



June 26, 2006

An Open Letter from Merck

Dear Valued Colleague:

At Merck, we are committed to rigorous scientific research conducted under high standards of ethical behavior. This is at the heart of who we are and how we do business. When we identified an error in the description of one of the statistical methods used to analyze certain data in the APPROVe study, we promptly notified the study authors and the New England Journal of Medicine (NEJM), which had published the study in 2005. We explained the basis for the correction and set forth our view that the correction did not change the result that, in the APPROVe study, an increased relative risk for confirmed thrombotic cardiovascular (CV) events for VIOXX compared to placebo was first observed beginning after 18 months of ongoing daily treatment. NEJM published its own correction notice regarding the APPROVe study on June 26, 2006. As that correction does not reflect our view of the data and the results of the study, we are providing our assessment as part of the on-going scientific discussion of this issue.

Merck recently identified the need for a correction in the description of one of the statistical methods used to analyze certain data in the APPROVe study (NEJM, 2005). Merck communicated this finding to the study authors, the Journal, regulatory authorities and the public in a timely and transparent manner on May 30, 2006. When we first identified the need for a correction to the article, we carefully assessed whether the error in describing one of the statistical methods affected the scientific results reported in the article. The CV data from the APPROVe trial were to be analyzed as part of a combined analysis with 2 additional placebo-controlled trials (ViP and VICTOR). The CV Data Analysis Plan (DAP) for this combined analysis specified a battery of statistical assessments to evaluate whether the relative risk of confirmed thrombotic CV events on VIOXX compared with on placebo was constant over time. Within this battery of statistical assessments, the use of the variable, logarithm of time, was the primary method specified.

In the manuscript submitted to NEJM, the methods section referred to the use of the logarithm of time. This description of the method used for the report of the p-value for the test of proportionality of hazards was in error. The reported result (p-value = 0.01) came from a method using linear time, not logarithm of time. Results of diagnostic analyses indicate that a model using linear time is more representative of the data than one using logarithm of time. Thus, the linear time analysis is an appropriate method to assess the changes in relative risk over time. Recent tests show that the result using logarithm of time has a p-value = 0.07. Even this borderline significant result justifies concern regarding changes in relative risk over time.

The battery of statistical assessments together indicate that the relative risk was not constant over time. The APPROVe article showed this non-constant relative risk in Figure 2 and Table 3. These data are separate from and not subject to the correction in the description from log to

linear time. Further, these results have not changed. As indicated in the original paper, there was an increased relative risk for confirmed thrombotic cardiovascular events for VIOXX compared to placebo observed in the APPROVe study beginning after 18 months of ongoing daily treatment. Therefore, we conclude that this correction to the description of the statistical method does not change the results of the APPROVe study. A correction notice and rationale based on this assessment was submitted by all of the authors to the Journal on June 9, 2006. We reaffirm our assessments that the correction does not change the results of the APPROVe study.

A detailed assessment of this statistical issue by scientists within Merck is attached for your review.

We hope this letter and the attached detailed assessment have addressed any questions you may have had about the correction in the description of one of the statistical methods used to analyze certain data in the APPROVe study. We know that you and the scientific community expect Merck to adhere to high scientific and ethical standards and we intend to continue to work hard to meet those expectations.

Sincerely,

A handwritten signature in blue ink, appearing to read "PSK", with a horizontal line extending to the right.

Peter S. Kim
President, Merck Research Laboratories

APPROVe: ASSESSMENT

Background

As recommended by its safety monitoring board, the APPROVe trial was unexpectedly stopped before completion because of an increase in risk of CV events for rofecoxib compared to placebo. The DAP for the APPROVe trial referred to the DAP for the combined CV data from APPROVe plus that from 2 additional trials (ViP and VICTOR), known as Protocol 203. Protocol 203 was designed to demonstrate non-inferiority of CV risk for rofecoxib in comparison to placebo; the primary analysis to demonstrate non-inferiority was based on the 95% CI on hazard ratio (HR) from a Cox proportional hazards model. That model requires the assumption of constant HR over time. To test that assumption, the DAP for Protocol 203 called for numerous statistical and graphical methods to assess the relative risk (RR) of rofecoxib compared to placebo over time -- that is, to show whether the data were consistent with a null hypothesis of no difference from constant HR over time. Although the results of a single test may be consistent with this null hypothesis of constant HR over time, it is common statistical practice to conduct more than one assessment because no single test alone can prove that null hypothesis. Rather, the purpose of the analyses testing proportionality of HR over time in the DAP is to determine whether there is evidence to challenge the assumption of constant HR over time. For this reason, any test result demonstrating non-constant HR is relevant. In that context, the term “primary” as ascribed to the treatment-by-log(time) interaction test in the Protocol 203 DAP means that it is the initial approach, but not the only approach, to assessing proportional hazards. Thus, the aim of the Protocol 203 DAP was to demonstrate that a battery of statistical assessments yielded no substantial evidence of departure from the assumption of constant HR over time.

In the absence of a specific plan for the analysis of CV event HR over time in APPROVe alone, the analysis of the proportional hazards assumption for the APPROVe CV data proceeded in a manner consistent with good statistical practice and consistent with the DAP for Protocol 203.

Detailed Summary of Review and Findings

Merck performed numerous statistical analyses on the preliminary APPROVe data which led to the withdrawal of VIOXX. The CV results from this preliminary data showing significantly greater risk on VIOXX than on placebo in the APPROVe trial were surprising. Therefore, consistent with good statistical practice, many post hoc analyses aimed at understanding and appropriately describing those results were carried out on those preliminary data. These analyses were consistent with the Protocol 203 DAP, and included the K-M plot, the test for treatment-by-log(time) interaction in the Cox proportional hazards model, and computations of HR by time interval. Examination of the K-M plot showed similar cumulative incidence rates over the first 18 months and higher rates on rofecoxib than on placebo beginning after 18 months, suggesting non-constant hazard ratio over time. The test for interaction yielded a statistically significant departure from constant HR over time ($p=0.006$). While analyzing the newly obtained, off-drug follow-up data, it was observed that the latter result for the preliminary data was from a test using linear time; the test for treatment-by-log(time) interaction on the preliminary data yielded $p=0.048$. Time interval computations revealed RR close to 1 and CI encompassing 1 over the first 18 months, and $RR > 4$ with $95\% CI > 1$ over the > 18 -month time period. Over the 36 month period of the study, the RR was lowest in the first 3 six-month time intervals and highest

in the last 3 six-month time intervals. All of these analyses yielded evidence of non-constant hazard ratio over time.

The CV DAP specified splitting the three year study period into time intervals with approximately equal numbers of events in order to assess the HR over time. The CV DAP also specified computing HR in successive 6-month intervals, that is, equally spaced on the linear time axis; this latter analysis yielded approximately equal numbers of events in those intervals. Thus the distribution of observed events was approximately evenly distributed on a linear time axis; the distribution of events was heavily skewed on the log-time axis. However, as previously mentioned, the RR was lowest in the first 3 six-month time intervals and highest in the last 3 six-month time intervals. In order to statistically test whether the observed differential HR in the 6-month intervals were proportional, the treatment-by-(linear)time interaction test was carried out. Thus use of the linear time proportionality test derived from the analysis strategy set in the CV DAP.

While preparation of the final data set and its analysis was ongoing, additional assessments aimed at modeling the HR over time proceeded on the preliminary APPROVe dataset. The modeling effort proceeded since the initial results strongly supported non-constant HR over time. Results of the modeling showed that a Cox model with terms for treatment and treatment-by-(linear)time interaction fit the HR over time data better than one with terms for treatment and treatment-by-log(time). In addition, a piece-wise model, with a constant HR over the first 18 months and terms for treatment and treatment-by-(linear)time interaction beginning at 18 months, fit the data better than both of those models. These modeling results provide additional evidence for non-constant HR over time.

The finalized APPROVe locked dataset for the 2005 publication contained 72 CV events, 2 more than the preliminary dataset on which the extensive analyses of HR over time were carried out. The K-M plot, a statistical test of proportional hazards, and time-interval RR computations were repeated on the locked dataset. All these supported a conclusion of non-constant HR over time; these results were reported in the NEJM paper.

Overall Assessment

We recently discovered the error in the description of the test of proportional hazards reported in the publication. The reference to treatment-by-log(time) interaction should have been to treatment-by-(linear)time interaction. The reported p-value = 0.01 came from that test using linear time, not logarithm of time. Recent tests show that the result using logarithm of time based on the locked dataset has a p-value = 0.07. The other assessments relating to proportionality of hazards, which were carried out consistent with the DAP, remain unchanged: the K-M plot remains unchanged; the time-interval RR's and CI's remain unchanged; the better fit of the Cox model with linear time rather than logarithm of time remains unchanged; and the statistical significance of treatment-by-(linear)time interaction test remains unchanged. Thus, the null hypothesis of constant HR over time still remains unsupportable since all these analyses carried out on the locked dataset are consistent with non-constant HR over time. Even the treatment-by-log(time) interaction p-value 0.07 approaches statistical significance, consistent with all the other analyses, which support a conclusion of non-constant HR over time.

Conclusion

We conclude, based on these facts, that the correction to the description of the statistical method does not change the result that, in the APPROVe study, the relative risk for confirmed thrombotic cardiovascular events was not constant over time, and that an increased relative risk for confirmed thrombotic cardiovascular events for VIOXX compared to placebo was observed beginning after 18 months of continuous daily treatment. We provide our views here as part of the on-going scientific discourse on these issues.