Medical and Regulatory Operations
Pfizer Inc
235 East 42nd Street 150/3/75
New York, NY 10017-5755
Tel 212 573 3940 Fax 212 808 8679
Email gretchen.dieck@pfizer.com

Gretchen S. Dieck, PhD
Vice President
Safety Evaluation and Epidemiology

May 29, 2003

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

Re: **Docket No. 02N-0528**; "*Premarketing Risk Assessment*," "*Risk Management Programs*," and "*Risk Assessment of Observational Data: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment*" (68 Federal Register 11120, March 7, 2003, and 68 Federal Register 25049, May 9, 2003)

Dear Sir/Madam:

The following comments on the above-captioned Risk Management draft Concept Papers and the Risk Management Public Workshops on April 9-11, 2003, are submitted on behalf of Pfizer Inc. Pfizer discovers, develops, manufactures, and markets leading prescription medicines for humans and animals and many of the world's best-known consumer brands. Our innovative, value-added products improve the quality of life of people around the world and help them enjoy longer, healthier, and more productive lives. The company has three business segments: health care, animal health and consumer health care. Our products are available in more than 150 countries.

Pfizer supports a Risk Management approach to ensuring availability of and access to safe and effective medicines by those who need them and we commend the Agency for actively engaging stakeholders in the development of guidance on the topic of Risk Management. Further, we support incorporation of Risk Management concepts early in the product development cycle as part of a continuum in the assessment of benefit-risk for each product and we believe that this should encompass a worldwide perspective. We endorse FDA's participation in Industry-Regulator consensus forums, such as the International Conference on Harmonization (ICH) and the Council for International Organizations of Medical Sciences (CIOMS), to maintain global consistency and harmonization on this important topic.

FDA Performance Goals related to PDUFA III include development of three guidance documents on Risk Management (drafts by September 2003 and final guidance to be published by September 2004):

- Premarketing Risk Assessment;
- Risk Management Programs; and
- Risk Assessment of Observational Data: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment.

In April 2003 FDA published a draft Concept Paper for each of these three topics and these papers were discussed at Public Workshops on April 9-11, 2003. We considered the published draft Concept Papers and discussion at the Public Workshops when developing our comments.

We have identified several areas that we believe require additional attention as FDA develops guidance from the draft Concept Papers and workshop discussions. Our general comments on these areas are:

- Stakeholder dialogue is essential. Relevant stakeholders should be involved in both the development of guidance and in the implementation of Risk Management Plans and Programs. Mechanisms should be established to ensure (a) Dialogue between the Agency, Sponsor, and others, when appropriate, and (b) Interaction within the Agency, e.g., Reviewing Divisions and the Office of Drug Safety. We believe that it would be appropriate to establish a schedule of opportunities for dialogue at various stages of a product's lifecycle. Collaborative discussion of strategy and interpretation of data should result in a common understanding of relevant issues. We believe that this will provide a platform for constructive interactions in the best interest of the public health and will minimize misunderstandings. Further, it should be emphasized that all data sources should be considered – no single source of data should be used in isolation;

- Any consideration of risk must be balanced with consideration of benefit. Risk Management is ultimately about ensuring proper balance in regulatory decisions. In assessing the balance between benefits and risks, it is important to give sufficient weight to the benefits.

- Risk Management is a continuum. We believe that the concept of Risk Management should begin early in product development and evolve at each phase of development as additional information is accumulated. However, all products are not the same and the need for Risk Management activities should be considered on a product-by-product basis;

- Terminology must be clear. Important phrases such as "Risk Management Plan" and "Risk Management Program" should not be used interchangeably. It is essential to clarify the meaning of these terms and others, such as Pharmacovigilance Plan, and to use them consistently. A Risk Management Program, for example, should be reserved for those very rare situations in which a known risk must be managed with a targeted intervention to maintain an acceptable balance of benefit and risk;

- International harmonization provides advantages. Risk Management is a shared global responsibility and stakeholders should endeavor to avoid multiple strategies to serve local needs. Thus, care should be taken to incorporate consensus definitions and approaches, e.g., those developed by ICH and CIOMS, wherever possible to ensure most efficient use of resources;

- Consensus must be reached on tools. Simplicity and flexibility are the cornerstones of appropriate tools. Agreement is needed on which tools are appropriate and how and when each tool should be used. For example, restricted distribution is an extreme form of a Risk Management tool. In those rare instances when restricted distribution is considered, it should be applied with care and only considered for drugs with high risk that is established (not theoretical). Further, a clear distinction should be made between tools that should be used to characterize risk versus those that can be applied to manage risk. For example, a case control study that is conducted as part of a Post-Approval Commitment may be useful to learn more about a certain risk, but such a study should not be considered useful as a tool to *manage* risk;

- A uniform approach to labeling is needed. Prescribing Information should be evidence-based and standardized where possible, e.g., agreement should be reached on what information goes into each section of product labeling and standard criteria should be developed for

bolded, *italicized,* and black box wording. We believe that this would facilitate product comparisons by prescribers;

- Prescribers should be permitted to practice medicine. Manufacturers have a role in educating prescribers, pharmacists, and patients on the best use of medicines, but manufacturers should not be expected to supervise physician prescribing practices or to enforce nuances in Package Inserts; and

- Good Guidance Practices are encouraged. Expectations in the three anticipated guidance documents should be tied directly to FDA's current legal authority to regulate the safety of drugs. Namely, FDA's expectations for regulated companies' Risk Management activities should be tied directly – and exclusively – to whether these activities help to ensure that marketed drug and biologic products are safe.

In summary, Pfizer endorses the thoughtful use of Risk Management concepts and practices throughout the continuum of a product's lifecycle, i.e., during the pre-approval, peri-approval, and post-marketing phases of product development. We believe that dialogue among stakeholders is key and we view Risk Management as a global process.

Our specific comments on the three draft Concept Papers are attached.

Finally, we support comments made by the Pharmaceutical Research and Manufacturers Association (P*h*RMA) at the Public Workshops and we also support P*h*RMA's written comments to Docket 02N-0528. We thank FDA for the opportunity to comment on this important topic and we would be pleased to respond to any questions that the Agency might have. We welcome the opportunity to join other stakeholders as a sounding board for Agency ideas as the three draft guidance documents are developed.

Sincerely,

Gretchen S. Dieck

Gretchen S. Dieck, PhD

GSD:sl

I. SPECIFIC COMMENTS

Premarketing Risk Assessment

These comments apply to the FDA draft Concept Paper titled "Premarketing Risk Assessment," dated March 3, 2003, and the Public Workshop held April 9, 2003, in Washington, DC. Comments are arranged by the topic of concern; line numbers refer to text in the draft Concept Paper.

1. Overall Comments

The Agency should clarify whether any or all of the concepts in this draft Concept Paper apply to activities (or data) that should be completed before the registration package is finalized or whether these activities can extend through the peri-approval period and be completed prior to marketing. The paper is entitled "premarketing," but, as early as line 13, text indicates that the paper is concerned with "analysis and presentation of safety data for approval." Since many new data are acquired between initial dossier submission and final approval, and the two are usually at least six months apart, this distinction can have large impact on the conduct of drug development programs.

In addition, the expectations set out in this draft Concept Paper and the two companion documents ("Risk Management Programs" and "Risk Assessment of Observational Data: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment") should be tied directly to FDA's current legal authority to regulate the safety of drug and biologic products. Namely, FDA's expectations for regulated companies' Risk Management activities should be tied directly – and exclusively – to whether these activities help to ensure that new drugs and biologic products are safe. FDA should avoid incorporating redundant or ineffective activities in guidance; we encourage the use of Good Guidance Practices in the development of the three anticipated guidance documents.[1]

2. Size of the Pre-Registration Database

On lines 110 –114 of the Concept Paper, it is suggested that there are several circumstances under which a safety database larger than that in the ICH E1 consensus guideline would be needed:

- "when there is a need to quantitate [sic] the occurrence of an expected low-frequency adverse event" or
- "when the benefit is small or experienced by a fraction of treated patients, or of uncertain magnitude"

We believe that the need for a larger safety database should be triggered by a specific concern raised by Phase II or Phase III data for a specific product. Furthermore, we think it prudent to define models for safety assessment and pharmacovigilance that apply to the entire drug development continuum, rather than to a single segment. The concepts set forth in a recent publication by Patrick Waller and Stephen Evans suggest that "safety can only be provisionally established at the time of authorization and that safety experience needs to be extended (in clinical practice)" to specific milestones or pre-specified levels of exposure, which are dependent on the type of drug, its use and the evidence concerning its safety. (Reference: Waller, PC and Evans, SJW: 2003. *A model for the future conduct of pharmacovigilance.* Pharmacoepidemiology & Drug Safety, 12:7-29).

To better understand the Agency's perspective on deviating from consensus agreements on sample size, it would be helpful for FDA to provide specific examples of past situations that needed larger pre-registration databases, the approaches that were used in these situations, and the resulting outcomes (i.e., FDA should provide examples patterned on that included in Appendix D of the CIOMS Working Group IV Report, as applied to Benefit–Risk evaluation). If there has not been any precedent, a detailed

[1] Indeed, these comments apply equally to the "Premarketing Risk Assessment" draft paper and to the "Risk Management Programs" and "Risk Assessment of Observational Data: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment" draft Concept Papers.

approach for a hypothetical example of a product intended for long-term treatment of non-life-threatening conditions would be helpful (i.e., FDA should provide a description of a chronic or recurrent intermittent use situation, triggers for a larger database, and possible approaches to address the identified issue).

In addition to sample size, there are many other factors that may impact the interpretation of safety data collected during the pre-marketing period of product development. In some instances, these factors differ from those applied to interpretation of post-marketing data. A decision analysis approach to address potential safety issues that arise at different stages of drug development, performed with the assistance of a model, may be of benefit, especially when the approach is evidence-based. At each stage of drug development, there is a set of decision criteria including assumptions as to whether a threshold should exist to proceed with additional safety investigation. For Phase III trials, Data Monitoring Committees may be chartered to address safety issues in conjunction with pre-specified decision rules. The latter is especially relevant for large blinded Phase III trials. Under the decision analysis approach, it will be useful for the sponsor and Agency to agree on the criteria to evaluate safety risk with sufficient precision at pre-specified milestones; criteria should be based on sample size to establish whether the lower bound of a risk estimate exceeds a certain risk tolerance or risk threshold. In any decision analysis approach, however, it is very important to emphasize that consideration of risk must be balanced with consideration of benefit. Risk Management is ultimately about ensuring proper balance in regulatory decisions. In assessing the balance between benefits and risks, it is important to give sufficient weight to the benefits.

We agree with the Agency that mathematical modeling can be useful for quantitative risk assessment in certain instances. However, it would be helpful to know which methodological approaches the Agency favors. Does the Agency support formal hypothesis testing concerning population frequency? It would be helpful for the guidance to describe situations in which precision of a risk estimate assumes greater importance.

The quantitative risk assessment and estimation issues that are posed by the different scenarios presented in the draft Concept Paper can be quite complex and multiple influencing factors must be considered. Risk models can differ not only in their mechanistic and mathematical bases, but also in ways that affect their application to data. Characteristics of risk assessment models and mode of their application (e.g., use of central vs. upper bound estimates, projection of risk over a defined time period or lifetime) make a considerable difference in the final estimate of risk. There is the potential to arrive at largely divergent quantitative estimates, depending upon the methodology and assumptions employed. A special case is presented when there is discordance between animal-based estimates and human data – considering the situation in which human data show much lower risks than would be predicted by extrapolations from animal data.

The Agency should also be more explicit on how it factors in other aspects of the safety database, such as strategies to manage risk (both pre and post approval), into its recommendation on overall database size. Some of these considerations might be strategies to manage risk (both pre- and post-approval) and the quality and relevance of safety data collected post-approval. Some guidance as to how these safety database size recommendations would apply under 21 CFR 314 subpart H would be helpful, since the stated intent is to use limited, interim safety data. These interim data should be coupled with an appropriate development plan to expand the safety database in the post-marketing period.

Similarly, the Agency should reconcile recommendations in this draft Concept Paper with those in previous guidance, which describe stepwise expansion of safety data for an oncology product that initially achieves accelerated approval based on Phase II trials with a total database of 200 patients ("*Guidance for Industry Cancer Drug and Biological Products-Clinical Data in Marketing Applications*," March, 2001, Section IV.)

In both the draft Concept Paper (line 136) and during the Public Workshop, the Agency suggested that it would be appropriate to develop a larger pre-registration database for agents likely to have a rapid uptake in a large population. The presumption here is that rare events will happen relatively early in the drug's marketed life due to the rapid uptake. There are several difficulties inherent in this approach, however:

1. Detecting rare events in a pre-market database is unlikely, depending on the rarity of the event. Table 1, below, gives the probabilities of observing at least one event (3rd column) or at least two events (4th column), given various sample sizes and event rates. As illustrated in the table, even doubling the size of a safety database will not markedly increase the likelihood of finding such an event. In order to find two events (one initial, one confirmatory) at the 1 in 10,000 rate, for instance, a database of 20,000 patients on the investigational drug would only provide about a one in two chance of detection, with about a 15% chance that the event would go entirely unobserved.

   Table 1. Probabilities of observing at least one event or at least two events for various sample sizes and event rates.

Event Rate	Sample Size	Probability (at least 1 event)	Probability (at least 2 events)
1%	500	0.993	0.960
0.5%	500	0.918	0.713
	1,000	0.993	0.960
0.1%	1,500	0.777	0.442
	3,000	0.950	0.801
0.01%	6,000	0.451	0.122
	10,000	0.632	0.264
	20,000	0.865	0.594

2. Pre-market predictions of rapidity of uptake are very difficult to make. For example, independent financial analysts that follow the pharmaceutical industry spend considerable effort trying to make such predictions and often differ from the real data by a factor of several-fold.

Finally, the approach, while not adding significantly to understanding of risks, will serve as a barrier for entirely new products that have the potential to address unmet medical needs. These are exactly the situations when the Agency needs to work with industry to lower barriers. Concerns around the safety of these agents once introduced to the marketplace are better addressed with a carefully constructed post-marketing safety surveillance program. In this way, rare events, likely to have been missed even in a 2-3 fold enlarged pre-registration database, are much more likely to be uncovered early, and in proportion to the true volume of use, rather than a guess as to what uptake will truly be. Such plans can be crafted, as they currently are, during the time of labeling negotiation.

The draft Concept Paper and Agency comments at the Public Workshop appear to endorse the suggestion that an enlarged pre-registration database (line 134) is indicated when a "very safe alternative to the investigational product is already available." For standard safety information, the usual pre-registration safety database will provide answers regarding the safety of the new agent, which can either be directly compared to existing agents within the dossier, or by comparison of dossier data with what is known of marketed agents. For very rare events, the same issues apply as mentioned in the paragraph above, and no reasonably sized pre-registration database is likely to address this.

As pointed out by some of the comments at the Public Workshop, (e.g., Judith Jones, who represented the International Society of Pharmacopepidemiology and Epidemiology, among others) a clear and consistent approach to what "very safe" means must be agreed. If this could only be established after many millions of patient exposure years, this would represent an unfair barrier to the innovative agent, which will always have a much smaller safety database at the time of registration. Furthermore, this construct does not allow for factoring in therapeutic trade-offs. A marketed product may be "safer" than the experimental one, but at the cost of less efficacy, which may, for people with more severe disease, translate into a worse outcome for this subpopulation than would be observed with the "very safe" agent. Alternatively, the marketed agent may be "very safe" in the population in general, but the new agent may have improved safety in certain significant populations of intended users. Hence, the concept seems difficult to apply, both in terms of arriving at a definition, and in making clear comparisons.

Rather than attempt to obtain this information in the pre-marketing phase of development, it may be more efficient to obtain the information in the post-marketing phase. Post-marketing surveillance can better assess safety of the new drug in actual use and in specific Phase IV studies; Phase IV studies can assess comparative usage in real-life clinical settings.

3. Use of Long-Term Controlled Clinical Safety Trials

In lines 143 ff, the draft Concept Paper discusses the contribution of long-term controlled safety studies as a desired segment of most, if not all, pre-registration safety databases. Although the benefits of controlled data over open-label data are described in the paper, there are several serious practical and ethical constraints on conducting such studies. These constraints must be balanced against the potential information that such studies can produce. In general, given acceptable available therapy (or even when the investigational agent itself shows promise of benefit based on the result of Phase II studies), it would rarely, if ever, be ethical to conduct long-term, placebo-controlled safety studies in any indication where there is morbidity or mortality benefit. Even with a standard of care or active comparator, we feel that significant practical constraints limit conduct of these types of studies to situations where certain conditions can be met:

- There is a specific safety-related hypothesis that is being tested so that the study has an adequate statistical power. A pre-specified hypothesis is important in view of an inflated overall false positive rate as a result of testing multiple endpoints without statistical adjustment. This point was well expressed during the Public Workshop by Dr. Simon (National Cancer Institute), who pointed out the potential for "innumerable endpoints and subset analyses;"

- The standard of care being used as a control is sufficiently effective relative to the experimental agent that differential dropout rates will not confound the desired comparisons. As pointed out by both the Agency and other participants during the workshop, imputation of missing safety data is a very difficult issue. Any efficacy differential in a long-term study could increase the need to impute safety data, severely compromising any conclusions that can be drawn by the study;

- There is a clear understanding that it is feasible to conduct the trial and that the population to be studied will tolerate such a trial. This entails a prospective determination that the trial is feasible and informed consent is clearly written for patients. In order for this to occur, the disease under study must not be sufficiently rare that a controlled study will compromise getting adequate exposure on the investigational agent due to inability to recruit and maintain study subjects in long-term, controlled, safety studies; and

- There is an agreed process for accepting and reviewing long-term safety data generated during the review period (e.g., data generated post-submission and up to review period).

If these conditions were met and two adequate and well-controlled studies were conducted against an active comparator, with pre-specified research objectives and hypotheses and the results of the studies support the hypothesis, relative safety claims (whether the claims are on superiority or non-inferiority) should be allowed. This would be in the context of performing these studies for specific hypothesis testing, with agreed *a priori* endpoints and outcomes, as is done for efficacy studies.

4. Degrees of Diversity in the Pre-Registration Database

Lines 159 ff in the draft Concept Paper discuss desired diversity of a pre-registration database, and this was the topic of some discussion at the workshop as well. It remains unclear, however, how these suggestions differ from current existing guidance for sex, race, age, disease severity, and other diversity factors that are already intentionally studied in current development programs. The suggestion that inclusion criteria be broadened to include very elderly, or concomitant medications should be discussed on a case by case basis at the End of Phase II meeting, so that the proper balance can be struck between (a) state of knowledge of drug safety at that time in development vs. risk to patients by widening criteria, i.e., when to broaden criteria during Phase III and (b) diversity vs. subgroup size. For more extreme conditions specified (e.g., very elderly patients taking a particular concomitant medication of interest), only a very small number of subjects may be in the database even with the use of intentionally broad inclusion criteria. Having information on a dozen or so patients in a pre-registration database, even if all patients did well, should not give a false sense of security, or undermine the need to assess safety in these groups in the post-marketing setting. Conversely, if one of the patients in these very small subsets has an adverse event, this does not provide definitive information either.

Certain types of diversity, such as a diversity of diet or the diversity of genetic loci of the pre-registration population simply cannot be studied on a systematic, across-the-board basis, as certain comments at the workshop implied they should be. Diet, for instance, is notoriously difficult to collect accurately. Hence, as noted above, specific drug-diet or drug-supplement interactions need to be identified on a for-cause basis (i.e., based on the metabolism of the drug) and studied in a controlled, hypothesis-guided, drug-drug interaction setting. Similarly, it is not practical, and very prone to Type II error, to study a random large set of genetic loci against safety and efficacy in general. Again, this must be hypothesis-driven, so that specific genetic loci are studied against specific outcomes that are based on a plausible hypothesis. For example, if a drug is metabolized by CYP 2C9 and hepatotoxicity is observed in a small group of patients, a plausible hypothesis for testing against the genetic material banked in Phase III may be focused on a relationship of 2C9 polymorphisms to hepatotoxicity. "Fishing expeditions" of hepatotoxicity against a large number of genetic loci simply because they can be studied, however, will be more likely to mislead due to false positives than to enlighten.

Hence, we agree with a judicious broadening of entry criteria as Phase III progresses, to be done on a case-by-case basis discussed with the Agency at the End of Phase II meeting. However, we caution against trying to analyze some forms of diversity (e.g., diet, genetics) outside of a hypothesis-testing framework and without a validated methodology to study the subgroup or interaction. We also caution against drawing too many inferences and taking a false sense of security out of very small numbers of subjects representing the "fringes" of the diverse population enrolled in the database.

In recent years the Agency seems to have been moving away from requesting certain formal interaction studies (except for probing CYP metabolism, and more recently pGP substrate interactions) and relying more on population PK analysis and use of *in-vitro* tests. It is not clear whether this is still the underlying philosophy since the draft Concept Paper (line 191) advocates many DDI studies based more on concomitant usage than on metabolic pathways. At the workshop, the Agency re-endorsed population PK methods. We would like to see the current approach continued, where sponsors and FDA agree, on a case-by-case basis, on the appropriate combination of database diversity analyzed by population PK techniques with specific drug-drug interaction studies. The safety database at registration should, therefore, comprise such a blended approach. It should be recognized that some interactions such as drug by diet are not practical to obtain on a Phase III population PK basis (see paragraphs above) and, if there are specific concerns with diet or dietary supplements, such components probably need to be tested in specific food-drug interaction studies.

5. Range of Doses Examined in Phase III

In both the draft Concept Paper (lines 170ff) and in comments at the workshop, the Agency advocates the examination of a range of doses in Phase III in order to fully characterize safety and efficacy of the new agent at those doses. The argument is made as well, that the safety information from patients given doses lower than the ultimate labeled doses is often not as meaningful as the safety information from

patients treated with doses in final labeling (or above). We agree with many aspects of this statement, in particular, that doses at or above the recommended labeled doses contribute more meaningfully to the safety database than doses below these levels. We also acknowledge that in some therapeutic areas, such as CNS or cardio-vascular diseases, large Phase III dose response studies are often conducted. Such studies, however, are in the setting of doses, all of which, based on the knowledge acquired at the end of Phase II, had the potential to be labeled doses. If properly conducted, Phase II studies should rule out many doses that, either because of inadequate efficacy or because of tolerability or adverse events, are deemed sub-optimal. This may leave only one dose possibly suitable for labeling, for study in Phase III, or it may leave several. In the latter case, multiple doses are often brought forward into Phase III. It may well turn out that of three doses studied, only one or two are suitable for labeling at the end of Phase III. Hence, in this case, dose selection would have gone on during Phase III and one or more doses eliminated. If the eliminated doses are higher than the ones selected for labeling then substantial data will exist on higher than recommended doses, which enhances the safety database at registration. This will **not** have come about, however, because the sponsor chose to extensively study a dose known to be sub-optimal. Rather, the dose was extensively studied on the assumption that it could be a labeled dose and only subsequently determined to be sub-optimal.

In our view, however, this scenario represents somewhat of a failure of the drug development process rather than something to be emulated. In the ideal drug development process, sub-optimal doses would not be used beyond Phase II. To the extent that in the real process such doses can be determined in Phase II, they should not be carried forward into Phase III. If insufficient information exists at the end of Phase II to make this determination, then a given dose should or studied in additional Phase IIb studies to make this determination, or, failing that, carried forward into Phase III. The dose should not be brought into Phase III solely to provide additional dose-response information, however. It is our recommendation that the approach at the End-of-Phase II meeting remains that the sponsor and Agency agree on which doses are clearly sub-optimal and these should not be studied in Phase III. At the end of Phase II, only those doses that appear to have potential to be the labeled doses should be carried forward into the Phase III program.

6. Use of Specialized Safety Data Collection Instruments

In both the draft Concept Paper (lines 269ff) and during the discussion at the workshop, the Agency advocated the use of specialized instruments in Phase III to elicit and better characterize less obvious adverse events. We would recommend that the Agency clarify what type of Phase II data, or other data, would trigger use of such instruments in Phase III and better clarify the circumstances where the Agency feels such testing should be done. In general, we would advocate that these recommendations be hypothesis-driven, not hypothesis testing, and be based either on extant Phase II data with the drug candidate, or known effects of the drug class or chemotype. The program of specific testing, as with genetic or diet/dietary supplement testing discussed above, should be a topic for discussion at the End-of-Phase II meeting, both in terms of the hypotheses to be tested, and how to best test them in Phase III.

7. Large Simple Safety Studies

In lines 279 ff of the draft Concept Paper, the Agency proposes that under certain circumstances, large simple safety studies (LSSS) should be conducted in Phase III.

It would be informative for the Agency to provide examples that illustrate FDA's experience with LSSS that have been conducted as part of post-approval commitments. This and other information could help to determine the appropriate frequency of rare adverse events to identify (as outlined in our comments above regarding overall pre-registration database size). LSSS are rarely simple, despite the name. These studies have frequently generated hypotheses rather than answering questions as intended and, hence, they are not a simple and direct means of assessing key safety issues. The recent COX-2 related CLASS and VIGOR studies, which were designed as "simple" studies to assess time to an event, are striking examples of this. Furthermore, despite their large size, such studies are still not large enough to determining risk factors around very rare events where routine post-marketing surveillance would be a more effective tool. At best, LSSS will only ascertain these rare events on a sporadic basis, but not provide any information about risks for them. Hence, LSSS size considerations must be based on the

need for meaningful precision of a safety risk estimate (or width of the confidence interval) or the desire to distinguish rates from the background or control group rates. We believe that size requirements be addressed in a meaningful manner only when the expectation is clearly specified.

Prior to the conclusion of Phase III, it is rarely possible or ethical to conduct a LSSS without knowing the dose(s) that have demonstrated favorable benefit/risk information and substantial knowledge of the safety profile. Therefore, conduct of a LSSS may be most appropriate as a Phase IV commitment as part of the ongoing approach to managing risk. In the pre-approval setting, LSSS are probably best used when there are specific events (perhaps identified safety signals) which are discernable in Phase II and no other Phase III approach seems adequate to explore such signals.

8. Hepatotoxicity Testing

Pfizer agrees with the draft Concept Paper recommendation (line 343) that potential hepatotoxicity should be evaluated in all development programs. Further, Pfizer agrees that detection of hepatotoxicity is an important safety component of each drug development program. FDA, P*h*RMA, and the American Association for the Study of Liver Disease (AASLD) have initiated a collaborative effort to detect and further understand drug-induced hepatotoxicity. The presentations and the papers from the initial conference sponsored by this effort are available via the FDA web site (i.e., *Drug-Induced Liver Injury: A National and Global Problem* - 12-13 February 2001, Westfields Conference Center, Chantilly, VA; http://www.fda.gov/cder/livertox/default.htm). The initiative sponsored by FDA/P*h*RMA/AASLD and the Hepatotoxicity Working Group continues and P*h*RMA has already agreed that specific hepatotoxicity proposals should follow the recommendations of this consensus group. Pfizer also endorses these recommendations on how to approach hepatotoxicity testing.

9. Groupings of Dictionary Terms

In lines 389-393, the draft Concept Paper recommends that the sponsor, in consultation with FDA, develop case definitions for medical conditions that may be of interest and prospectively group adverse event terms for data retrieval. We agree that it may be beneficial to lump dictionary terms to better characterize certain adverse events or a constellation of signs and symptoms in some circumstances. However, we suggest that the Agency provide clarity on the following points in formulating a strategy to group adverse event terms. Specifically, we recommend that:

- The Agency states the best timing for this consultation. For example, will it be at the end of Phase II? Also, it will be helpful if the Agency could suggest a mechanism to carry out this collaboration in a timely fashion, balancing the need for early planning with that for making data-based decisions. The ultimate goal should be to work together without causing any delay in regulatory decisions and patient's access to beneficial pharmaceutical and biologic products; and

- The Agency applies a uniform grouping strategy to drugs within a class to make sensible comparisons between them. If, at the time a specific strategy for grouping terms is being developed, consensus panels, e.g., CIOMS, have already proposed relevant groupings, these predefined groupings should be considered for use but not necessarily without modification. When comparing across databases, coding conventions and other aspects of data input must be considered.

10. Assessment and Validation of Safety Biomarkers

In its discussion of safety biomarkers (line 218 ff) the draft Concept Paper states that the "same dataset would not appropriately be used to both validate and assess the use of the new marker." The Agency should specify the meaning of the term "validate" and compare this with the intended meaning of the phrase "assess the use," since this is somewhat ambiguous. We interpret the statement as the following: "*The same dataset would not appropriately be used to both validate the biomarker and assess the safety of the NCE.*" We generally agree with this statement if that is the concept the Agency intended to convey. The operative word is "dataset" since this implies that the "learn and confirm" paradigm can be applied within a single program on *sequential* datasets. This concept is implied in the draft Concept Paper,

although the Agency's intended meaning is again somewhat ambiguous. However, it is highly unlikely that there would be sufficient data in a single program to "sufficiently validate a new safety biomarker, and is probably not the desirable strategy either since sponsors would want to assess sensitivity and specificity to a variety of true positive and true negative drug classes which are not likely to coexist in a single program." Furthermore, even if one program could be used to validate a new biomarker, the program's data would be unavailable to actually use the biomarker. What we would hope to do would be to validate new biomarkers over multiple programs, prior to their regulatory use. For that to be realized, however, the Agency would need to have an approach towards sponsors using studies and drug programs as vehicles to validate biomarkers *for future use*. However, this raises the possibility that the Agency could take precautionary and premature use of a not-yet validated biomarker to make safety inferences about the drug candidate being used in the validation process. We strongly believe that failure to address this issue will impede investment in clinical safety biomarker research.

There are alternate approaches to dealing with the issue of drawing inferences from partially validated biomarkers, and further work in this area should address which strategy the Agency would encourage. Here are three options:

- Sponsors make limited and cautious use of clinical safety biomarker research/validation across multiple programs for fear of the issue above. Without other options, this would probably be the default position of most sponsors;

- Sponsors conduct safety biomarker research on anonymized datasets post-hoc, following a set of principles that prevents internal inappropriate use (e.g., by setting performance standards and agreed action levels in advance) and do not share with the Agency until/unless fully validated; and

- Sponsors conduct safety biomarker research in collaboration with the Agency within a safe harbor that prevents the issue above by predefining the biomarker performance standards. The Agency must agree not to use the biomarker for regulatory purposes until those performance standards are met. Any type of harbor that would incorporate arbitrary conditions would not be considered "safe" and many sponsors would almost certainly default to strategy 1 or 2.

Finally, we would like to comment that the substantial attention given to safety "events" in the document is confirmation of the absence of effective safety biomarkers (most of which are continuous variables rather than discrete events) that would detect safety issues before they became clinical events. As is pointed out in the statistical table, above, detection of rare events is appallingly difficult, which underscores the need for a more proactive approach than strategy 1, above, for safety biomarker research.

11. Pre-Registration Research on Preventable Medication Errors

In the discussion of how sponsors can minimize preventable medication errors (lines 313-335), the draft Concept Paper advocates analyzing information, which is not readily available to sponsors (e.g., research concerning potentially confusing trade names). We agree with the Agency that pre-approval activities to minimize medication errors need to be developed and validated. Agency experiences and data for evaluation could help inform sponsors and also help focus efforts on processes and activities that would lead to the desired outcome.

The controlled clinical trial situation is designed to enforce compliance with the appropriate medication regimen. In the pre-marketing setting, there is no obvious way to simulate the relatively uncontrolled sequence of activities involved in drug prescription and drug acquisition from the general pharmacy (of choice or convenience). Some of these data can be acquired through simulations, while some data will have to be acquired post approval. Even then, the peri-approval period may not be the optimal time to determine some of the errors. The Agency should also consider the role of electronic prescribing in helping to minimize medication errors in the future.

12. Analysis of Pooled Data

In line 481 it is noted that person-time can be used as a denominator for pooled data. While we agree that this may be appropriate in some instances, it should be noted that there are instances when this is not appropriate. For example, such an analysis applied to the evaluation of idiosyncratic reactions could be misleading.

In lines 495-499 the draft Concept Paper states that when the results of a pooled analysis show a diminished statistical association and/or less risk compared to the safety signal originally obtained from one or more of the contributing clinical trials, it could suggest inappropriate use of data pooling. We do not agree with this cautionary statement.

While we agree that pooling should be based on sound scientific rationale, when more data are pooled based on pre-specified principles, the previously observed event rates could go up or down due to sampling fluctuations. The fact that a particular rate goes down after more data become available does not necessarily imply the loss of sensitivity or inappropriate pooling as long as the pre-specified pooling strategies are followed. This is especially true if an earlier high observed rate was a result of observing one or two events in a small sample.

We suggest that lines 495-499 be removed. In their place, we suggest that FDA reiterate the importance of following an appropriate pooling strategy that is determined in advance based on good scientific considerations.

13. Cost-Effectiveness of the Proposals

There were several comments made during the Risk Management Public Workshop that questioned statements regarding the effect of FDA's proposed changes in risk assessment and risk management methods on drug safety at launch of a new product. Articulating the underlying assumptions that the Agency will use to balance the cost of these proposals with the benefits likely to be achieved in terms of safer drugs at the time of approval would help all parties involved in the process better understand where to focus their efforts. Articulating such an argument will also be necessary for FDA to justify the proposals to the Office of Management and Budget.

Sponsors and the Agency share in the desire for safer drugs and a lower burden of preventable medication errors in our Society. Considering all of the newly proposed initiatives and suggested practices, as outlined in the three draft Concept Papers, it is not clear which components have the highest priority and how much of admittedly limited additional sponsor resources the Agency expects to be required to implement these new initiatives. Careful assessment of the impact of these new initiatives should include evaluation of potential benefits that could be realized by the alternative application of required resources.

Finally, it is important to note that the adoption of good risk assessment and risk management practices in drug development could be a prelude for instituting earlier "conditional" approval of medicines on the basis of a smaller volume of clinical data, for example, after completion of an expanded Phase II program or on the basis of a single pivotal Phase III study.

II. SPECIFIC COMMENTS

Risk Management Programs

These comments apply to the FDA draft Concept Paper titled "Risk Management Programs," dated March 3, 2003, and the Public Workshop held April 10, 2003, in Washington, DC. Comments are arranged by the topic of concern; line numbers refer to text in the draft Concept Paper.

1. General Comment

Pfizer believes that the framework for Risk Management Programs is an excellent first step toward creating standards for sponsors to follow when considering how to minimize risks associated with their medicines. We suggest renaming the proposed document "Risk Minimization Programs," since Risk Management refers to the broader process of risk assessment and pharmacovigilance that occur throughout the drug development process. We believe, based on statements from FDA at the Public Workshop, that a Risk Management Program (RMP) would be required only in rare cases when a drug carries unusually severe risks as exemplified by such drugs as Accutane®, Clozaril®, and thalidomide. This point needs to be stated more clearly in the document.

2. When Is a Risk Management Program (Risk Minimization Program) Appropriate?

We agree with the Agency that in most cases the decision to implement an RMP should be made on a case-by-case basis (lines 120-2). It is important that when FDA initiates a request for an RMP (lines 108-10), the Agency and sponsor will need to have a thorough discussion so that the parties reach a common understanding as to the nature and magnitude of the risk and the appropriate intervention needed to minimize it. It is also important to avoid situations where the Agency simply requests an RMP without any discussion around the precise risk of concern and what the major goals and objectives of a potential plan would be. Real-time input of key stakeholders, including physicians, pharmacists, and patients, may be necessary in some cases to ensure that a potential RMP will be accepted in practice and can be successfully implemented.

The draft Concept Paper states that submissions to FDA to revise the package insert for adverse events would not automatically lead to an RMP (123-4). We suggest adding that it would be anticipated that a package insert change for an adverse event would result in an RMP only in *rare* situations, as was conveyed at the Public Workshop.

3. Examples of Current Risk Management Tools

We encourage FDA to build a resource on its website (lines 179-180) that describes tools that are currently in use and how they have been evaluated. This is critical to enable sponsors to build on the vast experience base that already exists, but may not be well publicized or sufficiently documented. Access to such a resource would also serve to increase the probability of success for new RMPs. It would also be important to include a review that shows outcomes of the RMPs and how these programs have been evaluated for ability to achieve their intended objectives. We consider this a critical aspect of guidance, as it would provide an evidenced-based rationale for the initiative proposed in the draft Concept Paper.

4. Categories of Risk Management Programs (Risk Minimization Programs)

We feel that a categorization of risk management plans (lines 244-260) based on levels (and choice of tools) does not actually help in defining this process. The characteristics of the risk that is being managed, its severity, frequency, and ability to be recognized early and prevented, as well as other criteria, should be the primary determinants of the choice of tool(s). Labeling the program with a precise "level" number may serve to stigmatize the medicine and have the unintentional effect of steering prescribers to other, perhaps riskier medicines, a result that may not be in the best interest of the patient.

The FDA has not addressed how it will ensure that new products are not held to a different and more rigorous standard than older products used in a similar therapeutic area. If newer products are held to a higher standard in terms of risk minimization than older products with similar or greater risks, we believe that patients may be unintentionally and inappropriately exposed to products associated with a higher risk because of the ease with which they may be prescribed. We believe the Agency must ensure that it conducts a full risk analysis of older products when an RMP is indicated for a new product used for a similar indication. The sponsor of the new product usually does not have access to complete information on competitive products that would enable the sponsor to do such a comparative assessment (lines 426-34).

5. Evaluation of Risk Management Programs (Risk Minimization Programs)

We believe that there should be flexibility in the approach to the evaluation of RMPs (lines 315-341). We assume that the intention is not to have every package insert change be accompanied by an evaluation, but this is not actually stated in the draft Concept Paper. Similarly, it is not clearly stated whether even Level 2 programs involving a Dear Healthcare Professional letter would also require an evaluation program. We believe this should be decided on a case-by-case basis, depending on the nature of the risk.

There may not exist well-defined, validated metrics for evaluating all interventions, especially if the frequency of an adverse event is the primary outcome measure of interest. The precise quantification of rarely occurring adverse events is frequently not possible post-marketing because of all the well-known limitations inherent in existing databases. The example cited (in lines 321-2) of using the rate of hospitalization for a particular adverse event is potentially problematic, since it may not be possible to find the etiology of that event (drug-induced or other cause) in administrative databases. In the case where certain laboratory testing is required prior to prescribing a drug (lines 323-4), many administrative databases lack such data and would not be of use in performing an evaluation. This type of specific information may only be available in situations where individual patients are formally registered in a program and the prescribing physician documents compliance.

The requirement of at least two different evaluation methods that are quantitative and representative (lines 326-333) is not a realistic goal, since two validated methods may not be available. We believe that the Agency should clearly indicate that this may not be possible in all situations and, particularly, in those situations that involve rarely occurring events.

The meaning of the statement, "Consider using evaluation methods to assess if each RMP tool is performing as intended" (lines 340-1) needs to be clarified. We believe that in many, if not most, cases the evaluation will rely on surrogate health outcome measures (lines 351-2). In situations where an increased risk is associated with co-prescribing a contraindicated drug, we believe it should be possible to use administrative databases that are linked to prescribed medications to assess whether prescribers are complying with the contraindication.

When there is an urgency to communicate new safety information (lines 360-4), the pretesting of materials will not ordinarily be feasible for a drug that is already marketed. This should be acknowledged in FDA guidance. Pretesting may only be possible for products that have not actually been launched. Similarly, sponsors may be hampered from getting input from key stakeholders when there is an urgency to communicate new information.

6. Desired elements of a Risk Management Program (Risk Minimization Program)

Rather than requiring a description of the rationale for choosing a particular level over other levels (418-19), we believe that it is more important to provide the rationale for choosing a particular *tool*, which would be based on the characteristics of the risk to be minimized. We do not think it is appropriate for the Agency to include a section with specific guidance about conditions or outcomes that would lead to an RMP revision in the future, since we think that such determinations can only be made in the course of the evaluation process of the existing RMP. We believe it would be difficult, if not impossible, to predict particular conditions or outcomes (419-420) that would lead to an alternate intervention and this should be eliminated from subsequent guidance documents.

III. SPECIFIC COMMENTS

Risk Assessment of Observational Data: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment

These comments apply to the FDA draft Concept Paper titled "Risk Assessment of Observational Data: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment," dated March 3, 2003, and the Public Workshop held April 11, 2003, in Washington, DC. Comments are arranged by the topic of concern; line numbers refer to text in the draft Concept Paper.

1. Identifying Safety Signals from Postmarketing Spontaneous Reports

Data collected for a good case report should contain, when available, the age and sex of the patient, information regarding de-challenge and re-challenge, the possibility of drug interactions, and the suspected causality in addition to those elements described by the Agency in lines 79-104. These elements clearly constitute the 'ideal' case report, and it should be acknowledged that in practice most cases reported voluntarily would not contain all of this data, in part because the reporter is rarely the primary care physician. Other points the Agency may wish to consider regarding the collection of spontaneous reports include:

- Direct contact (line 81) may constitute many forms of contact (e.g., telephone, facsimile, electronic mail, face to face, etc.) and it would be helpful if the Agency clarified the meaning of direct contact;

- The need for clear criteria to assist sponsors in determining which reports require an attempt to collect additional information or further validation (lines 85-88); and

- Direct contacts regarding medication errors may lead to defensiveness and inadequate cooperation by the reporter of the case, suggesting that methods perceived as non-threatening need to be developed and tested for use in these cases.

2. Other Circumstances When Pharmacoepidemiologic Studies Are Useful (line 477, Question 6)

Conducting pharmacoepidemiologic studies prior to approval provides valuable data for understanding and characterizing post-marketing safety. Pharmacoepidemiologic studies that examine patient characteristics, patterns of drug use in the therapeutic class of interest, and the natural history of disease provide, at the very least, data on the background rates of mortality and morbidity in the potential patient population as compared to the general population. Studies can be conducted in diverse patient populations (e.g., private/public assistance insurance or varying geographical areas), which permit comparisons of disease rates, based on differences in clinical practice or access to health care. When results from these studies are available around the time of approval, these data provide a context for interpreting spontaneous post-marketing reports. Indeed, these data can be used to help improve the design of the clinical development program and can provide "real world" estimates to design post-marketing studies. Pharmacoepidemiologic studies can also be used post-approval to describe the patterns of use and characteristics of patients receiving the new drug; these studies may also be used to provide measurements of the drug's effectiveness at the population level. Finally, the Agency should also note in subsequent guidance documents that case-control and case-crossover study designs are sometimes performed using clinical trial data to address safety concerns that arise pre- or post-marketing.

3. Strengths and Limitations of Observational Pharmacoepidemiologic Studies

In the subsequent guidance document, FDA should discuss the challenges of designing and interpreting pharmacoepidemiologic studies. Observational pharmacoepidemiologic studies are not always the most appropriate method of evaluating safety signals or comparing the safety profile of different medications, especially when there are concerns of confounding by indication. Confounding by indication occurs when

the risk of an adverse event is related to the indication for medication use but not the use of the medication itself. The result, in observational studies, is a form of selection bias, where patients taking a particular medication are selected in a fashion that makes them at unequal risk of the outcome under study. As with any other form of confounding, one can, in theory, control for its effects if one can reliably measure the severity of the underlying illness but in practice this is not easily or completely done. This is especially so when a drug may have specific properties affecting the type of patient it is used for within its indication. In these cases, studies using randomization, whether experimental or observational in design, may be necessary.

Pharmacoepidemiologic studies are also limited by the calendar time required to obtain enough patient exposure to a medication to estimate rates of an outcome, whether in an automated database or in a *de novo* study. Sufficient patient exposure to a new medication may not be available for one to three years following approval, which may limit the usefulness of pharmacoepidemiologic studies for addressing immediate safety concerns. Finally, it is important to recognize that conflicting results are sometimes obtained from pharmacoepidemiologic studies, whether using different databases, or using the same database with different study designs. In some instances, it may be appropriate for the sponsor to propose a scientific committee of experts to guide the development of a study and/or to assist in interpreting results.

4. Large Automated Databases for Pharmacovigilance

Pfizer agrees with the Agency that large automated administrative, medical record or hospital-based databases are an important resource for conducting "real world" safety studies (lines 164-184). However, it is important to note that automated databases do not collect data uniformly or for research purposes. Information on important confounders (e.g., socio-demographic factors, health behaviors or over the counter medications) is only rarely available through automated databases; these databases are useful to classify drug exposure based on prescription refill data, but they do not permit accurate estimates of drug exposure when medications are used intermittently. Thus, despite the frequent use and relative ease of accessing these databases for research, it is ultimately the nature of the research question (or specific safety issue) that will guide the design, data collection method, and population/data sources used. Automated databases will not be feasible for studying many risks or safety concerns, and primary data collection methods may be more appropriate.

Finally, although one of the advantages of automated databases is a data source with large numbers of subjects, it is worth noting that the power of a study to examine rare events may still be limited. Sample size and power estimates are an essential element of study planning and should be included in a study protocol (lines 159-165). The release of interim and/or final pharmacoepidemiologic study results significantly influences patient care, sometimes with unintended consequences. The Agency may want to provide guidance on how and when data from a study are released (i.e., interim results) and whether the approach adopted should be described in the study protocol.

5. Medical Record Validation of Pharmacoepidemiologic Study Endpoints

In the draft Concept Paper, the Agency indicates that validation of diagnostic findings in claims database studies through detailed review of at least a sample of medical records is essential for all pharmacoepidemiologic studies (lines 182-184). In light of recently enacted legislation on personal medical data privacy and the existence of standard, validated outcomes in many databases, the Agency may want to consider this requirement's impact on a sponsor's ability to conduct pharmacoepidemiologic studies, particularly in diverse patient populations. In many cases, outcomes have been validated in other studies conducted prior to the proposed study (e.g., the UK General Practice Research Database (GPRD)). Additionally, it is increasingly possible to use the full range of data in automated databases to validate the occurrence of disease states or outcomes without medical record review (e.g., propensity scores, comorbidity scores, or automated review of health care utilization patterns). For example, hospitalization for biliary tract surgery is well described when both the diagnostic code found among hospital discharge diagnoses and the procedure codes associated with the evaluation and management

provided by the attending physician or surgeon are related in an administrative record. For other outcomes, administrative claims or outpatient/inpatient codes that include laboratory test results are often successfully used to characterize disease states without resorting to a medical record review. Finally, it should be noted that for some databases, medical record review is extremely difficult or impossible (e.g., Medicaid). To exclude these databases, particularly when they provide information on important populations such as the uninsured, would limit our ability to understand if the risks of a medication differ across patient populations.

6. Risk Estimates from Pharmacoepidemiologic Studies

While the relative risk (RR) obtained from cohort studies and the odds ratio (OR) obtained from case-control studies are important from a causation perspective, they are often confusing to the general public and physicians, in particular, since they do not incorporate information on the background rates of outcome within the study population. The Agency may want to consider emphasizing the importance of reporting incidence rates and/or the risk difference or attributable risk, which is the arithmetic difference between incidence rates in the exposed and the unexposed, in cohort and nested case-control studies to help physicians and patients contextualize RRs or ORs reported in pharmacoepidemiologic studies (line 165, methods for analysis).

7. Registries

In general, patient registries are most appropriate when the information needed could not be obtained from any other source (Line 471, Question 4). If selected, a registry may be a useful surveillance tool when an appropriate comparison group is selected to maximize internal validity and data collection is standardized to minimize information bias. In addition, outcomes to be studied must be clearly defined in the design stage to ensure that the study is adequately powered. It is important to note that findings from registries are difficult to interpret, due to the inherent bias and other limitations of this study design. Registries generally involve voluntary patient enrollment and typically do not have concurrent internal comparison groups. Because of this, rates estimated in registries are usually compared to rates from external populations. It would be helpful for the Agency to provide suggestions on appropriate control groups and the rationale for their selection. Strengths and limitations of historical controls and contemporaneously enrolled controls, for example, should be described as they apply to specific designs of patient registries.

8. Surveys

Surveys (lines 213-235) are useful tools to assess drug exposure, knowledge, and medication adherence because of their ability to provide extensive information from a large number of people. In general, surveys also use relatively few resources. It would be helpful if the Agency discussed surveys within the context of specific study designs, whether epidemiologic or qualitative. When considering surveys, it may be helpful for a sponsor to consider:

- Use of validated surveys and/or piloting new surveys before implementation, including whether translation and cultural validation is applicable;
- The most appropriate method of survey distribution (e.g., mail vs. telephone);
- Whether special statistical methods are needed to account for sampling methodology (e.g. multi-stage sampling) in survey design; and
- Whether it is practical, or desirable, to validate self-reported data against medical or pharmacy records.

9. Analyzing Post-marketing Spontaneous Reporting Rates

As the Agency notes in lines 267-274 and in line 281, spontaneous reporting systems should be used primarily to generate, rather than test, hypotheses because these systems rely on voluntary reporting from health professionals and consumers. This type of reporting is influenced by a range of medical and social factors that are not directly related to the actual rates of adverse events associated with a drug. These factors include the length of time a drug has been on the market, the reporting environment, the type of adverse event, the class of drug, and publicity in the mass media or professional literature. As such, the comparison of reporting rates should not be used for hypothesis testing nor should the true background incidence rate be inferred from the reporting rate. These limitations apply, since no direct or consistent relationship has been demonstrated between the reporting rate and the background incidence of an event in the population using the medication. For example, while the Agency notes that a high reporting rate may be indicative of a high incidence rate (lines 259-296), high reporting rates are frequently caused by factors that stimulate reporting. Stimulated reporting is a well-described effect that has been demonstrated across medications and drug classes (Lasagna, 1987; Rossi, Hsu and Faich, 1987; Meinzinger and Barry, 1990). In these cases, a high reporting rate would not correspond to a "true" high incidence rate in the patient population.

The Agency may want to consider two points related to its recommendations for commonly performed analyses of post-marketing spontaneous reports (lines 308-322). First, race/ethnicity is clearly an important risk factor for exploration, but it is rarely reported. When it is, it often does not directly correspond to the racial/ethnic categorization used in epidemiologic studies that are designed to determine risk. Second, analyses of lot-to-lot variation in medicine or differences between formulations are often difficult to interpret with certainty due to incomplete information. Formulation type is rarely reported in a spontaneous report, even when multiple dosage formulations and/or routes of administration are marketed for a medication. When it is possible to conduct analyses, these are usually performed on small subsets of cases for which information is available. Small subsets usually make it difficult to draw meaningful conclusions. Finally, information on lots is generally not reported, unless the reporter specifically suspects that there is a quality control issue involved.

10. Assessing Post-marketing Spontaneous Reports

In the draft Concept Paper, the Agency states that safety signals may inform us about new unlabeled adverse events, an increase in the severity of a labeled event, an increase in the frequency of a labeled event, or new interactions (lines 250-259). Post-marketing spontaneous reports do have the potential to inform us of new unlabeled events or interactions, but the Agency may want to clarify that spontaneous reports do not provide information regarding the actual frequency of an event or varying levels of severity of an event, as indicated by the Agency in lines 285-286. Thus, in general, it is more accurate to state that signals indicate the need for further investigation rather than inform us of important new conclusions.

Clearly, the extent of investigation of a potential signal will need to be decided on a case-by-case basis. For example, the first step in the evaluation of many signals (i.e., those detected from post-marketing spontaneous reports) is case-level causality assessment. In certain cases, this may be adequate to refute the signal (i.e., there may be obvious confounding by indication or the presence of important comorbidities associated with the event). Therefore, the Agency may wish to recommend that the initial evaluation of a potential signal generated from post-marketing spontaneous reports be an expert case-level clinical review/causality assessment of cases that generated the potential signal; this should occur before any other investigations or formal analyses.

Further, it is not always necessary to assemble extensive or multiple sets of data in those cases where individual case-level causality assessment demonstrates that the "signal" is due to other factors. This highlights the difficulty and potential confusion due to non-standardized nomenclature (the Agency recommends such an action in lines 347-349). Because of the confusing and variable nomenclature associated with the detection of signals and related safety phenomena, the Agency may want to consider whether the suggested mechanism of reporting signals (lines 376-381) is too broad for the actual range of

signals and related phenomena that are commonly encountered. Many signals that are generated are related to obvious confounding factors and known adverse events. It is unclear whether these should require a regulatory submission or an analysis of data from the clinical or pre-clinical development program.

Finally, in describing causality assessment, the Agency proposes three causal categories: probable, possible, and unlikely (lines 343-345). Since these definitions will have varying interpretations, the Agency should provide clear definitions of these terms. These definitions should be consistent with those already developed through the ICH and CIOMS consensus processes. In addition, other causal categories should be considered, such as definitely unrelated/drug excluded (i.e., for use in reports of drug-drug interactions when an association between drug and the adverse event is biologically or pharmacokinetically implausible.)

11. Automated Signal Detection/Data Mining Methods (Line 464, Question 2)

The development and use of signal detection methods represent a rapidly growing field in pharmacovigilance. It would be particularly beneficial if the Agency addressed the issue of imprecise and highly variable terminology that is being used to describe similar but related concepts in the field of signal detection. Examples of terms that create confusion are: "signal", "alert", "combination", "associations", etc. Regulatory authorities, health care professionals, pharmaceutical companies, and drug monitoring centers may have different views of what these terms mean and this topic should be carefully discussed in consensus forums to develop definitions for worldwide use. This is particularly important given recent developments in the application of various statistical methodologies for signal detection, such as empirical Bayesian screening. The various methodologies and even users of the same methods have varying definitions of what constitutes a signal for a given method. The joint FDA-P*h*RMA Safety Evaluation Techniques (SET) working group has prepared a document that delineates many of the unresolved issues and questions. This document is available to the Agency, including the Office of Drug Safety.

The Agency may want to recommend that sponsors consider the following when evaluating the utility of competing signal detection methods:

- The comparative performance characteristics of the methods, including analyses within a given data base as well as between smaller pharmaceutical company databases and larger regulatory databases (NB: conventions for coding with MedDRA and other aspects of data input may have important implications when comparing performance across databases);

- The need to achieve a consensus on basic, common elements of signal detection protocols/strategies, such as the choice of database, signal metric/thresholds, role of data stratification, and decisions on whether to mine data on the suspect drug alone versus the suspect plus concomitant medications. In addition, the need for consensus on global data base mining versus drug-specific data base mining versus event-specific database should be considered; and

- Circumstances when it is or is not acceptable to compare "signals" between drugs.

Finally, regardless of the methods used to detect signals or to mine data, it is important to note that these mechanisms only provide a statistical score of disproportionality. This statistical score is based on the selected model of statistical independence between drug and event. Pfizer feels it is important for the Agency to emphasize in the subsequent guidance document that signal detection/data mining methods are preliminary and they merely generate hypotheses rather than test them; none of the currently available signal detection/data mining methods necessarily have any correlation with clinical reality.