



July 6, 2004

Dockets Management Branch (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Rm. 1061
Rockville, MD 20852

RE: [Docket No. 2004D-0189] – Draft Guidance for Industry on Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment - Request for Comments

Merck & Co., Inc. is a leading worldwide, human health products company. Through a combination of the best science and state-of-the-art medicine, Merck's Research and Development (R&D) pipeline has produced many important pharmaceutical products available today. These products have saved the lives of or improved the quality of life for millions of people globally.

Merck supports regulatory oversight of pharmaceutical product development and welcomes guidance for compliance that is based on sound scientific principles and good judgment. As a leading pharmaceutical company, Merck has extensive experience in thoroughly evaluating our products from discovery to approval and throughout their marketing life to assure that they continue to provide health benefits with minimum risk. Therefore, we are well qualified to comment on the risk assessment and risk management draft guidance documents issued by FDA on May 5, 2004¹. Herein, we are providing comment on the draft guidance for industry entitled: *Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment*.

General Comments

We commend the FDA for its efforts in the development of guidance for industry on good practices for risk assessment, risk management, and pharmacovigilance, and particularly for its prior issuance² of the three concept papers to encourage discussion of these important topics. The practice of issuing concept papers describing novel regulatory approaches, prior to issuance of draft guidance documents is fully supported by Merck.

69 FR 25130, Docket No. 2004D-0189
68 FR 11120, Docket No. 02N-0528

The concept paper provides one additional opportunity for interested parties to provide comment and is a valuable tool when guidance documents describing new regulatory concepts are developed. We fully recognize the extra efforts that the concept paper precipitates and we appreciate the agency's continued commitment to this approach.

The definition of pharmacovigilance provided in the draft guidance (Line 115 – 117) appears to limit the term to post-approval activities. We are concerned that this definition does not fully harmonize with the definition of pharmacovigilance contained in the ICH E2E draft on Pharmacovigilance Planning. ICH employs the WHO definition “the science and activities relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug related problem”. We fully agree with taking a life-cycle approach to pharmacovigilance throughout the drug development continuum, encompassing post-approval activities. ICH and WHO do not limit their definition of pharmacovigilance to post-approval activities. Both the ICH, WHO and FDA definitions encompass pharmacoepidemiologic studies, but the FDA limits the definition to pharmacoepidemiologic safety studies. We strongly recommend that the FDA adopt the internationally accepted definition of pharmacovigilance which includes pre-marketing, post-marketing and all pharmacoepidemiologic activities and that the term will be used consistently throughout the document.

We suggest that FDA include a definition of the term “signal” in the guidance document. The lack of a definition may cause the reader to over-interpret the term and to look upon signals as establishing or reflecting association or causality. Signals do not establish association, let alone causality. What constitutes a signal may depend on the data used to support the assertion that a ‘signal’ has been sent. A ‘signal’ from a prescription event monitoring database may not be, or mean, the same thing as a ‘signal’ from a spontaneous reporting database. A conclusion that a ‘signal’ represents a real clinical association generally would require detailed medical or epidemiological follow-up.

Specific Comments

IV. IDENTIFYING AND DESCRIBING SAFETY SIGNALS: FROM CASE REPORTS TO CASE SERIES

A. Good Reporting Practice

Lines 142-144: “*Spontaneous case reports of adverse events submitted to the sponsor and FDA, and reports from other sources, such as the medical literature or clinical studies, are potential signals of adverse effects of drugs.*” It would be more accurate to state that spontaneous case reports “... *may be potential signals ...*” because spontaneous report data analyses provide an indication of disproportionality of reporting, which could occur for a variety of reasons. The recommendation to obtain complete follow-up information (Line 146) is appropriate, but the implication for follow-up on reports from

spontaneous report databases is unclear. Additionally, as not all reports are equal, there may be more intense follow-up for some reports. Triage of follow-up efforts is appropriate and this should be noted in the draft guidance. It should be recognized that obtaining follow-up information, as appropriate, is already a standard course of action.

C. Developing a Case Series and Assessing Causality of Individual Case Reports

Line 222 – 223: *It is important to remove any duplicate reports.*

Since FDA can obtain reports from several sources, especially if a report mentions products from different sponsors, we recommend that this sentence be reworded to: *It is important for the Agency to remove any duplicate reports.*

Line 247 – 248: *FDA recommends that sponsors carefully evaluate confounded cases and should not simply dismiss them.*

We request that the term “confounded case” be clearly defined.

D. Summary Descriptive Analysis of a Case Series

Line 283 – 284: *After individual cases are assessed for causality, one or more of the cases may suggest a safety signal warranting additional investigation.*

We suggest rewording to: *After individual cases are reviewed, one or more of the cases may suggest a potential safety signal warranting additional investigation.* Individual cases should not be assessed for causality; in aggregate form, a search for overall trends is more appropriate. This is noted in Line 252 and should be re-emphasized within the document: *For any individual case report, it is rarely possible to know with a high level of certainty whether the event was caused by the product.*

E. Use of Data Mining to Identify Product-Event Combinations

Line 311 – 317: *At various stages of risk identification and assessment, looking systematically into the data by using statistical or mathematical tools, or so-called data mining, can provide additional information about the existence or characteristics of a signal. By applying data mining techniques to large adverse event databases, such as FDA’s AERS or VAERS, a sponsor may be able to identify unusual or unexpected product-event combinations warranting further investigations. Data mining is not the only technique used to make causal attributions between products and adverse events.* Along with “signal”, data mining is an undefined term in the draft guidance. At best, data mining is a tool that identifies potential reporting associations between drugs and events that may be worth further investigation. Consequently, it is not clear what ‘data mining techniques’ means. Does the reference to AERS and VAERS indicate that data mining is only applicable for spontaneous reporting databases? Line 317 implies that data mining could make causal associations, which is outside the capabilities of a data mining tool.

Line 319 – 323: *A method of data mining currently in use is the comparison of the fraction of all events reported for a particular product (e.g., liver failure), the "observed rate," with the fraction of reports for all drugs that are for that same event, the "expected rate." This analysis can be corrected for such characteristics as reporting year, age, and gender, and it is also possible to do the analysis for drugs of a specific class or for drugs that are used to treat a particular disease. We do not believe this paragraph describes a data mining method. The comparison described is a statistic that expresses the degree to which the adverse event report rate among patients whose reports mention a particular drug differs from rate at which the adverse event is reported among all patients whether their reports mention the drug or not.*

We believe the assumption that the statistic can be corrected for characteristics such as age, gender, etc. is inappropriate. One may calculate a standardized rate like a standardized mortality rate to adjust for differences among population, but this really is not a correction (which would imply that the uncorrected rate was in some sense 'wrong'). We suggest that the reference to data mining and "correction" for characteristics be eliminated from the paragraph.

Line 325 – 331: *The statistic (or score) used to quantify the disproportionality between the observed and expected values for a given product-event combination is compared to a threshold that is chosen by the analyst to optimize sensitivity and specificity. A signal is operationally defined as any product-event combination with a score exceeding the specified threshold. It is not unusual for a product to have several signals identified using these methods. The lower the threshold, the more likely it is that signals of true effects will be detected, but these lower thresholds will also result in more false positive signals.*

Since there is no definition of 'truth', nor any 'gold standard', it is not at all clear what 'sensitivity' and 'specificity' mean here. Also, the definition of a 'signal' used herein is not appropriate for the overall definition of the term as used elsewhere in the draft guidance and may create confusion.

Line 333 – 342: *Several data mining methods have been described and are worth considering, such as the Multi-Item Gamma Poisson Shrinker (MGPS) algorithm, the proportional reporting ratio (PRR) method and the Bayesian neural network approach. Except when the observed number of events is small (e.g., less than 20), these methods will generally give similar scores. These approaches are inherently exploratory and may provide insights into the patterns of adverse events particular to a given product relative to other products in the same class or to all other products. FDA recommends exercising caution when making such comparisons, however, because voluntary adverse event reporting systems such as AERS or VAERS are subject to a variety of reporting biases,*

because some observations could reflect concomitant treatment, not the product itself, and because the disease being treated may cause the events.

We note that MGPS is a commercial product, not an algorithm, per se. The correct term for the algorithm is the Empirical Bayes Gamma-Poisson Shrinkage algorithm, well-described in the refereed scientific literature. MGPS has not been described in the same way and we suggest elimination of the reference to MGPS. We suggest that Line 337: should read “patterns of adverse event *reports* ...” as reports, not events, are what are available. It is not true that all of the reports actually represent adverse events as such. Lines 339-42: We recommend that the statement be made stronger. Spontaneous reporting databases never should be used to make treatment comparisons, especially in publications.

F. Safety Signals That May Warrant Further Investigation

Line 357 – 360: *FDA believes that the methods described above will permit a sponsor to identify and preliminarily characterize safety signals. The actual risk to patients cannot be known from these data because it is not possible to characterize all cases definitively and because there is invariably under-reporting of some extent and incomplete information about duration of therapy, numbers treated, etc.*

Insufficient evidence is brought forth to support this belief. At best, these methods may be helpful as part of the tools a sponsor might use in carrying out pharmacovigilance. There is no evidence to support that these methods (e.g. data mining) are in any way necessary, let alone sufficient, for identifying potential safety issues.

G. Putting the Signal into Context: Calculating Reporting Rates vs. Incidence Rates

Line 395 – 415: *In contrast, for spontaneously reported events, it is not possible to identify all cases because of under-reporting, and the size of the exposed population is at best an estimate. Limitations in national denominator estimates arise because:*

- 1. National estimates of the number of patients exposed to a medical product and their duration of exposure may not be available;*
- 2. It may be difficult to exclude patients who are not at risk for an event because their exposure is too brief or their dose is too low; and*
- 3. A product may be used in different populations for different indications, but use estimates are not available for the population of interest.*

Although we recognize these limitations, we recommend that sponsors calculate crude adverse event reporting rates as a valuable step in the investigation and assessment of adverse events. FDA suggests that sponsors calculate reporting rates by using the total number of spontaneously reported cases in the United States in the numerator and estimates of national patient exposure to product in the denominator. FDA recommends that whenever possible, the number of patients exposed to the product nationwide be the

estimated denominator for a reporting rate. FDA suggests that other surrogates for exposure, such as numbers of prescriptions or kilograms of product sold, only be used when patient-level estimates are unavailable.

Lines 395-406 indicate that both the estimate of the number of cases of adverse events and the total number of patients exposed to the drug are unreliable. As such, the FDA recommendation (Lines 408-9) to calculate crude event reporting rates will result in an unreliable figure. Given the deficiencies of the numerator and denominator, it is entirely possible that the result of the calculation could become regarded, entirely incorrectly, as a number with meaning, and that actions could be taken as a result, without any real foundation. If a reporting rate is large enough to warrant further investigation, then it is worth undertaking a serious investigation using other databases (not spontaneous reporting databases) to provide a credible estimate of the incidence or risk of adverse event.

Line 417 – 423: Comparisons of reporting rates can be valuable, particularly across similar products or across different product classes prescribed for the same indication. However, such comparisons are subject to substantial limitations in interpretation because of the inherent uncertainties in the numerator and denominator used. As a result, FDA suggests that a comparison of two or more reporting rates be viewed with caution and generally considered exploratory or hypothesis-generating. Reporting rates can by no means be considered incidence rates, for either absolute or comparative purposes.

Lines 417-8: This statement is very likely to be misinterpreted by the public, including medical professionals. The inadvisability of using spontaneous reporting system information for comparisons of drugs needs to be stated much more strongly. Comparisons of drugs or drug classes on the basis solely of data mining computations carried out using data from spontaneous reporting databases are scientifically unjust, regardless of how elaborate the analyses are and regardless of what caveats might accompany such reports.

Conclusion

We commend the Food and Drug Administration for issuance of the three concept papers, followed by draft guidance documents, on premarketing risk assessment, risk minimization action plans, and good pharmacovigilance practices and pharmacoepidemiologic assessment. These documents, along with the public workshop on April 9, 10, and 11, 2003 represent an extraordinary effort on the part of the Agency to convey its preliminary thoughts on these issues and to stimulate discussion with stakeholders.

The call for guidance on risk assessment, risk management, and pharmacovigilance activities in the PDUFA III goals is neither an expression of concern that current efforts are inadequate nor a call for more intense surveillance. It simply a call to document those practices that represent the best of what we are doing now. Risk management, itself, is not new to drug development. As an industry, in conjunction with the FDA, we have been conducting pre-approval tests of increasing intensity and complexity on potential products for decades; we have been collecting, monitoring, and evaluating spontaneous reports on marketed products and taking appropriate action to minimize risks. Likewise, we have carried out Phase 4 programs based on commitments made to the Agency at the time of approval to address potential, often theoretical, risks that had not been resolved at the time of approval. It is the best of these practices that the guidance is intended to capture, along with fostering international harmonization with the approach.

We appreciate the opportunity to share our comments with respect to FDA's Draft Guidance for Industry: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment. Please do not hesitate to contact me, should you have any questions.

Sincerely,



Donald M. Black, MD, MBA
Vice President
Global Regulatory Policy