

18 North Iroquois Lane
Chester Springs, PA 19425

5271 '02 MAR 12 09:30

March 8, 2002

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

Dear Sir/ Madam:

Enclosed find my comments to the draft guidance entitled: "Integration of Study Results to Assess Concerns About Human Reproductive and Developmental Toxicities" (Docket No. 99N-2079). These are my personal comments only and do not include input from any other individual, organization or company.

Sincerely yours,



Thomas A. Marks, Ph.D., D.A.B.T.

99N-2079

C10

As written, this guidance is likely to result in even more ambiguity in labels/package inserts. The document is heavily weighted toward declaring that the animal study(ies) conducted on a therapeutic are not adequate to evaluate the risk for adverse reproductive effects in humans. The present situation of assigning a Pregnancy Category to a therapeutic is bad enough. More often than not, therapeutics are assigned a Pregnancy Category C, which indicates that the animal developmental and reproductive toxicology (DART) studies performed are not adequate to assess risk, even though the studies may have been negative and were carried out according to regulatory guidelines. This guidance will likely result in a significant increase in the number of studies judged to be inadequate, a conclusion likely to be inserted into the label, if the therapeutic in question is approved by the FDA. Since such information may be of little use to physicians, and especially the public, little is likely to be gained.

Although one can conclude that therapeutics found to affect human development and/or reproduction have also been found to affect development and reproduction in animal models, often such effects have been discovered in humans first. The classic example is thalidomide. Another example of a human teratogen is 1,3-cis-retinoic acid (Accutane), which, like vitamin A, was found to be teratogenic in animal models. However, Accutane was identified as a human teratogen after FDA approval, whereas evidence that therapeutic doses of vitamin A are teratogenic in humans is lacking. In spite of such a poor track record, the perception is that therapeutics are a major risk to human development and/or reproduction, resulting in the term "therapeutic orphan" for women because of the reluctance of physicians to prescribe, and women to take, medication during pregnancy (Shirkey, H. [1968] *J. Pediatrics* 72:119-120). In turn, such fears often result in unnecessary therapeutic abortions and/or unfounded legal actions (Czeizel, A and Racz, J [1990] *Teratology* 42:505-512).

In the ninth edition of *The Catalog of Teratogenic Agents* (T.H. Shepard, The Johns Hopkins University Press, Baltimore, 1998, page XV), it states: "There are more than 3000 agents in this catalog. About 1,200 can produce congenital anomalies in experimental animals, but only about forty of these are known to cause defects in humans. Therefore, there exists a wide difference between our knowledge of experimental teratology and the role that external agents play in producing human malformations." Clearly the use of animal models has, historically, led to the identification of hundreds of agents that are a threat to embryonic/fetal development in one or more animal model, but which have not been found to be a threat to the human conceptus. One reason for this is that humans are not usually exposed to such agents, with the likely exception of cancer chemotherapeutic agents, at levels approaching those used in animal studies (e.g., maximum tolerated dose). The end result is that it is difficult for the medical community, let alone pregnant women, to assess the risk of a therapeutic to reproductive outcome.

It seems apparent on reading this guidance that reviewers are likely to give weight to every positive finding; e.g., ignoring a lack of a dose-related increase in incidence, and/or the absence of statistically significant differences from the control, and/or placing too much emphasis on increases in variations (minor anomalies), especially skeletal, as doses approach lethal levels. Very often the maximum tolerated dose (MTD) is very close to a lethal level and sometimes results in one or more parental deaths. One would expect developing embryos/fetuses to be more susceptible than the pregnant animal to the effects of a therapeutic at doses approaching the MTD. In fact, minor, especially skeletal, anomalies or variations are often observed at such doses. Thus, one should expect to see increases in such variations over what is found in the control concepti. To conclude that such effects are drug related may be true, but hardly reason for much concern. If such incidences are deemed worthy of being mentioned in the label for an approved therapeutic, such information is only likely to confuse the reader while being of little use in accessing risk. Worse, if it becomes apparent that this is likely, it probably will be perceived by the developers of a therapeutic as reason to repeat DART studies and/or use other species (animal wastage) that are even less likely to be predictive of human effects, with little or no increase in information of value. Since little historical information in the DART areas is available for species other than the rat and rabbit, the use of alternative models could further confuse things and lead to the conclusion that studies were inadequate.

The present requirement for ADME/TK studies is another factor likely to increase the number of studies judged to be inadequate, leading to efforts/requirements that alternate species be tested. Also, the emphasis on employing the clinical route in DART studies increases the likelihood of an inadequate study. The s.c. route often has been used, at least in past years, in place of the i.v. route in DART studies, as it was considered impractical to dose 96 or more (144-192 in a Study of Fertility and Early Embryonic Development to Implantation study) rats, daily for several weeks by the latter route. The stress on the animals, associated with i.v. treatment, can be expected to compound the effects of the drug. To increase the number of daily treatments when testing a drug with short plasma half-life may further influence the results and likely increase the number of technical errors. The fact that such studies are often completed, or are at least in progress, before human PK information is available can be expected to result in even more inadequate studies. Thus, this guidance, as written, is likely to increase the number of studies judged not to be adequate to evaluate the risk for adverse human reproductive effects. Such a conclusion in itself may be no worse than assigning a Category C to the therapeutic. However, if it means inserting into a label a lot of information of little use to physicians, and especially the public who have little understanding as to how to interpret such information, little has been gained.

Reproductive and Developmental toxicologists are well aware of the poor track record for predicting human problems with animal DART studies. Obtaining

adverse effects in DART studies has not been difficult. Increasing the likelihood of judging such studies inadequate most likely will result in an increase in the number of studies repeated, the amount of material placed in the label, the degree of confusion in the medical community and with the public, with little likelihood of our being better able to predict human risk. Although hundreds of reproductive hazards have been identified in animals, a few of which have proved to be human hazards as well, what value is there in increasing the likelihood that the DART studies are inadequate? Why not just summarize the important findings, as well as any limitations, without implying that the studies may have no value by declaring them to be inadequate.

Because of the poor track record for predicting the risk of a therapeutic to human development and/or reproduction, one can conclude that, historically, DART studies have not been adequate. Far more false positives have been identified in DART studies than those that have been found to adversely affect human development and/or reproduction. It has been more than 40 years since thalidomide was identified as a human teratogen, an event that led to worldwide changes in the way toxicology studies are performed. Regulatory requirements have steadily increased over the years, one result of which has been the identification of hundreds of agents that produced positive findings in DART studies. The problem is not being able to identify potential threats to human development and development, but rather to sort out the false positives and identify those likely to be a real threat. Being unable to do this has meant that the human is still the ultimate test animal for such effects. Thus, it is an academic exercise to find more and more ways for us to judge that animal DART studies may not be adequate to evaluate the risk for adverse human reproductive effects.

Although this guidance may be an effort to put FDA reviewers on the same plane when reviewing DART studies, toxicologists responsible for such testing surely will also use it as they continue to make every effort to perform adequate studies. However, the increased likelihood that an FDA reviewer will conclude that one or more of such studies are inadequate can only result in increased efforts by the pharmaceutical company toxicologist to meet the requirements of the guidance. Pharmaceutical company toxicologists generally have years of experience in performing DART studies and can be expected to use such knowledge and experience in evaluating such studies. However, he/she is likely to recommend further study if, on following the flowcharts that are the backbone of this guidance, he/she concludes that one or more of DART studies do not follow a path likely to result in a judgment of no known concern. It would be nice if such additional studies will increase the likelihood of successfully predicting human risk. However, often a repeated study will generate more questions than answers. Worse, the development of potentially worthwhile therapeutics may be discontinued because of the increased costs and delays of performing additional studies, especially if our rat and/or rabbit models appear to be inconsistent with findings in humans, resulting in the perceived need for additional studies in

animals (e.g., monkeys), which may be of little additional value in predicting risk to humans.

More specific concerns about this guidance follow. In section III.A.2 (lines 210-230), the term 'relevant' is used. Routinely, the rat has been employed in DART studies, along with the rabbit in one of the developmental studies. Thus, one assumes that such models are relevant. However, since DART studies are generally completed or ongoing before ADME/PK data in humans are available, the questions in Section III.B.2 (lines 287-305) can only be answered on performing human clinical trials. If the answer to one or more of the questions is no, then the conclusion is that the study(ies) may not be adequate. To so conclude is no worse than being placed in Category C in the label. Thus, it should be made clear somewhere in the guidance that a 'no' answer does not mean that one must continue performing more tests in order to obtain evidence that the studies are adequate.

How does one determine whether or not the exposures were significantly greater than those demonstrated in humans at the maximum recommended human dose (Section III.B.2)? It is generally understood that the highest dose tested in a DART study is the MTD. Thus, one could test a drug at the MTD and not get significantly greater exposures, especially with an oncology drug. Also, to successfully carry out DART studies, one needs surviving conceptuses/offspring. Thus, one may have to lower the dose below human exposure levels. Since such a study appears to be inadequate, is there any point in doing it? This situation was not addressed in the guidance.

The use of the phrase "rare events", is a major flaw in this guidance. On the one hand, the guidance states (lines 470-471): "If the positive signal occurs only during processes that are of limited relevance to humans (rare), there would be less concern for adverse human reproductive outcomes." Later on (lines 537-538) it states: "Thus, an increased frequency of positive signals for rare events in drug-exposed animals increases concern for reproductive or developmental toxicity in humans." Clearly, the use of the word "rare" needs to be clarified. The guidance does not specify whether one specific rare event is involved or whether two or more rare events constitute a problem. Thus, in any study, one can expect to identify a multitude of events, especially minor skeletal anomalies (variations) at the MTD. Such anomalies are likely to occur infrequently in the control group. Thus, the frequency in one's historical control may be low, leading one to conclude that they are rare events. Since the MTD is often close to a lethal dose, the presence of such minor anomalies should not be judged to be rare events, especially if they only appear at the MTD. The embryo/fetus can be expected to be more susceptible to adverse effects of a therapeutic than the pregnant animal, especially at doses that approach lethal levels. Since such anomalies can be expected at the MTD, this fact should be considered when assigning a signal. Also, undue significance should not be assigned to events that occur only once, even though several anomalies may occur in the same

fetus, and even if one or more of these events are considered to be rare. Another fetus in the same litter, or a fetus from another litter, even from another dose group, may experience anomalies that are different from the above fetus. One could conclude that such effects represent a positive signal, even though there is no consistency between the anomalies found. Known teratogens can be expected to cause the same anomalies in different fetuses. Thus, different anomalies in different fetuses are more likely to have occurred by chance than to be drug related. In short, unless one is familiar with what is considered important in evaluating a DART study, one can assign undue importance to what may be considered to be a rare event. In the absence of a dose response and biological significance, one should not consider the presence of “rare” events to be a positive signal, especially if they occur only at the MTD.

The use of the $TI_{10/90}$ ratio (lines 555-587) appears arbitrary and of questionable value in assigning a signal. Obviously, the $TI_{10/90}$ for an oncology drug will often be low and handled differently than therapeutics with other indications. Also, it will often be the case that different species will be used to assess efficacy than to assess toxicity. In DART studies, treatment is usually given over relatively long periods (e.g., organogenesis or fetal period to weaning or prenatally to implantation), thereby often resulting in a lower MTD than if the drug was given over shorter periods, while, clinically, the therapeutic often is given for shorter periods (e.g., the MTD in an acute study is generally higher than in a chronic study). Thus, the TD10 might be expected to be low and the ED90 higher than would be the case if the situation were closer to what would be expected if the test circumstances were more closely associated with what would be the case on human exposure to the therapeutic.

Although the relative exposures ratio (lines 669-700) also employs a relevant metric, in order to show decreased concern a ratio ≥ 25 must be obtained. This may be difficult to achieve especially considering the low levels likely to be obtained in a DART study at the NOAEL. Thus, both the TI and relative exposure ratios are more likely to result in values indicating increased concern, making it unlikely that the decreased concern designation would be assigned.

In summary, this draft guidance provides a rather inflexible step-by-step approach to the evaluation of findings from developmental and reproductive toxicology studies for purposes of identifying (categorizing), but not managing potential human risk. It appears heavily weighted toward assigning positive signals. The end result likely will result in evaluations indicating that the animal studies conducted are inadequate to fully assess the risk for adverse human reproductive effects. Such findings are consistent with the present situation where most therapeutics approved by the FDA are given a designation of Category C. However, the guidance is also likely to result in few therapeutics being judged to be of “no known concern.”

Because DART studies, historically, have generated hundreds of apparent “false positives,” it is likely that employment of this guidance will generate even more findings of “low” to “significant concern” for future therapeutics. It would be comforting to conclude that such diligence will result in reduced incidences of ill effects in the areas of human reproduction and development. However, it seems more likely that using this guidance will result in the inclusion, in the labels of approved therapeutics, of information that will result in even more reluctance on the part of physicians to prescribe a therapeutic to a woman of child-bearing age, especially if she is pregnant. Such information also is likely to further increase the level of concern in a pregnant woman, a situation that already is apparent (Koren, G., et al., [1989] *Am. J. Obstet. Gynecol.* 160:1190-1194). Such concern frequently results in a pregnant woman requesting a therapeutic abortion, and even situations where a physician will recommend such a procedure. The latter individual likely is aware that at least 3% of children have identifiable malformations at birth. Thus, even if evidence of an association in humans has not been shown, the presence of adverse information in the label of a therapeutic may be enough to precipitate litigation.

At first glance, it seems like a good idea to issue a guidance putting investigators and FDA reviewers on common ground in reviewing DART studies. However, as written, this guidance will likely only serve to further point out the deficiencies in such studies while increasing the level of concern by increasing the likelihood that a finding(s) will be judged as positive. Worse, it probably will increase the number of studies that are judged to be inadequate, resulting in the additional investigations in the same and/or another species. This will further result in delays in drug development, animal wastage, and the increased likelihood that such development will be discontinued because of increased costs, in addition to unnecessary concerns about safety and/or possible litigation. We are probably better off if we recognize that, inherently, our animal models have flaws when it comes to predicting human reproductive risks. This guidance is likely to provide further evidence that this is the case, without increasing our ability to assess human risk.

M
18 N Iroquois Ln
Chester Springs PA 19425

WILMINGTON, DE 19850

MAR

8

2002



Pockets Management Branch

HFA-305

Food and Drug Administration

5630 Fishers Lane

Rockville, MD

20855