensitivity.

[Slide.]

As Dr. Shashaty pointed out this morning, in December of 2006, the label for Trasylol was revised. One of these revisions included adding patients who were at increased risk for blood loss and blood transfusion.

[Slide.]

Also reviewed for you this morning, and I won't spend time with it; Dr. Corso did an excellent job talking about the risk factors of CABG surgery and particularly the risk of blood transfusion. This has become so significant that the STS, the Society for Thoracic Surgery, has developed blood-management guidelines for cardiac surgery.

[Slide.]

Last year, in anticipation of an advisory committee, we did review transfusion-associated mortality. I won't spend time with this now but we do have Dr. Shander who was part of that consensus

panel in our group with us if you would have additional questions.

[Slide.]

One thing that we have heard is that aprotinin reduces transfusion rate. When using the U.S. clinical trials that were the basis for approval of this product, you can see, for both primary and repeat CABG, that both the full and the half dose significantly reduced the percent of patients receiving red blood cells and platelets.

In addition to decreasing the percent of patients that are transfused, the mean units of red blood cells, platelets, fresh frozen plasma and cryoprecipitate are also reduced. This data is consistent with the data that was collected by Bayer in Europe as well as other studies in the literature.

[Slide.]

When looking at reexploration, Dr. Smith will elaborate on this in his presentation as well as you heard Dr. Corso mention it this morning. Reexploration for bleeding is a serious

complication for the CABG patient. Aprotinin has been demonstrated to limit reexploration.

If you look at the first line here, this is a publication by Keubler. This is based on the Bayer clinical trials and you can see that full-dose aprotinin reduces reexploration.

In a meta-analysis conducted by Dr. Brown in 2007 looking at what was available in the literature, he has published that full-does aprotinin limits reexploration. Half-dose aminocaproic acid and tranexamic acid did not share this property.

These serve as the basis, these data serve as the basis, for the STS guidelines that give a Class I recommendation for full-dose aprotinin limiting reexploration. I should make note that the half dose aprotinin aminocaproic acid and tranexamic acid do not carry the same level of evidence.

[Slide.]

As you heard this morning, the cardiothoracic surgeon is faced with more and more

patients being on anti-platelet therapy going into the operating room. Dr. Smith will also elaborate on this in his risk-benefit conclusion.

In the Bayer clinical-trial data, aprotinin did reduce transfusion in patients receiving aspirin at the time surgery. This study is by Dr. Van der Linden and is not sponsored by Bayer but is an independent study, double-blind placebo-controlled study that demonstrated that full-dose aprotinin significantly reduced the percent of patients transfused as well as the mean number of units of red blood cells and platelets transfused.

[Slide.]

When looking at the totality of the evidence across randomized clinical trials, you can summarize that aprotinin is effective in reducing blood loss in transfusion and CABG surgery. Both the full and the half dose are recommended with the Class I recommendation in the 2007 STS Guidelines for blood management and cardiac surgery.

In addition to this, the STS Guidelines

for blood management also give a Class I recommendation for limiting reexploration for bleeding. In addition, the 2005 STS Guidelines for anti-platelet therapy also recommend full-dose aprotinin as being effective in patients who were pre-treated with aspirin or clopidigrel.

[Slide.]

I would now like to focus on the safety of Trasylol in CABG procedures.

[Slide.]

First, looking at in-hospital mortality. I should point that, when you look at the Bayer randomized controlled trials or the Brown meta-analysis of randomized controlled trials or any of the observational studies based on clinical-trial databases, you can see that there is no increase in in-hospital mortality associated with aprotinin.

However, the only study that does show an effect is the i3 drug safety study that we just heard presented by Dr. Schneeweiss. I would like to elaborate a little bit more on Bayer's position

on the limitations of this database and of this result.

[Slide.]

When you look at the STS database, which has been published by Shroyer in 2003, and you look at the mortality model for comparison of the major risk factors in the CABG-only model, you see that there are 28 risk factors. Here we only displayed those that have an odds ratio greater than 1.3 and they are in decreasing order of odds ratio.

You can see that many of these variables are missing from the i3 drug safety database. Only about half of the variables are present.

You can also see the surrogate markers are used. Discharge diagnosis is the basis for history rather than in the clinical trial where one may have the medical history. In doing this, you can see that consistently the odds ratios estimated from the i3 drug safety database tend to be lower, especially for first reoperation which is indicated by re-do cardiac being significantly lower than that identified in the STS database.

I make note of this because patients who are at increased risk for bleeding include those patients who have gone through multiple reoperations for their procedure.

[Slide.]

Another oddity about this database that makes us not want to trust it is the fact that if you look at the mortality risk model from the STS and you look at the risk for prior myocardial infarction, hypertension and prior stroke, the i3 database would lead one to believe that old MI and hypertension were protective against mortality. This is contrary to what we would expect clinically and what has been identified by a well-established database with the STS.

[Slide.]

So the summary of the in-hospital mortality findings are that the randomized controlled trials reported no increased risk of mortality related to aprotinin. The observational studies, with the exception of the i3 study, which is based on administrative database, also have not

shown an association with aprotinin and in-hospital mortality.

The conclusions from the i3 drug safety study are unreliable mainly because of the limitations of an administrative database with many key risk factors not being included and the diagnosis being recorded by discharge diagnoses only.

The odds ratios are also contrary to clinical experience as they pertain to hypertension and prior MI. Bayer's conclusion is that, when you look at the totality of the evidence with a hierarchical approach, the data do not support an increased risk for in-hospital mortality associated with aprotinin.

In regards to long-term mortality, Dr. Makuch will elaborate on that in his presentation but we find that there are also significant statistical flaws to Mangano's 2007 publication and we also believe that this is not a basis for a mortality signal.

[Slide.]

Now to focus on stroke and myocardial infarction. This was reviewed for you in detail at last year's advisory committee meeting including graft patency. I won't go into that today but if you should have questions we would be happy to respond.

When we look at stroke and myocardial infarction, when we look at the Bayer randomized controlled trials as well as the Brown meta-analysis, there is no statistically significant finding for stroke or myocardial infarction. In addition, Dr. Mangano, in his observational study, does report an increased risk of cerebrovascular event and cardiovascular event as a composite endpoint in primary CABG patients but not in complex surgery patients.

I should note that the FDA reanalysis of Dr. Mangano's data does not show a statistically significant finding for MI or stroke. Dr. Karkouti's data did not show an increased risk for myocardial infarction or stroke and a recent publication by Dr. Coleman out of Hartford Hospital

also showed no statistically significant risk for stroke or myocardial infarction.

But what they did find in their database is there was a statistically significant reduction of neurologic outcomes associated with aprotinin.

In the i3 drug safety study in the preliminary report, they determined that they could not determine myocardial infarction from the database and the authors concluded that there wasn't a safety signal for stroke. In the final report, they do not address this.

The FDA has concluded that the stroke findings were likely associated with uncontrolled confounding in this study.

Therefore, when looking at the totality of the evidence in a hierarchical basis, Bayer concludes that the data do not support that there is an increased risk of stroke or myocardial infarction associated with aprotinin.

[Slide.]

Now to focus on renal function.

[Slide.]

I would like to show you a slide that was shown to the committee last year at the advisory committee. This is an analysis of the Bayer randomized controlled trial global database on safety for patients that had a baseline serum creatinine. I should make note that there are 2,249 patients in that full-dose aprotinin group, 2,164 in the placebo. As you can see, the majority of them did have creatinines available for baseline comparison.

When looking at this data, 9 percent of full-dose aprotinin patients versus 6.6 percent of placebo patients had a serum creatinine elevation greater than 0.5 mg/dL over baseline. This was statistically significant as discussed at last year's advisory committee.

This information has now been reflected in the revised label of December, 2006 to reflect this difference.

I would also like to point out, when looking at dialysis, and this was by bleeding CRFs because it wasn't prospectively collected, you could see that

there were no differences between groups.

When looking at the median time to resolution of serum creatinine elevations, it was on an average of nine days for aprotinin and six days for placebo. Further subsequent analysis to try to establish the patients at particularly increased risk, it was determined that patients with preexisting renal impairment or those who had received aminoglycosides were at even further risk of serum-creatinine elevations if they also received aprotinin.

Given this information, this was also reflected in the revised label of December, 2006 that now reads that this risk may be especially increased for patients with preexisting renal impairment or those who receive aminoglycoside antibiotics or drugs that alter renal function.

[Slide.]

When looking at the outcome of renal dysfunction, whether you are looking at the randomized controlled trials or the publications, it is important to note that every author has used

a slightly different definition of renal dysfunction.

I apologize for the business of this slide but I think it will reflect better in your written version. But I have provided for you those various definitions.

One thing that is consistent; regardless of which definition for renal dysfunction is used, the randomized controlled trials have shown a statistically significant increase of renal dysfunction associated with aprotinin and the observational studies are also consistent with that with one notable exception.

Although Dr. Mangano reports an association with the composite renal outcome that is statistically significant for primary CABG and for complex CABG, when the FDA did their analysis, they did not find that renal dysfunction was statistically significant.

[Slide.]

Now looking at renal failure, which also uses various definitions, but most definitions

include the requirement for new-onset dialysis. You could see, looking at the randomized controlled trials, that there is no increased risk of renal failure requiring dialysis.

When we look at the publication by Dr. Mangano as well as confirmed with the FDA analysis, there was found to be a statistically significant increased risk of renal failure requiring dialysis associated with aprotinin. However, that finding was not presented by Dr. Karkoti and just yesterday, published in Circulation, is an article by Dr. Furnary where he looked at this and determined that aprotinin was not an independent risk factor for renal failure but, rather, it was an increased number of transfusions in the high-risk patient population that were receiving this drug just for that reason, that they were high risk.

I think that Dr. Karkouti also elaborated on that this morning in his presentation.

So now let's focus on the i3 drug safety study. As reported in Part A of the report, the

odds ratio for the entire patient population was 1.65. However, when excluding patients who had known dialysis preoperatively and redoing the analysis, and this is reported in Table 7, page 19, of the Addendum dated August 31, 2007, that most unfortunately was not available at the time of the briefing document but was provided to you this morning, you will see that odds ratio, when excluding those patients that did not have preexisting dialysis, is 1.04 and is not statistically significant.

[Slide.]

So, when looking at the outcome of renal failure and renal dialysis and trying to make some conclusions of it, let's look a little closer at i3 study. Again, let's talk about STS database and the odds ratio and the rank order of those odds ratios that are above 1.3.

Of the 28 risk factors, only approximately half of them are present in i3. Once again, you don't necessarily see the same magnitude as one would have expected from these odds ratios from the

STS database. But, once again, you are also using discharge diagnoses rather than medical history to be able to evaluate this.

[Slide.]

Once again, there is an oddity that makes you believe that this database is not reliable. If you are to look at the renal outcome and you are to look at the i3 covariates, you would see that old MI, hypertension and smoking were protective against renal failure in a statistically significant fashion. Once again, this is contrary to what we would expect from our clinical experience.

[Slide.]

So, when looking at the totality of data for renal dysfunction and renal failure findings, the randomized controlled trials as well as the observational studies have demonstrated an increased risk of renal dysfunction associated with aprotinin. However, the randomized controlled trials, as presented by Bayer and as well as the meta-analysis by Brown, have not demonstrated an

increased risk of renal failure.

The results of the observational studies are variable with Dr. Mangano's study with the FDA analysis showing an increased risk of renal failure but not renal dysfunction, and the i3 study having the limitations that we have just discussed.

When looking at this, Bayer concludes, and it is reflected in the December, 2006 label revision, and this is a direct quote, "Trasylol administration increases the risk for renal dysfunction and may increase the need for dialysis in the perioperative period."

[Slide.]

Now to focus on hypersensitivity which we spent quite a bit of time discussing at last year's advisory committee, but I just would like to point to you a few key label revisions that were implemented in December of 2006.

Trasylol is now contraindicated in patients with known or suspected aprotinin exposure during the last 12 months. It is also emphasized that aprotinin may be a component of some fibrin

sealant products and that fatal reactions have been seen with the initial test dose as well as other components and fatal reactions have been seen when the initial or the test dose was well tolerated.

Trasylol is also recommended to be administered only in the perioperative setting where cardiopulmonary bypass can be rapidly initiated. Bayer has also implemented a risk-minimization plan including prescriber education and we continue to work on an aprotinin-specific IgG assay available as a point-of-care device.

That development is ongoing and dialogue with the FDA has been initiated.

[Slide.]

So in summary, aprotinin provides an important clinical benefit for the CABG patient. As we have heard this morning from Dr. Corso and as reflected in the STS guidelines, it is vital to have a multi-modality approach for blood management and Trasylol is an important component of that armamentarium for blood management.

The recent label changes of December, 2006 reflect updated safety analyses. Bayer continues its effort for risk minimization including the development of an IgG assay in order to reduce the risk of hypersensitivity.

In totality, Bayer remains convinced that the benefits of aprotinin outweigh the risk when used in accordance with labeling.

I would now like to introduce Dr. Bob Makuch who will give a statistical overview for us.

DR. HARRINGTON: Just as a reminder, the sponsor has about 20 minutes left.

Aprotinin Studies: Weight of Evidence

DR. MAKUCH: Good morning.

[Slide.]

So I will speak very quickly.

[Slide.]

This slide gives an overview of the generally accepted weight of evidence associated with various types of study designs for efficacy and safety. RCTs carry the greatest weight since randomization ensures that both known and unknown

covariates are balanced on average. Confounding and other biases are avoided through proper randomization.

Finally, the studies are planned to address well-defined hypotheses leading to straight-forward statistical analysis.

Observational studies usually are preplanned to address the research question but are subject to well-known biases such as confounding and channeling. Thus, one is less certain whether any treatment differences can be attributed to treatment as opposed to other factors.

Administrative databases are not pre-planned to address a specific scientific hypothesis often leading to a variety of complex analyses. They also are subject to all of the problems of observational studies as well.

I return to the schematic at the end of my talk to summarize the studies results.

[Slide.]

The 2006 Dr. Karkouti study was discussed at the previous ad com. It is a single-center

study in which major imbalances in baseline risk factors were properly addressed through propensity-score matching. Essentially, propensity-score matching is an approach in which one attempts, for an observational study, to mimic an RCT and to eliminate confounding. This is done by selecting factors at the decision-making level and then matching on these factors to select similar subjects for each treatment group being compared.

Creatinine elevations were consistent with the RCT data supported by FDA reanalysis. Finally, no significant risks of cardiovascular or cerebrovascular or short-term mortality were identified, again similar to the RCT findings.

[Slide.]

A second 2006 study was done by Dr. Mangano with 69 centers in roughly 47,000 patients. Unlike Dr. Karkouti, he used the fundamentally different regression-modeling approach to address issues of confounding and significant baseline imbalances between treatment.

Regression models, unless coupled with the correct application of methods to achieve balance cannot be expected to give reliable results. Mangano performed supplemental analyses using a propensity score but it was done improperly in which a single propensity score was derived for all treatments.

Because no diagnostic displays or analyses were shown to support claimed covariate balance, other baseline factors confounded with treatment represent viable alternative candidates for causality attribution.

Finally, the FDA re-analysis is shown in the final bullet and the results are provided in your slide.

[Slide.]

In 2007, Dr. Mangano published long-term mortality data in a subset of 62 of the 69 centers and a much smaller number of roughly 3800 patients. The same limitations apply as in his previous publication. In addition, significant confounding exists between treatment and geographic regions

with no patients receiving aprotinin in Asia or the Middle East.

RCTs are explicitly designed to prevent all patients getting only one treatment in the a center precisely because this confounding cannot be reliably corrected for. Thus, any bi-region comparison must be carefully constructed to eliminate this issue of confounding.

Thus, one cannot ascertain whether any adverse outcomes are due to aprotinin or to country differences in standard of care in different patient populations. In addition, there was marked differential lost-to-follow-up between treatments. The FDA concluded there were no significant differences in mortality at six weeks, six months, one year and two years. Mortality differences appear significant or nearly so at Years 3 through 5.

[Slide.]

To evaluate the study, two general areas were considered. The first is the database, itself, with some points mentioned in the slide,

and the second area is design and analysis which I will now discuss briefly.

[Slide.]

As you have seen, i3 used the premier prospective database. Characteristics of this claims database are summarized on the slide and I will give you a moment to look at some of the points in the slide without discussing it for purposes of saving time.

I think that bottom line, though, that I do wish to emphasize is that, despite its size, because it is a claims database, there are numerous deficiencies present, many not presently seen in prospectively designed studies.

[Slide.]

Because it is a claims database not designed for this research effort, there are important covariates unmeasured or misclassified. Many have higher odds ratios for risk of the adverse outcome than the risk of aprotinin and any of these factors provide valid alternatives to treatment as causal factors.

In addition, a Bayer analysis showed that the important predictor, re-do surgery, was misclassified as primary surgery in over 50 percent of the patients seen in the same hospital. No regression model, however sophisticated, can rescue a database with these issues.

[Slide.]

Originally, six adverse outcomes were proposed for use by i3. But this evolved over time in which MI was dropped even before the September, 2006 preliminary report. MI was not reported because i3 could not be sure whether MI was a treatment-emergent outcome or not.

Second, the August report dropped all outcomes except acute renal failure and death.

[Slide.]

Except for death, all outcomes were defined by surrogates. Renal failure requiring dialysis was defined by charge codes for hemodialysis, peritoneal dialysis or hemofiltration.

For renal-failure dialysis outcome, the

last bullet, patients with known dialysis prior to surgery were originally not excluded. But this primary cohort did include more patients with recorded pre-op dialysis patients treated with aprotinin than aminocaproic acid. This issue was first addressed in the i3 addendum of August 31st.

[Slide.]

i3 addressed the significant issues of confounding by treatment and baseline imbalances using the regression models. As noted previously, though, regression models cannot be expected to give reliable results unless coupled with correct application of methods to achieve balance.

i3 considered instrumental variable and propensity-score methods to augment the capabilities of their model but neither were appropriately applied to address the limitations of their regression model.

[Slide.]

The issue of confounding is what makes use of observational and claims databases so challenging. Randomization in RCT virtually

eliminates this issue and propensity scores attempt to mimic the RTC. Regression modeling, though, as a methodologic approach to address confounding is very complex.

In a recent paper by DeLong involving a claims database analysis, she concluded that, the second bullet, treatment selection criteria vary across centers which confounds the treatment outcome relationship. So geographic lumping may not be enough to address this second layer of confounding above and beyond the first in the first bullet of confounding with treatment.

The Karkouti study is unaffected by this issue since it was a single-center study.

Proper treatment of these issues, as pointed out in this paper, requires more complex regression models than used by i3. In fact, DeLong showed that standard logistic models overestimated the true treatment effect by roughly 35 to 40 percent compared to complex models that appropriately took into account the full layers of confounding.

[Slide.]

i3 also performed a sensitivity analysis to explore how strong an unmeasured confounder would have to be to explain the findings and concluded there was no plausible candidate. However, their analysis is flawed by the inherent assumption that the analysis characterizes association of a single unmeasured confounder.

There are multiple baseline important risk factors, as we have seen, whose combined odds ratios far exceed the i3 single covariate value implying the set of unmeasured covariates are more than sufficient to negate the i3 findings.

[Slide.]

This slide suggests that aminocaproic acid has a dose-response effect on in-hospital death in the i3 study with a low-dose odds ratio of 0.83 and high-dose odds ratio of 1.35. A more likely explanation, however, is that confounding effects have not been removed fully by the regression model.

[Slide.]

i3, in their August 31st addendum, performed an analysis of renal failure and showed that the association between aprotinin and renal failure was reduced from 1.39 to 1.04, no longer statistically significant when properly excluding patients with pre-existing dialysis.

[Slide.]

Dr. Schneeweiss also showed the C-statistic, another technique to augment the credibility of the regression model. But the recorded C-statistic overestimates the true predictive ability of the model since it was not validated on an independent data set.

Comparison with the STS C-statistic is inappropriate since STS used an independent dataset for proper validation of predictive ability. Biased, inflated C-scores in administrative data are well known.

[Slide.]

Finally, a medical-record review is underway to address data-accuracy concerns in the claims database, but there are sampling issues that

require attention if this chart review is to be valid and useful.

Because time is short and you have the slide, I direct you to the last bullet in which I conclude that no valid inference can be drawn from this review essentially because the methods for doing this are not described and it is done in only two hospitals that are highly selected.

[Slide.]

In summary, here is our schematic with populated data for perioperative mortality. Except for the i3 study which is associated with the least weight of evidence, there is no evidence of a statistically significant association with perioperative mortality.

Also, when i3 did a propensity-matched analysis, the risk was reduced 20 percent and it is now just statistically significant.

[Slide.]

For renal dysfunction, there is consistent evidence to support the hypothesis that aprotinin is associated with a statistically significant

increase in renal dysfunction. This conclusion is consistent with that presented at the 2006 advisory committee.

[Slide.]

For renal failure, the i3 study gives mixed results. The regression model showed a statistically significant association with renal failure of 1.65. No association, however, was found with an odds ratio of 1.04 when patients with dialysis before surgery were excluded in a propensity-matched subgroup.

For the Mangano study, an FDA analysis is ongoing. The remainder of the studies all associated with the greater weight of evidence showed no statistically significant association between aprotinin and renal failure.

Thank you for your attention. I now introduce Dr. Peter Smith from Duke University.

Trasylol (aprotinin injection): Risks and Benefits from a Surgeon's Perspective

DR. SMITH: Thank you.

[Slide.]

I am here to provide a surgeon's perspective on the importance of this drug and its safety and efficacy. It is allowing us to adapt to a changing world that I will try to describe to you. I will remind you that this is the only drug that is specifically indicated for use in heart surgery to minimize transfusions.

That is not really in question here, with abundant randomized controlled-trial evidence that transfusion is reduced.

[Slide.]

The STS database shows the change in our patient population from comorbid features going up. These are 1.6 million records out of the 3.5 million record database. It was really created so especially it wouldn't be evaluated by administrative databases, rather by prospectively obtained clinical databases.

This is changing patient characteristics here, in general, but there are lots of characteristics that are being measured and they get coalesced into predictive algorithms that allow

us to properly risk-adjust these patients. You can see that the risk-adjustment prediction of mortality and all these other significant complications has been increasing the real measure of how these risk factors come together.

[Slide.]

We have some special risk factors. These are the last 300,000 patients, 2005 and 2006. Lots of our patients have had angioplasty urgency, almost half of them myocardial infarctions recently. They are all getting aspirin. This is a performance-measure indicator right now. Obviously, these anticoagulants have been discussed and platelet inhibitors are very common.

[Slide.]

This is why they are really good for our patients, obviously, overall and that is why they are important. They are not great for surgeons.

Here are the unstable angina ACCHA Guidelines that just came out last month and you can see six agents here that interfere with hemostasis are recommended initially in these

patients.

[Slide.]

Then when they are eventually treated, either by medical therapy, Bayer stents, drug-eluting stents.

They all get aspirin, clopidogrel for a month, clopidogrel for at least a year if not a lifetime for patients with drug-eluting stents. There are 5 million of these estimated worldwide, 2.5 million in the United States alone. A lot of them are going to come to cardiac surgery and clopidogrel is being used in a myriad of other patients as well that we see.

[Slide.]

Clopidogrel causes increased bleeding, as was already mentioned this morning. Here is no 5-day-delay patients from a registry, significant increase in bleeding. Obviously, not all our patients can tolerate a delay. But, in this randomized trial when no delay could be done but aprotinin was used, there was a significant reduction in bleeding afforded by aprotinin.

Aprotinin is also effective here in patients not on aspirin but also in patients on aspirin. Here is the tranexamic acid data showing it is effective here but when aspirin is added, tranexamic is not effective. There are no data regarding aminocaproic acid in aspirin use.

[Slide.]

So aprotinin is effective in our patients getting aspirin or clopidogrel. It has important antifibrinolytic effects as we have mentioned, but it is important to realize its anti-inflammatory and it is effective by preserving platelet from being injured during cardiopulmonary bypass and from drugs. And that is why it is an adaptive thing. It is being effective in new agents as they are being developed.

The another antifibrinolytic drugs don't have these proven properties.

[Slide.]

The risk of re-exploration for bleeding is obviously a very serious complication. Here is data from 1.3 million patients. This is the

predictive algorithm for risk factor for re-exploration for bleeding. It is increased 25 percent when normalized back to 1995.

[Slide.]

Obviously, this is a serious event. This is comparing 17,000 patients re-explored, 660,000 who were not. You can see they have a higher incidence of the use of these anti-platelet agents when they get re-explored and they have large increases in the use of blood and blood products.

They have a double ventilation time, almost two days on the ventilator, four days in the ICU. Here is the bottom line, an about a five-fold increase in the risk of mortality when re-exploration occurs.

[Slide.]

Full-dose aprotinin significantly reduces this. It is obviously a morbid and mortal event and the other fibrinolytic agents don't have any indication there.

[Slide.]

Now the safety of the drug has been called

in question, obviously, from these observational data despite the abundance of randomized-trial data. So I will make a few comments as a surgeon on this

The surgeon's decision to use aprotinin is complex and it is not corrected for these observational trials. Aprotinin is characteristically used in high-risk patients and it is because they are high risk. These observational trials are not blinded, we can't forget, and the treatment selection is known and influences the patient-management decisions and the measured outcomes.

Finally, the cardiac surgeon is an important influence on both the outcome and the selection. Now, in a randomized controlled trial, each surgeon contributes to all the study arms by randomization and that eliminates that effect.

[Slide.]

What about some of these observational trials, like here is the EPI 2 database on which Dr. Mangano bases his conclusions. Half the

no-treatment group patients came from Europe. 70 percent of the aprotinin patients came from Europe and none of the aminocaproic acid patients came from Europe because it is not used there.

96 percent of the Amicar patients came from the United States. 25 percent of the TA patients here came from probably mostly Canada. So, when you look at this, it is virtually impossible for a single surgeon or center to have contributed patients to all four of these arms and it puts a big imbalance confounder into this that probably can't be corrected for.

[Slide.]

The safety also has been called into question for mortality risk. Obviously, the randomized controlled data show no signal for mortality. The Mangano, in primary surgery, had a signal but this has been taken away with the FDA analysis we will be hearing and Dr. Karkouti and Coleman saw no signal.

We obviously have the signal from the i3 database that we have discussed and it comes back

toward neutral with propensity matching, but we disagree with the validity of this database.

[Slide.]

Renal dysfunction, obviously, is not news. This has been known from the randomized controlled trials for a long time and is on the label. Dr. Mangano didn't show renal dysfunction. Dr. Karkouti did even with his latest correction. Dr. Coleman saw the renal dysfunction agreeing with the randomized controlled trials.

I would like to point out, though, that these data are based entirely on in-hospital serum creatinine rises and they don't take into consideration anything that happens to these patients after they are discharged from the hospital.

[Slide.]

Here are 216 patients who were randomized between aprotinin and placebo, aprotinin full dose, in the randomized Bayer trials where creatinine was measured specifically. Preoperatively, the patients, of course, since they are randomized,

start the same.

There is a drop during surgery that is common. Over the seven days of the hospitalizations, you see this typical rise in the serum creatine on average in the aprotinin patients compared to the placebo. This achieved statistical significance at seven days.

However, at follow up, this is 35 days later on average, the serum creatinine is exactly the same in the two groups. So this is a transient problem.

I would also mention that urine output was measured in the first 48 hours in these patients and it was exactly the same. So this is a non-oliguric type of renal-dysfunction pattern that we are seeing due to this drug.

[Slide.]

Renal failure, of course, is another matter. When dialysis, which is usually part of measure, occurs, that is a significant complication. That wasn't seen in the randomized controlled trials.

We see it in Dr. Mangano's data here, barely statistically significant. Dr. Karkouti, he saw a trend but he has eliminated that and even more so in the data that he presented today.

Then, of course, with i3, when you eliminate those who were on dialysis before surgery, you see that there is no dialysis or renal-failure figure. SO we are left with Dr. Mangano's assertion that there is increased renal dialysis.

Now, how can that be possible if there is not increased renal dysfunction.

[Slide.]

So some clue might be found here from looking at this locked database that he is using, Epi 2, which was reported on by Dr. Ott for other purposes in another publication where she was looking at the country distribution. Dr. Mangano presented these data earlier.

You can see that renal dysfunction here and renal failure have different incidences. For renal failure, it is much less frequent than in

renal dysfunction, less than half. Same in the U.S. Same in the U.K. But, in Germany, for unknown reasons, there is more renal failure than there is renal dysfunction.

This calls into question the veracity of this outcome variable because many patients in this group must have had dialysis when they failed to meet criteria for having renal dysfunction.

Recall that 70 percent of the aprotinin data in Dr. Mangano's dataset comes from Germany. This may be driving that renal-failure outcome variable by Dr. Mangano.

[Slide.]

So, in conclusion, aprotinin has a favorable risk-benefit profile when used in accordance with the label. It reduces blood loss, transfusion, re-exploration and it may reduce stroke. It is effective in patients with complex and evolving anti-platelet therapies that we have to deal with every day.

It is associated with renal dysfunction but not renal failure and that appears to be

temporary and resolves in a few weeks. It is an essential therapeutic option for CABG-surgery patients at increased risk for bleeding and transfusion.

Thank you.

DR. HARRINGTON: Thank you. I want to thank the sponsor for maintaining the time constraints.

Next, we are going to hear for the next hour from the FDA presenters and then we will have about 15 minutes for questions before lunch, and plenty of time after lunch.

FDA PRESENTATION

Aprotinin: Observational Studies

DR. OUELLET-HELLSTROM: Good morning.

[Slide.]

My name is Rita Ouellet-Hellstrom. I am an epidemiologist in the Office of Surveillance and Epidemiology Division of Drug Risk Evaluation.

[Slide.]

This morning, we have heard a summary of the aprotinin's approval history and the

observational studies. The objective of my presentation is to briefly compare and contrast study designs and biases, summarize the results and show the consistency as well as the inconsistency of the study results and identify the questions that can be addressed by an FDA analysis of the data and those that cannot.

[Slide.]

Clinical trials considered to be the gold standard are usually designed and powered to evaluate efficacy whereas observational studies are powered to assess safety. Study subjects in the efficacy trials may be younger, have fewer comorbidities. They are randomly assigned to treatment groups whereas study subjects in observational studies represent clinical practice and are selected as treated.

Because of the random assignment in clinical trials, it is expected that all known and unknown population characteristics are randomly distributed across treatment groups and assumed balanced. Observational studies, on the other

hand, are more prone to biases and confounding that need to be controlled.

Because data are collected prospectively and require direct patient contact, clinical trials are usually of short duration and more expensive.

[Slide.]

The three observational studies under consideration have different designs or different procedural complexities, population sizes, comparator and control groups and exposure and outcome definitions.

[Slide.]

The studies also differed in the proportion of non-primary surgeries performed such as emergency re-do or CABG and valve surgeries. These complex surgeries range from a high of 72 percent in the Karkouti study to a low of 31 percent in the Mangano study.

[Slide.]

The three studies differed in the numbers of study subjects evaluated. The Mangano 2006 study, a multicenter international study, reported

on over 4,000 study subjects. The Mangano five-year mortality study included patients from 62 of the 69 original centers. In general, 90 percent of the patients were included in the long-term mortality study but only 62 percent of those receiving texamination acid.

The Karkouti study was smaller and the i3 Premier study provided information on over 30,000 patients in each treatment group. The large size is a major strand of this study.

[Slide.]

Label indications differ for the three products in the United States. Differences have already been presented. There is very little use for tranexamic acid in the U.S. for cardiac surgery. Use of aminocaproic acid has been steadily decreasing and aprotinin's use increasing through the end of 2006, and we see a reversal.

[Slide.]

The studies also differed in their exposure and outcome definitions.

[Slide.]

Mangano and associates compared use of any antifibrinolytic with no treatment. Analytical details will be discussed later by Dr. Levenson. Both the Karkouti and the i3 Premier studies compared use of aprotinin with use of at least one other antifibrinolytic. Only the Mangano study compared treatment patients with those not treated.

The preliminary i3 Premier analysis excluded information on over 60,000 patients who received no therapy. Data, however, has been made available to the agency and will be discussed by Chris Holland in his presentation.

All three studies evaluated cardiovascular and cerebrovascular and renal outcomes as well as in hospital deaths, but they differed in how these safety outcomes were defined. The observation period was the length of the hospital stay. Study definitions follow.

[Slide.]

Both the Mangano and Karkouti studies included new Q-waves in the cardiovascular definition. Otherwise, their definitions differed.

Mangano and associates used a competent definition that included myocardial infarction and heart failure. Karkouti and others only considered myocardial infarction as defined here. i3 Premier study based their definition on utilization codes separately for acute coronary revascularization and for heart failure.

Consequently, the results from the i3 study would be expected to vary significantly from the other two studies.

[Slide.]

Definitions for cerebrovascular events were also different. The Mangano study used a competent definition but also presented the results separately. Karkouti and associates defined stroke as any post-operative neurological deficit and the i3 Premier study identified stroke based on utilization codes beginning on the day of surgery.

Based on these definitions, results would be expected to differ across studies.

[Slide.]

All three studies included new dialysis as

the definition for acute renal failure. Two of the three studies also considered creatinine measures. Mangano and Karkouti provided varying clinical endpoints in their definition of renal dysfunction. Of note, however, Karkouti and associates included a new requirement for dialysis and the definition of both renal dysfunction and acute renal failure.

Codes for hemo and peritoneal dialysis or hemofiltration are not specific for renal events and are likely to include patients treated for excess fluid. Overall, renal definitions across studies were more similar and would be expected to capture some of the same renal events.

[Slide.]

Death was defined as any in-hospital deaths for all three studies. Mangano's long-term mortality study considered deaths from any cause over five years. There is no question about the validity of death when observed and there is no variability in its definition.

However, the i3 interim report states that some patients were transferred soon after surgery

to other medical facilities and observation appears to have stopped at the time of transfer.

In Dr. Mangano's long-term mortality study, deaths were identified by patient contact during follow-up period and by querying death registries. The quality and completeness of follow up and death-registry queries may vary by state and by country.

[Slide.]

The overall proportion of patients that could not be found in Mangano's long-term study were somewhat high and the proportion of patients lost-to-follow-up differed across treatment groups. If any of the lost patients were deceased, the mortality results could be very different. Differences in the lost rate were seen as early as six months of follow up.

Information on completed comorbidities over the five years following surgery is not provided in the published paper although that information might be available since long-term follow up was done through patient interviews.

[Slide.]

Because of the way study subjects are selected in observational studies, there is a need to identify and control for confounding and residual bias. Concerns about the validity of these studies to assess safety centered around use of medical codes to identify outcomes, imbalances in the baseline characteristics across treatment groups, geographical, regional and/or provider clustering resulting from variations in regulatory actions across countries, institutional recommendations and medical-provider preferences.

Other biases, less frequently mentioned, include observation time and comprehensive follow up.

These concerns eventually question the validity and usefulness of the observational studies to assess safety, as we have heard this morning.

[Slide.]

In the preliminary report, the i3 Premier study reported on outcome and confounder covariates

based on medically unvetted utilization codes. The sensitivity and specificity of diagnostic and procedural codes varied depending on the seriousness and knowledge of the disease under study and have been shown to be more reliable for myocardial infarction than heart failure or renal dysfunction.

Differences are summarized in this slide.

[Slide.]

Interpretation of discharge summary codes used in the i3 Premier study to identify past medical history is also subject to misinterpretation. It can be readily assumed that codes for diabetes accurately represent pre-existing medical conditions. Codes for other medical conditions such as liver disease, however, assumed in the i3 study to represent a pre-existing condition could also represent a developing disease complication that occurred during the hospital stay.

[Slide.]

Because aprotinin use is indicated for

patients undergoing CABG who are at risk of bleeding, concern has been raised that aprotinin-treated patients require more complex surgery and that observational studies cannot adequately adjust for these differences.

To address these concerns, all investigators have used or mentioned use of multivariate modeling and use of propensity score. Karkouti used a 1:1 matching design. These analytical tools will be discussed by Mark Levenson and Chris Holland in their presentations.

[Slide.]

Other biases of concern in the three observational studies include observational time and patient follow up. The importance of these biases were addressed for death but also applied to other outcomes evaluated. Hospital stays can vary by several days. In most cases, it can be assumed that the longer the stay, the sicker the patient and, therefore, more time to observe the outcome.

Outcome in patients discharged to other facilities soon after surgery reported in the i3

Premier study are missing. Other than under-ascertainment, missing observations are not a major problem when assessing relative risk unless the rate of loss differs across treatment groups.

The Mangano and the i3 Premier papers do not specify the length of observation period and no investigator performed a time-to-event analysis with the short-term in-hospital data.

Karkouti's paper does provide information on the average length and range of stay for each treatment group and these were nearly identical.

[Slide.]

Results are summarized on the following slides.

[Slide.]

For cardiovascular events, Mangano and associates present risk estimates for primary and complex CABG separately. Results show an increase in risk, between 10 and 50 percent, across studies. Subanalyses in the i3 Premier study show a decreasing risk with increasing medical information although patients with longer hospital stays before

CABG surgery may represent a different population altogether.

The elevated point estimates are more likely to reach statistical significance when sample sizes are large. The studies suggest an increase in cardiovascular risk but results, themselves, remain inconclusive, perhaps due to differences in outcome definitions.

[Slide.]

The same observations can be made for cerebrovascular events or stroke. Results also differ across studies, range from no increase to over 100 percent increase. However, postmarketing safety reports indicate an increase in cerebrovascular adverse events associated with aminocaproic acid and tranexamic acid.

A class effect would be masked in a study design that uses a 1:1 matching but could be observed in a study design to compare use versus no-use.

Risk estimates for cardiovascular events vary across studies are less suggestive of an

increased risk for aprotinin use exclusively and results remain inconclusive likely due to differences in definitions.

[Slide.]

In contrast, despite differences in study designs, outcome and exposure definitions, the risk estimates for renal advance are consistently elevated above 1 across studies. Attempts to control residual confounding by considering the more homogenous group undergoing complex surgery results from subanalyses or patients with normal versus abnormal baseline renal failures do not seem to change the risk estimates for renal events. Definitions are also more similar across short-term.

[Slide.]

The risk estimates observed are persistent and consistent across studies and, because of the inability to characterize the exact renal compromise at this time, may actually represent an under-ascertainment of renal safety concerns.

[Slide.]

The increased risk of death in two of the three studies ranged from 40 to 60 percent but were statistically significant only in the i3 Premier study and the Mangano long-term follow-up study.

Information on death events, varied across studies, appear to be somewhat correlated with renal events and is suggestive of an increased risk although results remain inconclusive. Differences may be related to small numbers and missing endpoints in some studies.

[Slide.]

The risk estimates for cardiovascular renal events and death seen in the observational studies are consistent with treatment emergent safety concerns first observed for high-dose aprotinin versus the placebo in the two pivotal efficacy studies and the safety clinical trials in the U.S.

The number of patients observed was small in these trials and results were not statistically significant with the exception of heart failure and renal events in the safety study.

Results of efforts to pool data from national and international clinical trials to increase power should be considered cautiously, however, since pooling also increases the probability of introducing baseline imbalances across treatment groups similar to those seen in observational studies.

[Slide.]

Results from the observational studies suggested an increased risk of adverse events and death with the use of aprotinin but, due to concerns about adequate controlled--of bias and confounding, the differences observed were insufficient for regulatory action.

The Office of Surveillance and Epidemiology recommended confirmation of the study results in an independent analysis. My colleagues Mark Levenson and Chris Holland will present results for the FDA analysis.

I would like to emphasize, however, that the FDA analysis attempts to reproduce study results, standardize analytical approaches, assess

comparability of risk across treatment groups, perform time-to-event analysis and compare aprotinin treatment with no treatment.

[Slide.]

But the FDA analysis cannot redefine outcome and exposure criteria and provide missing information on patients from excluded centers, patients lost-to-follow-up and uncollected data identifying comorbidities and competing risks.

[Slide.]

In summary, the observational studies, despite different designs and outcome definitions, some with large numbers of patients, others with access to medical records and some with direct patient contact all show a consistency of results for renal events and confirmed treatment emergent safety concerns identified in the first clinical trials.

Results with the other outcomes including death are suggestive of an increased risk with aprotinin but remain inconclusive. Reanalysis provides some answers but final results may need a

larger clinical trial sufficiently powered to assess safety and death.

[Slide.]

Thank you.

Now, Dr. Mark Levenson will present the FDA analysis provided by Drs. Mangano and Karkouti.

Statistical Review of the Observational Studies of Aprotinin Safety Part II: the i3 Safety Study

DR. LEVENSON: Hello.

[Slide.]

My name is Mark Levenson. I am a statistician in the Quantitative Safety and Pharmacoepidemiology of the Office of Biostatistics in CDER. Today, I will be speaking on the statistical review of the observational studies of aprotinin safety.

In particular, I will discuss the methods used in the statistical review and the results of the statistical review of the Mangano and Karkouti studies. My colleague, Chris Holland, will follow me with a presentation on the statistical review of

the i3 drug safety study.

[Slide.]

First, I would like to acknowledge the cooperation of Dr. Mangano and Dr. Karkouti in providing data and other materials that enable us to perform this review.

[Slide.]

My presentation consists of five parts. First, I will briefly state the objectives of the statistical review. I will then discuss the statistical methods used in the review. Then I will present the findings of the reviews of the Mangano and Karkouti studies. Finally, I will summarize the findings from the review of the two studies.

[Slide.]

There are two objectives of the statistical review. The first objective is to confirm the reported findings based on the investigators' methods. By this I mean, using the methods employed by the study investigators, can we reproduce the reported numerical results.

The second objective is the more important of the two. The objective is to evaluate the statistical robustness of the findings. FDA analyzed the data from the three studies. The same methods were applied to all three studies to place them on an equal footing. The methods were designed to be robust in that they required minimal assumptions and we made use of effective diagnostics to support the findings of the review.

[Slide.]

Now I will discuss the statistical methods used in the review.

[Slide.]

All three studies in the FDA review made use of propensity scores. In a randomized controlled trial, baseline risk factors are expected to be similar between the treatment groups. In observational studies such as the three studies under review, baseline risk factors may not be similar between the treatment groups.

Propensity scores are used to adjust for differences in baseline risk factors between two

treatment groups. Presently, I will call the two treatment groups Treatment A and Treatment B. The definition of the propensity score is the probability of an assignment for a patient to Treatment A versus Treatment B based on measured risk factors.

The intuition behind propensity scores is as follows. Suppose Treatment A patient and Treatment B patient have the same propensity score. This means they had the same probability receiving Treatment A. One may assume it was just random that one patient received Treatment A and one received Treatment B. Thus, a comparison in outcomes between these two patients is fair or unbiased.

[Slide.]

In practice, there are several issues to consider in the use of propensity scores. First, I introduce the concept of balance. Treatment groups are balanced for a risk factor if the distributions of the risk factor are similar for the two groups.

Propensity score methods cannot account

for unmeasured confounders. This is in contrast to a randomized controlled trial in which unmeasured confounders can be expected to be balanced between the treatment groups.

The propensity scores are unknown values. They must be estimated based on statistical modeling. Diagnostics are important to judge the effectiveness of propensity scores in achieving balanced comparison groups.

[Slide.]

Given propensity scores, there are several methods that could be used to estimate treatment effects. I matching, for example, for each aprotinin patient, you search for a control patient with a similar propensity score. Treatment effects are estimated by comparing outcomes within matched peers. This method was used by the Karkouti investigators.

A similar approach is called stratification. In stratification, patients are divided into strata so that, within a stratum, patients have similar propensity scores. Treatment

effects are estimated by comparing outcomes within each strata. This method was used by FDA in our analysis of the three studies.

The final method discussed is based on multivariate regression. In this method, the propensity score is used as a covariate in the regression estimate of the treatment effect. This method may not be robust to the form of the regression model. This method was used by the Mangano and i3 investigators.

[Slide.]

Now I describe the methods used by FDA in the analysis of the three studies. The methods and analysis plans were prespecified prior to the analysis of the data. Propensity scores with stratification was used to adjust for baseline risk factors.

Medical epidemiological and statistical expertise was used to choose the important risk factors. Diagnostics, both analytical and graphical, were used to evaluate balance and explore findings.

[Slide.]

Now, I present the results from the statistical review of the Mangano study.

[Slide.]

You have already heard summaries of the Mangano study. Here I present some key points relevant to the statistical review.

The analysis plan was prospectively specified. For the in-hospital analysis, the analysis plan was specified after the database was locked. For the long-term mortality analysis, the analysis plan was specified before the unblinding of the mortality data. The specification included the inclusion criteria for the patient, the definitions of the outcomes, the subgroups to be analyzed and the methods used to estimate treatment effects.

The study considered in-hospital outcomes which were the subject of The New England Journal of Medicine paper and long-term mortality follow-up outcomes which were the subject of the JAMA paper. Note that seven of the 69 centers did not

participate in the long-term follow up.

[Slide.]

Multivariate regression, with and without propensity score as a covariate, was used by the investigators to estimate treatment effects. Logistic regression was used for the in-hospital outcomes and Cox proportional hazard regression was used for the long-term mortality outcomes.

This regression was intended to account for the lost-to-follow-up in the post-hospital period. As stated earlier, regression estimation may not be robust to the form of the model.

The propensity score was defined as the probability of receiving any of the three active agents versus not receiving an agent. This propensity score may not result in proper balance between any two treatment groups such as aprotinin and the no-agent groups.

There were no adjustments for geographical differences among the treatment groups. I will discuss this further on the next slide.

[Slide.]

I now present some numerical summaries of the data. First, a summary of the geographical regions. This table gives the percentage of patients by geographical region for each treatment group. There were geographical differences among the four treatment groups. 57 percent of the no-agent patients and 69 percent of the aprotinin patients were in the European region. None of the aminocaproic acid patients were in the European region.

Comparisons between the aminocaproic acid and the no-agent groups that do not adjust for geographical differences would confound differences in patient population and standards of care with treatment effects.

[Slide.]

There were differences in the long-term follow up among the treatment groups. This table divides the patients into three categories based on follow up. The first category, completed five-year follow up or died, represents complete information for patients. 83 percent of the aprotinin patients

had complete information compared to 73 percent of the no-agent group.

The difference between the two groups is explained by the difference in the percentages of patients with no post-hospital follow up. These are the patients in one of the seven centers that did not participate in the long-term follow up.

10 percent of the no-agent patients had no post-hospital follow up compared to 1 percent of the aprotinin patients. The percentages of patients lost-to-follow-up in the post-hospital period were comparable between the two groups, 17 percent for the no-agent patients and 16 percent for the aprotinin patients.

[Slide.]

The mean age of the no-agent patients was 63 versus 65 for aprotinin patients. Both groups were roughly 80 percent male and 4 percent African American or Hispanic.

[Slide.]

The aprotinin patients appeared sicker than the no-agent patients. A higher percentage of

patients of the aprotinin group underwent some surgery, some other procedure, in addition to CABG, 19 percent for aprotinin patients versus 11 percent for the no-agent patients.

However, a higher percentage of the no-agent patients had non-elective surgery, 21 percent of the no-agent patients versus 15 percent for the aprotinin patients. The aprotinin group had higher percentages of history of liver disease, history of renal disease and previous sternotomy.

The two groups had similar percentages for elevated creatinine levels, ejection fractions less than 44 percent and preoperative MI.

[Slide.]

Now I will discuss the review of the findings and methods reported by the study investigators. The primary findings of the New England Journal of Medicine and the JAMA articles based on the investigators' methods were reproduced. This is not an endorsement of the methods but rather a check that the investigators' methods were implemented as stated.

Imbalances in baseline risk factors and geographical regions between the aprotinin no-agent groups after propensity-score adjustment were found. These imbalances may influence the estimated treatment effects.

A lack of overlap in propensity-score distributions between the aprotinin and no-agent group was found. Without good overlap, the estimated treatment effects are sensitive to the form of the regression model.

[Slide.]

Now I will present the results of the FDA analysis of the study.

[Slide.]

The propensity-score adjustment resulted in well-balanced comparison groups. This table shows the percentages of patients with baseline risk factors by treatment group before and after propensity-score adjustment. Note that the differences in percentages between the treatment groups after propensity-score adjustment was smaller for all factors.

For example, for the factors surgical procedure, CABG and other, the percentage went from 11 percent versus 19 percent to 15 percent versus 16 percent. Previous sternotomy was the only one of 28 risk factors in the propensity-score model with significant differences after the propensity-score adjustment.

However, again, the differences were smaller after the adjustment than before.

[Slide.]

This table show the adjusted treatment group estimates for the in-hospital outcomes for the no-agent and the aprotinin groups. It also shows the risk ratios of aprotinin versus no agent and the associated 95 percent confidence intervals. For the renal composite and the renal-failure outcome, aprotinin had a statistically significant effect as seen by the fact that the 95 percent confidence intervals did not contain the value of 1.0.

For none of the other outcomes was there a statistically significant effect. However, the

risk ratio estimates were greater than 1 for the cardiovascular and stroke outcomes.

[Slide.]

This plot shows the percentage of patients with a renal composite outcome for each treatment group by propensity-score strata. The blue squares give the percentages for the aprotinin patients and the green dots give the percentages for the no-agent patients.

The higher number propensity-score strata corresponds to patients more likely to get aprotinin. As was seen in the table on baseline risk factors, these patients appeared sicker. For nine of the ten propensity-score strata, the aprotinin patients had higher rates for the outcome. This demonstrates that the effect of aprotinin was seen across a range of patients.

[Slide.]

This table shows the adjusted treatment-group estimates for the long-term follow-up mortality outcomes. Starting at six months, the aprotinin group had higher estimated

rates than the no-agent group. Starting at three years, the effect was statistically significant or nearly statistically significant.

[Slide.]

Here the adjusted mortality estimates are plotted across time. The blue dashed line gives the estimates for the aprotinin group. The green line gives the estimates for the no-agent group. The separation between the two groups is seen starting at six months.

[Slide.]

I will now present results for the North American subgroup. Note that for this subgroup, all the associated centers participated in the long-term follow-up portion of the study.

[Slide.]

This plot shows the distributions of propensity scores by the European and North American subgroups. The green box plots show the distributions for the no-agent group. The blue box plots show the distributions for the aprotinin group. The no-agent group had similar

distributions for the two regions. However, the difference between the two treatment groups was greater in North America than in Europe.

In the North American subgroup, we were not able to achieve adequate balance between the aprotinin and no-agent groups. The analysis of the North American subgroup used the aminocaproic-acid group as the reference group.

Adequate balance between the aprotinin and the aminocaproic-acid was obtained.

[Slide.]

Here are the in-hospital results for the North American region subgroup. Note that there are only 342 aprotinin patients in this subgroup compared to over 1200 in the full group. The estimates were thus more variable for the subgroup than the full group.

For all three renal outcomes, the renal composite, the renal-failure and the renal-dysfunction outcomes, aprotinin had a statistically significant effect as compared to aminocaproic acid in North America.

For none of the other outcomes was there a statistically significant effect. However, the risk ratio estimates were greater than 1.0 for all outcomes.

[Slide.]

Here are the long-term mortality results for the North American subgroups. For all time points, the aprotinin group had a higher estimated rate compared to the aminocaproic-acid group. Note for the period from six months to four years, the aprotinin effect was statistically significant.

[Slide.]

In summary, renal outcome effects, particularly renal failure, was seen in a range of patients and in the North American region subgroup. Effects for cardiovascular, cerebrovascular and in-hospital death outcomes were not statistically demonstrated. Long-term mortality effects were seen in a range of patients and in the North American region subgroup.

[Slide.]

Now I present the results of the

statistical review of the Karkouti study.

[Slide.]

The study was a retrospective study of five years of patient data from a single center. In contrast to the other studies, the patient population in the Karkouti study consisted for CABG and non-CABG procedures such as valve surgery. All procedures entailed cardiopulmonary bypass.

By the hospital guidelines, aprotinin was used for high-risk patients and tranexamic acid was used for other patients. The investigators used propensity scores with 1:1 matching to estimate treatment effects. 440 of the 586 aprotinin patients were matched.

[Slide.]

The mean age of the tranexamic-acid patients was 63 versus 55 for aprotinin patients. 75 percent of the tranexamic-acid patients were male versus 65 percent for the aprotinin patients.

[Slide.]

The aprotinin patients appeared sicker than the tranexamic-acid patients. A higher

percentage of the aprotinin patients underwent some other procedure in addition to CABG or some procedure other than CABG, 89 percent for the aprotinin patients versus 33 percent for the tranexamic-acid patients.

The aprotinin group had higher percentages of non-elective surgery, previous sternotomy and abnormal creatinine levels. The percentages of ejection fractions less than 40 percent were similar between the two groups. The tranexamic-acid group had a higher percentage of pre-operative MI.

[Slide.]

Now I will discuss the review of the findings and methods reported by the study investigators. The primary findings of the Transfusion article using the investigators' methods were reproduced. Based on the matching approached used by the investigators, the observed risk factors were well balanced.

[Slide.]

Now I will present the results of the FDA

analysis of the study. Because of severe differences in risk factors between the two treatment groups, an analysis subgroup was defined. The subgroup was defined based on the overlap region in an initial set of propensity scores.

The subgroup contained a vast majority of the initial aprotinin patients, 553 out of 586 patients. Baseline risk factors were more similar between the two treatment groups in the subgroup than in the full group.

[Slide.]

The propensity-score adjustment resulted in well-balanced comparison groups. This table shows the percentage of patients with baseline risk factors by treatment group before and after propensity-score adjustment.

Preoperative MI was the only one of 23 risk factors in the propensity-score model with significant differences after propensity-score adjustment.

[Slide.]

This table shows the FDA treatment effect

estimates of aprotinin versus tranexamic acid in the first column. Also shown are the estimates based on the matched paired analysis used by the study investigators in the second column.

The renal-dysfunction outcome was not available for the matched patients and could not be analyzed by the FDA methods. As found by the study investigators, the renal-dysfunction outcome was statistically significant based on the matched paired analysis.

There were no statistically significant effects found for the other outcomes using either analysis methods. However, the risk-ratio estimates from both methods were greater than 1 for the renal failure, myocardial infarction and stroke outcomes.

[Slide.]

This plot shows the percentage of patients with a renal-failure outcome for each treatment group and by propensity-score strata. The blue squares give the percentages for aprotinin patients and the green dots give the percentages for

tranexamic-acid patients.

Here five strata are used as opposed to ten strata used in the analysis of the Mangano study because of the smaller number of aprotinin patients. In four of the five strata, the aprotinin patients had a higher rate of renal-failure outcome.

[Slide.]

In summary, the renal-dysfunction effect was statistically significant. There was some evidence for the renal-failure effect. Effects for the myocardial infarction, stroke and in-hospital death outcomes were not statistically demonstrated.

[Slide.]

Now I will summarize the findings of the statistical review of the Mangano and Karkouti studies.

[Slide.]

The evidence for a renal effect of aprotinin including renal failure was consistent. Effects for cardiovascular, cerebrovascular and in-hospital death outcomes were not statistically

demonstrated. There was evidence for a long-term mortality effect.

Finally, as in any observational study, there was potential for unadjusted confounders between the two treatment groups. This may bias treatment-effect estimates.

My colleague, Chris Holland, will now present the results of the statistical review of the i3 drug safety study.

Thank you.

Statistical Review of the Observation Studies of Aprotinin Safety Part II: The i3 Safety Study

MR. HOLLAND: Good morning.

[Slide.]

My name is Chris Holland. I am a mathematical statistician in the Quantitative Safety and Pharmacoepidemiology Group in the Office of Biostatistics. I am going to discuss the statistical review of the i3 drug safety study.

[Slide.]

This is the outline of my talk and I will

point out that the analysis aprotinin in no-treatment is new. It was not a part of the briefing package since the data on the no-treatment group was just received last month.

[Slide.]

The objectives of this presentation are to examine the statistical robustness of the conclusions from the i3 drug safety study by implementing an alternative methodology and to compare aprotinin patients to patients who receive no I.V. antifibrinolytic with respect to all-cause in-hospital mortality, cardiovascular outcomes, cerebrovascular outcomes and renal outcomes.

[Slide.]

This is the title of the preliminary report. The data used for the analyses in this report were provided to the FDA and used for the FDA analyses that I am about to describe.

[Slide.]

We are going to use this slide to help clarify the different analyses that have been conducted, what the comparison groups were and how

many patients have been involved going back to last year, the preliminary report was released. It included 29,358 patients who received aprotinin and 37,077 patients who received either aminocaproic acid or tranexamic acid during their CABG surgery.

In March of this year, an analysis dataset of the preliminary report data was provided to FDA and this dataset was used to conduct the FDA analysis comparing aprotinin to the other I.V. antifibrinolytics.

In August, FDA received an analysis dataset of the patients who received no I.V. antifibrinolytic agent during their CABG surgery and, from this dataset, 51,588 patients were used to define the no-treatment group used for the comparison of aprotinin to no-treatment.

Also, in August, Part A of the final i3 safety report was released. FDA has not had sufficient time to review this report and has not received any datasets to verify the findings. This report is, therefore, not addressed in this presentation.

[Slide.]

So two datasets have been provided to FDA and used for analysis. The first one contains the 66,000 patients identified for analysis in the preliminary report. The 29,000 patients that represented the aprotinin dataset, the aprotinin group, came from this dataset. And there were also 35,719 aminocaproic-acid patients and 1,358 patients in the tranexamic-acid group. They, together, represent the 37,000 patients used to represent the group of the other I.V. antifibrinolytics.

The second dataset contains 69,176 patients who received no I.V. antifibrinolytic during their CABG surgery. But, of those, 17,588 received I.V. AFs after CABG. So 51,588 patients who received no I.V. antifibrinolytic agent during or after CABG represents the no-treatment group.

[Slide.]

So there are some limitations to the study and hospital claims data to consider. First off, the accuracy of the derivations and the outcomes of

the covariates have not yet been evaluated. A well-implemented chart review would allow for such an evaluation.

There was also the issue of covariates that were not addressed or not available in the analysis data. The covariates listed here are some of the ones that were used in the Mangano and Karkouti studies but not in this one. So adjustments for these factors could therefore not be made.

Regarding the outcome definitions, not all outcomes were explicitly collected in the source data and so, aside from death, they are defined by surrogates. For example, patients are considered to have renal failure if they underwent dialysis but we know that not all actual cases of renal failure will result in dialysis and not all dialyses are to treat renal failure.

The outcomes of heart failure and renal failure could not be evaluated on the day of surgery because of the inability to determine from claims data whether or not events on the day of

surgery occurred before or after surgery.

Death is explicitly captured but only in-hospital deaths are available.

[Slide.]

I will now describe the statistical methods.

[Slide.]

As described by Dr. Levenson, propensity-score methods were used for all FDA analyses. For this study, subgroup and sensitivity analyses were also conducted in order to assess the statistical robustness of the overall results.

The sensitivity analyses included an analysis of outcome rates per patient weeks. This was to adjust for the longer follow-up time among aprotinin patients who stayed in the hospital on average roughly one day longer than patients in the comparison groups.

Since more information pertaining to procedures and medication use becomes available in the database with longer pre-surgery hospital stays, covariate assessments can be improved for

patients with longer pre-surgery hospital stays.

So the analyses of patients who are admitted greater than or equal to one and greater than or equal to three days prior to surgery were therefore conducted to determine whether this improved covariate assessment could affect the results.

Lastly, since the tenth propensity-score decile could be considered to contain the most severe or high-risk patients, an analysis that excludes patients in this decile was conducted in order to determine if the overall results were, perhaps, largely driven by outcomes in this group of patients.

Subgroup analyses were also conducted base on age, gender and race.

[Slide.]

Also unique to the analysis of these data was the use of hospital characteristics to define strata for analysis. Hospital characteristics that were found to be predictive of aprotinin use and that also provided adequately sized strata were

chosen to create strata.

This resulted in eight strata for comparison between aprotinin and the other agents and four strata for the comparison between aprotinin to no treatment.

Propensity score modeling within each stratum was then performed. This allowed for better propensity-score estimation. Propensity-score deciles were then constructed within each stratum and final estimates are weighted averages across all the strata

[Slide.]

Not all patients were used for all outcome analyses. For the analysis of acute heart failure and acute renal failure, which are the outcomes mentioned earlier that could not be evaluated on the day of surgery, patients who had zero days of follow up were excluded from the analysis.

As you can see here, most of those were due to death. For the outcome of acute renal failure, patients who met the criteria for pre-existing renal failure, which was virtually the