Supporting Documents for Frank Sasinowski's May 20, 2014 Testimony Before House E&C Health Subcommittee

Appendix 1:
Proposed Advisory Committee Chart: Substantial Evidence of Effectiveness.

Appendix 2:
E-Filing on Regulations.Gov by Frank Sasinowski, Director, and Alexander VarondHyman, Phelps & McNamara, P.C. re: Docket No. FDA-2013-D-0575, Comment on Section VII. C.: "Evidentiary Criteria for Accelerated Approval" of the FDA "Draft Guidance for Industry: Expedited Programs for Serious Conditions—Drugs and Biologics" (Aug. 26, 2013).

Appendix 3:
King et al., *A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis*, N. Eng. J. Med., NEJM.org DOI: 10.1056/NEJNMoa1402582 (May 18, 2014).

Appendix 4:
Memorandum re Iloprost, NDA 21-779, Cotherix, Inc. from Robert Temple, Director, Office of Drug Evaluation I, CDER, FDA (Dec. 27, 2004).

Appendix 5:
Approvable Letter from Robert Temple, Office Director, Office of Drug Evaluation I, CDER, FDA, to Benjamin Lewis, Senior Director, Regulatory Affairs, Prestwick Pharmaceuticals, In. (Mar. 24, 2006).

Appendix 6:
Carl Peck & Jill Wechsler, *Report of a Workshop on Confirmatory Evidence to Support a Single Clinical Trial as a Basis for New Drug Approval*, 36 Drug Information Journal 517 (2002).

Appendix 7:
Frank Sasinowski, *Quantum of Effectiveness Evidence in FDA's Approval of Orphan Drugs: Cataloging FDA's Flexibility in Regulating Therapies for Persons with Rare Disorders*, 46 Drug Information Journal 238 (Mar. 2012).

Appendix 1: Proposed Advisory Committee Chart

Substantial Evidence of Effectiveness		
Quantum of Effectiveness Evidence Needed		**Source of Authority**
1	Two Adequate and Well-Controlled Studies	21 U.S.C. § 355(d) [1]
2	One Adequate and Well-Controlled Study with "Confirmatory Evidence"	21 U.S.C. § 355(d) as amended by FDAMA 115 [2]
3	One Study Providing Statistically Very Persuasive Evidence and Where a Second Study Would be Difficult to Conduct on Practical or Ethical Grounds	May 1998 Guidance [3]
Types of Therapies in which FDA Has Exercised Flexibility		
A	Accelerated Approval/Subpart H/Fast Track Therapies	Historical FDA Precedents [4]
B	Orphan Drug Therapies	Historical FDA Precedents [5]

1. Federal Food, Drug, and Cosmetic (FDC) Act § 505(d)
2. FDA Modernization Act § 115
3. FDA Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products (May 1998)
4. FDC Act § 506
5. FDC Act § 526

LAW OFFICES
HYMAN, PHELPS & MCNAMARA, P.C.
700 THIRTEENTH STREET, N.W.
SUITE 1200
WASHINGTON, D.C. 20005-5929
(202) 737-5600
FACSIMILE
(202) 737-9329
www.hpm.com

FRANK J. SASINOWSKI
Direct Dial (202) 737-4287
FSasinowski@hpm.com

August 26, 2013

BY E-FILING ON REGULATIONS.GOV

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

Re: Docket No. FDA-2013-D-0575
Comment on Section VII. C.: "Evidentiary Criteria for Accelerated Approval" of the FDA "Draft Guidance for Industry: Expedited Programs for Serious Conditions—Drugs and Biologics" (hereinafter, "Draft Guidance")

Dear Sir/Madam:

These comments are based on an analysis of FDA's Subpart H approvals from the 1992 promulgation of the Subpart H regulations to the present.

In Section VII. C. of FDA's June 2013 Draft Guidance, the Agency describes several factors that FDA weighs in assessing whether the available evidence is sufficient to allow FDA to conclude that the proposed surrogate is "reasonably likely to predict clinical benefit"[1] and thereby constitute the basis for a Subpart H[2] marketing approval.

---

1 In these comments, when the term "surrogate" is used, it is meant also to encompass what FDA in its Draft Guidance refers to as "a clinical endpoint that can be measured earlier than irreversible morbidity or mortality (IMM) that is reasonably likely to predict an effect on IMM or other clinical benefit." Draft Guidance at p.16, lines 511-513.

2 In these comments, Subpart H will be the short-hand term used interchangeably with 21 C.F.R. Part 314, Subpart H; 21 C.F.R. Part 601, Subpart E; "Accelerated Approval" and "Fast Track."

Subpart H authority has existed for well over 20 years. FDA created it on its own regulatory ingenuity to address the AIDS epidemic. However, the importance of Subpart H as a regulatory innovation and vehicle for providing patients suffering with serious and often rare diseases where there is inadequate available therapy has recently taken on significant added importance. Two milestone events within approximately the past year illustrate this.

1. In the FDA Safety and Innovation Act (FDASIA) of July 2012, Congress and President Obama revised the statutory provisions of Subpart H[3] to, as FDA in its Draft Guidance states, "facilitate somewhat broader use of accelerated approval to expedite patient access to important treatments for serious conditions[,] . . . provide additional flexibility[,] . . . provide clarification concerning the use of clinical endpoints[,] . . . [and] make clear that FDA has the authority to consider pharmacologic or other evidence . . . in determining whether an endpoint is reasonably likely to predict clinical benefit." Draft Guidance at p. 14, lines 445-453. While these were added in July 2012 by statute, this analysis establishes that here, as is often the case, Congress is merely codifying in statute the practices and policies that FDA had already put into place and acted upon previously. In the text of FDASIA, Congress however directed that FDA expand its use of this authority.

> "FDA should be encouraged to implement more broadly effective processes for the expedited development and review of innovative new medicines intended to address unmet medical needs for serious or life-threatening diseases or conditions, including those for rare diseases or conditions, using a broad range of surrogate or clinical endpoints . . . This may result in fewer, smaller, or shorter clinical trials for the intended patient population or targeted subpopulation without compromising or altering the high standards of the FDA for the approval of drugs. Patients benefit from expedited access to safe and effective innovative therapies to treat unmet medical needs for serious or life-threatening diseases or conditions. For these reasons, the statutory authority in effect on the day before the date of enactment of this Act governing expedited approval of drugs for serious or life-threatening diseases or conditions should be amended in order to enhance the authority of the FDA to consider appropriate scientific data, methods, and tools, and to expedite development and access to novel treatments for patients with a broad range of serious or life-threatening diseases or conditions.

---

3 See Footnote #2.

SENSE OF CONGRESS.—It is the sense of Congress that the Food and Drug Administration should apply the accelerated approval and fast track provisions set forth in section 506 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 356), as amended by this section, to help expedite the development and availability to patients of treatments for serious or life-threatening diseases or conditions while maintaining safety and effectiveness standards for such treatments."

2. In September 2012, President Obama became the first President to comprehensively address the complexities of developing new medicines for Americans when he released his report, "Propelling Innovation in Drug Discovery, Development, and Evaluation." In that report, FDA is instructed to expand the use of its Subpart H authority. See President's Council of Advisors on Science and Technology Report to the President on Propelling Innovation in Drug Discovery, Development, and Evaluation. ("Presidential Report") at pp. 59-68. Specifically, this Report recommended that:

- "The FDA should expand the scope of acceptable endpoints used to approve drugs for serious or life-threatening diseases with unmet needs. Under current law, the FDA has considerable discretion in deciding whether a surrogate or intermediate clinical endpoint is 'reasonably likely, based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict' clinical benefit. At one extreme, the FDA might be highly risk-averse, requiring near-certainty that the surrogate or intermediate endpoint will translate to clinical benefit. At the other extreme, the Agency might accept endpoints that are simply correlated with disease outcome or plausibly related to disease outcome based on current scientific understanding. Neither extreme would serve the public well. The FDA's interpretation of 'reasonably likely . . . to predict' can have a major impact on the pace of medical innovation and on patient safety . . . Historically, the use of [Subpart H] has been primarily used in a limited number of therapeutic areas—principally, HIV/AIDS, cancer, and inhalation anthrax (87 percent of cases) . . . We believe that the Nation would benefit if the FDA were to expand the use in practice of acceptable indicators to other serious or life-threatening diseases." (Presidential Report at p. 59).

- "Recommendation 3: Expand the Use in Practice of FDA's Existing Authorities for Accelerated Approval and Confirmatory Evidence. The FDA should make fuller use of authorities previously granted by legislation and not yet fully utilized. The FDA should expand the use in practice of its existing authority for Accelerated Approval. FDA should direct its staff, across all divisions, to make full use of the Accelerated Approval track for all drugs meeting the statutory standard of addressing an unmet medical need for a serious or life threatening

illness and demonstrating an effect on a clinical endpoint (other than survival or irreversible morbidity) or on a surrogate endpoint that is reasonably likely to predict clinical benefit." (Presidential Report at p. 61).

Given this renewed recognition of the promise of FDA's Subpart H authority to address those suffering from serious diseases without adequate available therapy, and given FDA's issuance of this Draft Guidance addressing the Agency's Subpart H authority, an analysis of FDA precedents in order to promote a better understanding of the circumstances under which Subpart H may be employed may be both timely and productive. In this way, it is hoped that the regulatory ingenuity of FDA in creating Subpart H and the recent Congressional and Executive exhortation to more fully mobilize this Subpart H power may find expression.[4]

METHODS

First, the FDA Draft Guidance in several places cites to the Subpart H precedents in AIDS and cancer, and there is little regulatory uncertainty as to the evidentiary criteria for a surrogate to be the basis for marketing approval in either of these two therapeutic areas. Therefore, this analysis is of the 19 Subpart H approvals identified by FDA on its website[5] that are for conditions other than AIDS or cancer.

Second, to maximize the usefulness of this analysis as a comment on the Draft Guidance, this analysis of each of these 19 precedents is organized according to the order of factors cited by FDA in Section VII. C., "Evidentiary Criteria for Accelerated Approval" of the Draft Guidance. Organizing this analysis according to the order of the

---

4 One of the two commentators here conducted an analysis of FDA orphan drug precedents that has, to some, proved of some utility (Frank Sasinowski, Quantum of Effectiveness Evidence in FDA's Approval of Orphan Drugs, 46(2) Drug Inf. J. 238-263 (Mar. 2012)); this analysis of Subpart H precedents, it is hoped, may prove to be of like usefulness.

5 This website is current up to September 2011 and the two commentators have supplemented it to include Subpart H approvals since September 2011. See "CDER Drug and Biologic Accelerated Approvals," available at http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/HowDrugsareDevclopedandApproved/DrugandBiologicApprovalReports/UCM278506.pdf.

factors listed in the Draft Guidance has the added benefit of providing a logical structure for this analysis.

- Part 1 of Each Analysis: Regulatory Factors Weighing into FDA Determination - Severity and Rarity of the Condition, Availability of Alternative Treatments, and External Expertise.

Under Section VII. C. 1., "Whether an Endpoint Is 'Reasonably Likely to Predict' Clinical Benefit," FDA acknowledges that "[w]hether a drug effect on a given endpoint is reasonably likely to predict clinical benefit is a matter of judgment" and then, FDA explains that the Agency "considers all relevant evidence and weighs the uncertainty [of the evidence, presumably] against the severity of the disease to be treated and the lack of available therapy. On a case-by-case basis, FDA will make informed judgments using both internal and external expertise." Draft Guidance at p. 18, lines 609-612.

This FDA statement in its Draft Guidance generally tracks what was inserted by FDASIA into the statutory authority for Subpart H. Specifically, FDASIA amended the Federal Food, Drug, and Cosmetic (FDC) Act to provide FDA with the authority to approve a therapy under accelerated approval when FDA determines "that the product has an effect on a surrogate . . . that is reasonably likely to predict clinical benefit . . . taking into account the severity, rarity, or prevalence of the condition and the availability . . . of alternative treatments." FDC Act §506(c)(1)(A) as amended by FDASIA §901 (emphasis added). Comparing the FDASIA text of July 2012 with the Draft Guidance statement from June 2013, there is one conspicuous incongruity, and it has huge implications for the 30 million Americans with rare disorders and their families and friends. Noticeably absent from the Draft Guidance statement is the over year-old statutory requirement that FDA must take into account, in addition to severity of the disease and availability of alternative treatments, whether a condition is rare.[6] Therefore, in the first part of the analysis of each of these 19 Subpart H approvals, consideration is given to each of these factors: the severity of the disease, its rarity, and whether alternative treatments exist. These three factors are, by statute, required to be taken into account by FDA in determining whether to grant Subpart H approval.

---

[6] To one of the commentators, who has devoted a career, both at FDA and since FDA, to aiding in the development of therapies for our brothers and sisters with rare conditions, it is impossible to overstate the degree of his apoplexy over this oversight in the Draft Guidance.

The first part of each analysis, again tracking the FDA Draft Guidance, also accounts for whether there is any evidence that FDA considered "external expertise" (which most often would have been by seeking the expert input of an FDA Advisory Committee on that therapy). Draft Guidance at p. 18, lines 611-612.

In the aggregate, we refer to this first set of factors in FDA's Draft Guidance as "Regulatory Factors Weighing into the FDA Determination" because each of these four factors is a regulatory decision by FDA: that is, whether the condition is serious, whether it is rare or an "orphan," whether "available therapy" exists, and whether to seek the input of an advisory committee.

- Part 2 of Each Analysis: Understanding of the Disease[7]

In Section VII. C. 1. a., of the Draft Guidance, "Understanding of the disease process," FDA explains the criticality of understanding the disease process as fundamental to achieving the "biological plausibility"[8] of the surrogate. Draft Guidance at pp. 18-19, lines 617-648. Therefore, the second part of the analysis of each precedent is the degree to which the underlying disease is understood.

- Part 3 of Each Analysis: Understanding of the Relationship between Drug Effect and Disease Process

Under Section VII. C. 1. b., "Understanding of the relationship between the drug's effect and the disease process," the Agency notes that "[t]he extent to which a drug's effect on the surrogate endpoint is known to predict an effect on the disease is critical."

---

[7] Parts 2-4 not only track the order in the Draft Guidance, but also are tied to the language quoted in FDASIA: "[t]he evidence to support that an endpoint is reasonably likely to predict clinical benefit . . . may include epidemiological, pathophysiological, therapeutic, pharmacologic, or other evidence . . . ." FDASIA § 901(a). Specifically, Part 2 relates to "pathophysiological" evidence; Part 3 relates to "epidemiological, . . . pharmacologic, or other evidence;" and Part 4 relates to "therapeutic evidence."

[8] This term is taken from a paper by FDA officials, Drs. Desai, Stockbridge and Temple, Blood Pressure as an Example of a Biomaker that Functions as a Surrogate, 8(1) AAPS J. E146-E152 (2006) ("[B]iological plausibility [is] sometimes intuitive, sometimes supported by animal data or by favorable response in extreme cases (e.g., malignant hypertension).").

Draft Guidance at p. 19, lines 653-654.[9] FDA, in its Draft Guidance, then lists several factors to consider in identifying and assessing a surrogate endpoint, including, "[w]hether there is reliable and consistent epidemiologic evidence supporting the relationship between the endpoint and the intended clinical benefit" and "[w]hether the effect on the [surrogate] endpoint has been shown to predict a clinical benefit with drugs in the same or closely related pharmacological class." Draft Guidance at pp. 19-20, lines 662-675. Therefore, the third part of the analysis of each precedent assesses the evidence for these factors, noting that, for the purposes of this analysis, epidemiological evidence is interpreted more broadly to include all observational studies, including long-term longitudinal studies and "natural history" studies. This part of each analysis essentially assesses the predictive potential of the surrogate.

- Part 4 of Each Analysis: Clinical Evidence for the Surrogate and for the Clinical Benefit

In the Draft Guidance, FDA acknowledges, as noted earlier in this comment, the primacy of clinical evidence of the drug itself, both on the surrogate and on the clinical benefit, but explains, "[h]owever, this guidance does not address clinical evidence requirements because they are not readily generalizable." Draft Guidance at p. 18, lines 614-615. Our analysis has the luxury of not needing to distill general requirements from the many precedents, but can assess the strength of the clinical evidence in each case for each drug's effect, both on that specific surrogate and on that particular clinical benefit. Accordingly, the fourth and final part of the analysis of each precedent is the strength of clinical evidence on the surrogate itself, as well as on the clinical benefit.

Lastly, with respect to the methods employed for these analyses, a word on the weights given to each of the factors in these four parts of each analysis: these weights themselves are a matter of judgment, as are each of the assessments or "scores." Other individuals may prefer either greater or lesser weights for any of these factors, and may even decide that some of these factors should not be included at all or still others be added. Similarly, others – especially the experts in the medical community, Sponsors

---

[9] FDA in its Draft Guidance continues (and in so doing helps to explain the relative value of what in this analysis we have divided into Parts 3 and 4): "Sometimes this relationship can be assessed epidemiologically but it is most persuasively established by knowing that a drug that affects the surrogate also affects a clinical outcome." Draft Guidance at p. 19, lines 653-656. In this analysis, we, accordingly, weight more heavily Part 4 (Clinical Evidence) as compared to Part 3 (which is, in part, epidemiology).

and, especially the FDA reviewers and supervisory officials – may disagree with the scores given to any factor or factors for any of these 19 Subpart H precedent approvals. All of these views would be fair, especially when based on a more thorough understanding of the science or evidence, and are understandable.

RESULTS

All available FDA source documents were gathered and analyzed for each of these 19 Subpart H approval precedents in order to "score" each according to the factors laid out in the FDA Draft Guidance according to the weights and scoring of the commentators. Figure 1[10] is a chart summarizing these, and in Appendix 1 there is a narrative text that describes some of the most relevant information pertinent to each of the FDA Draft Guidance factors for each of these Subpart H approvals.

DISCUSSION

Regulatory ingenuity, if not outright genius, led FDA on its own to create the concept of the Subpart H approval in order to address at first, the emerging AIDS epidemic in the 1980s and since then, all other serious conditions for which there is an unmet medical need. The linchpin of the FDA Subpart H system was, and is, the surrogate endpoint that is "reasonably likely to predict clinical benefit" (or intermediate clinical endpoint that is "reasonably likely to predict ultimate clinical benefit.")

There have been many misunderstandings of this Subpart H system. Some have thought that it meant that the quantum or quality of evidence was somehow reduced, and the statutory requirement of "standard evidence of effectiveness" was in some way, in whole or in part, skirted or deferred. While this seems not to be the case in statute, regulation or policy, the other extreme is just as likely not to "serve the public well" (quoting the Presidential Report at p. 59). The other extreme is the view that unless the surrogate is validated, it cannot be relied upon in a Subpart H approval decision. This is sometimes found in reviews that conclude that the Sponsor's evidence failed to satisfy the standard of approval because the trial(s) attempted to prove both the drug's effect on the surrogate as well as on the clinical benefit and the clinical benefit showing was not robust enough to validate the drug's effect on the surrogate.

---

[10] The drugs in Figure 1 are listed chronologically, from the most recent Subpart H approval, Sirturo, to the earliest, Betaseron.

Between these two extremes, there has existed a gaping hole that has begged to be addressed for nearly three decades and that is – what is the regulatory and evidentiary foundation for FDA's determination that an unvalidated surrogate is capable of supporting a Subpart H approval? Now FDA in its June 2013 Draft Guidance has tackled this and laid out clearly discreet principles and factors.

This analysis (that is, this comment on the Draft Guidance) attempts to apply those principles and factors to the 19 Subpart H approvals (that are not for AIDS or cancer) in order to discern, in that analytical process, the types and patterns of the evidence that FDA has found adequate to be the foundation for these Subpart H precedents.

Let's see what this analysis can tell us, and then, see if those findings can further our understanding, both of Subpart H in general and of when it may be applicable going forward.

- Part 1

Part 1 of the analysis of each precedent assesses the first set of factors that FDA describes on lines 609-612 of the Draft Guidance: severity of disease, lack of available therapy and external expertise (as well as, yes, rarity too). For each precedent, we present the assessment of these factors under the heading of "Part 1" in Figure 1 and in the narratives for each precedent in the Appendix. The consistency of findings across the 19 precedents with respect to these Part 1 factors is highly robust. In its Draft Guidance, FDA explains that it "weighs the uncertainty" of "all relevant evidence" against these Part 1 factors. Trying to predict whether any surrogate will indeed "reasonably predict" clinical benefit will never be an absolute certainty, and so there will likely be at least some residue of uncertainty in each case. This analysis confirms what some may have forecasted, which is that a strong showing in these regulatory Part 1 factors is nearly a prerequisite for qualifying for Subpart H consideration.

- Part 2

Understanding of the disease process is the next key factor listed by FDA in the Draft Guidance (lines 617-648). Part 2 of the analysis of each precedent describes our assessment of this factor for that therapy. For 12 of the 19 precedents, a maximum score of 3 was achieved. This is consistent with FDA's view stated in the Draft Guidance that this can be "an important factor in determining whether an endpoint is reasonably likely to predict clinical benefit." Draft Guidance at lines 631-632.

However, three precedents (Makena for pre-term birth, Luveris for pregnancy, and Remicade for Crohn's Disease) received scores of "1" on a scale of 0 to 3 because in each case the pathophysiology of the underlying disease is complex and not so clearly understood from lesion/dysfunction initiation through causal pathways and factors promoting deterioration in that condition. Nevertheless, a key take-away from this observation is that, although most of the time a clear understanding of the pathophysiology of the disease process will facilitate access to reliance upon a surrogate, the absence of a complete understanding of the disease process or even the existence of a relatively weak understanding of the disease process is not, in and of itself, incompatible with Subpart H.

- Part 3

With respect to the next key factor listed by FDA in its Draft Guidance, Part 3 of the analysis of each precedent reviews how well-understood the relationship is between the drug's effect on the surrogate and on the disease process. For this part of the analysis we searched the FDA reviews for evidence of reliance upon epidemiological associations (see, e.g., Sirturo and Makena), as well as the effect of another drug in the same or pharmacologically similar class of therapy to affect both the surrogate and the disease (see, e.g., Tysabri and Celebrex). Note that in several cases there was only relatively weak support for this relationship between the surrogate and the disease process, such as in the cases of Fabrazyme (in which little had ever been shown between clearance of substrate in particular cell types and renal function), Promacta, Remodulin, Synercid and Biaxin. Again, as in the case of Part 2, a weaker showing in this particular factor was not a bar to Subpart H qualification.

- Part 4

Finally, in its Draft Guidance, FDA noted the critical role of the clinical strength of evidence of the drug both on the surrogate and on benefit as well. While FDA was not able to articulate generalizable principles with respect to the strength of clinical evidence (Draft Guidance at lines 614-615), the power of this analysis is that by looking at the specifics of each of the 19 precedents, we may be able to ascertain that which may otherwise not be discernible. We divided the analysis of clinical evidence into two components: the clinical evidence of the drug on the surrogate and the clinical evidence of the drug on the clinical benefit.

With respect to the clinical evidence of each drug's effect on its surrogate, it is not surprising that 10 of the 19 precedents garnered the highest rating of 4 on a five point scale of 0 to 4. (Note that this factor was given the greatest weight in the overall analysis

because it was viewed by the commentators as the single most important factor.) However, even therapies such as Sulfamylon and Synercid, which had extremely weak strength of clinical evidence on their respective surrogates, were judged by FDA as appropriately qualified for Subpart H, carried mainly on the strength of other factors described for each of these precedents.

The latter half of the assessment of overall clinical evidence was the strength of evidence of clinical benefit. It was not anticipated that this would score high, and generally the Subpart H approval precedents had relatively little clinical evidence of benefit in the clinical data sets that were the basis for each approval. Eleven of the 19 precedents had essentially no substantial positive evidence of clinical benefit, and one of the precedents actually had a fairly strong negative numerical "lean" in clinical outcome evidence, suggesting that the therapy may have a negative impact on long-term clinical benefit.

- Overview

The FDA regulatory factors, which this analysis collected under the heading of Part 1, were remarkably consistently favorable for each of these 19 precedents. As for the relative strength of the FDA factors which this analysis housed under headings of Parts 2, 3 and 4, there were some noteworthy consistencies, especially within Part 2 (understanding of the disease process) and the component of Part 4 on the clinical evidence of the drug's effect on the surrogate. Also, of note, a weak assessment or contribution from Part 2 or Part 3 or even (remarkably to the commentators) the Surrogate Component of Part 4 did not prove to be a barrier to qualifying for Subpart H.

As with the prior analysis of FDA's orphan drug precedents by one of the commentators, this analysis of FDA's Subpart H precedents testifies to FDA's flexibility in applying its standards to therapies under its review. In 2013, both Congress and the President additionally and strongly exhorted FDA to extend and expand its use of Subpart H, especially beyond AIDS and cancer. By interpreting and applying the factors FDA laid out in its Draft Guidance to these precedents, the commentators hope that this analysis will help propel that endeavor.

CLOSING

In summary, this comment is meant to illustrate the various factors cited by FDA in its Draft Guidance, as well as present the strength of clinical evidence in each of the 19 Subpart H approvals. In so doing, this comment sheds light and provides vitality to the factors cited by FDA, in the Agency's Draft Guidance, as well as contributes to an

understanding of the strength of scientific and clinical evidence in FDA's reaching its prior Subpart H approval determinations. We hope that this will enable all to more easily and more frequently embrace Subpart H, this regulatory innovation created by FDA, as some of the veil obscuring the basis for FDA's determination when a surrogate is "reasonably likely to predict clinical benefit" has been, at least partially, now lifted.

Onward!

Sincerely,

Frank J. Sasinowski
Hyman, Phelps & McNamara, P.C.

Alexander J. Varond
Hyman, Phelps & McNamara, P.C.

Enclosures

Figure 1: Subpart H Analysis Keyed to Factors in FDA's Draft Guidance on Expedited Programs

By Frank Sasinowski and Alexander Varond to FDA, August 26, 2013

Drug	Part 1: Regulatory Factors Weighing into FDA Determination (Section VII.C.1)[1]				Part 2:[2] Understanding of the Disease Process (Section VII.C.1.a)[1]	Part 3:[2] Understanding of the Relationship Between the Drug's Effect on Surrogate and the Disease (Section VII.C.1.b)[1]	Part 4:[2] Strength of Clinical Evidence (Section VII.C.1)[1]		Total
	Statutory Factors			External Expertise (0-1)	Understanding of the Disease (0-3)	Epidemiological, Pharmacologic and "Other Evidence" on the Surrogate or Disease (0-3)	Surrogate Endpoint (0-4)	Clinical Benefit (0-3)	
	Severity (0-2)	Rarity (0-2)	Unmet Need (0-2)						
1. Sirturo	2	2	2	1	2	2	3	-1	13
2. Ferriprox	2	2	2	1	3	2	4	0	16
3. Makena	2	2	1	1	1	3	4	1	15
4. Promacta	2	2	2	0	2	1	4	1	14
5. Exjade	2	2	2	1	3	2	2	0	14
6. Levaquin	2	2	1	1	3	2	4	1	16
7. Tysabri	2	2	2	0	2	3	4	1	16
8. Luveris	2	2	2	1	1	2	2	2	14
9. Fabrazyme	2	2	2	1	3	1	4	0	15
10. Remodulin	2	2	2	1	3	1	3	1	15
11. Cipro	2	2	2	1	3	2	4	1	17
12. Celebrex	2	2	2	1	3	3	4	0	17
13. Synercid	2	1	2	1	3	2	2	0	13
14. Remicade	2	2	2	1	1	1	3	0	12
15. Priftin	2	2	2	1	2	2	2	0	13
16. Sulfamylon	2	2	2	1	3	2	1	0	13
17. ProAmatine	1	2	2	0	3	2	3	0	13
18. Biaxin	2	1	2	1	2	1	2	1	12
19. Betaseron	2	2	2	1	2	2	4	2	17
	Range for Part 1: 5 to 7 (out of 7)				Range for Part 2: 1 to 3 (out of 3)	Range for Part 3: 1 to 3 (out of 3)	Range for Part 4: 1 to 6 (out of 7)		

[1] These citations are to the FDA Draft Guidance

[2] Parts 2, 3, and 4 are also from FDC Act Sec. 506, as amended by FDASIA Sec. 901, specifically, FDASIA uses the following terms for each of these parts:
- Part 2 relates to "pathophysiological" evidence
- Part 3 relates to "epidemiological, . . . pharmacologic, or other evidence . . ."
- Part 4 relates to "therapeutic" evidence

Statistics	Score
Average Score	14.5
Min	12
Max	17
Median	14
SD	1.7

APPENDIX 1

1. SIRTURO (bedaquiline)

This December 28, 2012 approval for treating multi-drug resistant tuberculosis (MDR-TB) was based on a surrogate of time to sputum culture conversion.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

"Overall mortality still exceeds 10%, with a range of 8 to 21% for patients enrolled into good treatment programs." (Medical Review, Dec. 26, 2012, p. 22).

b. Rarity of the Condition

FDA granted Sirturo orphan drug designation on January 10, 2005. Furthermore, in FDA's determination that the time to sputum culture conversion is an acceptable surrogate on which to base accelerated approval, it appears that FDA may have taken into account specifically the rarity of MDR-TB in this country in that FDA acknowledged that: "In the United States, the total number of MDR-TB cases has fluctuated from 88 to 132 cases [since] 1993, with 88 cases reported in 2010." (Medical Review at p. 22).

c. Lack of Available Therapy

"Treatment of MDR-TB is more complex (than treating drug-susceptible TB or DS-TB) and prolonged and typically has a favorable outcome rate [of only] 41-70%. Cases of MDR-TB are currently treated with at least five second-line anti-TB drugs for an extended period of time that may last up to two years . . . The challenges of the treatment of MDR-TB include toxicities of the drugs, decreased potency, cost (50-200 times more expensive than DS-TB) and the need for possible hospitalization." (Medical Review at p. 22).

d. Use of External Expertise

FDA did turn for external expertise to the Anti-Infective Drug Advisory Committee, which on June 3, 2009, "voted 18 to 1, recommending that sputum culture conversion . . . could be used as a surrogate . . . [t]herefore, the committee recommended that approval of an antimycobacterial drug could be done under Subpart H regulations (Accelerated Approval) using sputum culture conversion as a surrogate endpoint. Further, traditional endpoints used to evaluate treatment response such as relapse, failure, and mortality should still be used . . . for traditional approval." (Medical Review at p. 28).

Part 2. Understanding of the Disease Process

In this case, the pathophysiology of MDR-TB is well-understood.

Part 3. Understanding of the Relationship Between Sputum Culture Conversion and Relapse, Long-Term Response and Mortality

Epidemiologic evidence exists that supports the relationship between sputum culture conversion and clinical outcome, in particular, mortality. See Shama D. Ahuja et al., "Multidrug Resistant Pulmonary Tuberculosis Treatment Regimens and Patient Outcomes: An Individual Patient Data Meta-analysis of 9,153 Patients," 9(8) PLOS Medicine e1001300 (2012).

Part 4. Clinical Evidence of Sirturo's Effect on Sputum Culture Conversion and on Relapse and Mortality

The FDA Medical Reviewer noted the existence of the epidemiological evidence, but stressed that the clinical evidence provided by the sponsor both on the surrogate and on traditional endpoints of clinical benefit, especially mortality, would be "most persuasive." In this case, the Medical Reviewer listed these traditional endpoints as relapse, long-term response, and mortality. (Medical Review at p. 16).

There were two Phase 2 clinical trials that comprised the clinical evidence for this drug on the surrogate and on clinical benefit, but only one of which was considered to be the single, pivotal trial: Study C208 Stage 2. Study C208 Stage 2 was a randomized, double-blinded, placebo-controlled trial with a 24-week treatment period in which both the drug and "placebo" arms received an optimized background regimen. (Statistical Review, July 26, 2012, p. 6).

a. Sputum Culture Conversion

The primary endpoint, which was the surrogate endpoint, of the time to sputum culture conversion was highly statistically significant (p-value of 0.0005) (N=160 randomized, with 67 and 66 subjects in the drug and placebo arms in the mITT analysis, respectively). Sputum culture conversion at week 24 was a key secondary endpoint (as well as another supportive measure of the surrogate endpoint of sputum culture conversion), and it too was statistically significant (p-value = 0.014) with 78% and 58% of drug and placebo arm subjects, respectively, achieving sputum culture conversion at week 24. (Statistical Review at p. 6). "Lastly culture conversions data after all patients completed 72 weeks in the study showed a statistically significant but diminishing improvement in the time to sputum culture conversion for [Sirturo-]treated patients compared to placebo-treated patients." (Medical Review at p. 44).

b. Relapse and Mortality

Relapse is a "traditional" measure of clinical benefit. The Medical Reviewer notes that in "the mITT population, five subjects (7.6%) in the [drug] group and eight subjects (12.1%) in the placebo group experienced relapse . . . [However,] the subjects in the placebo group appear to take a longer time from culture conversion to relapse than those in the [drug] group." (Medical Review at pp. 59-60). Therefore, the Medical Reviewer conducted an alternative analysis, and in this analysis, "the two treatment arms become more comparable with respect to relapse with 5 relapses on [drug] and 4 on placebo." (Medical Review at p. 60).

Survival is the most objective and clinically meaningful benefit in MDR-TB. In the pivotal study, 9 of 79 in the drug arm died (11.4%) compared to 2 of 81 (2.5%) in the placebo arm. (Medical Review at p. 70). Both placebo subjects died of TB as did 5 of the 9 subjects in the drug group. (Medical Review at p. 70). Signals of QT prolongation and serum transaminase elevation, with one death due to liver injury in the drug arm, were also observed. (Medical Review at pp. 70-71).

In the "summary and conclusions" section of the statistical review, FDA observed: "There was a statistically significant increase in mortality in the [drug] group. Despite the observed treatment benefit in time to culture conversion, it did not lead to a benefit in patient survival. This was a major concern both for efficacy and safety." (Statistical Review at p. 60).

The relationship between the traditional clinical endpoints of relapse and survival and the surrogate endpoint of sputum culture conversion were not robust in this case. In fact, the clinical evidence on survival was actually and strongly in the wrong numerical direction.[1] Notwithstanding this, FDA appears to have, as noted in its Draft Guidance, relied in part on the "external expertise" of the June 2009 Anti-Infective Drug Advisory Committee as well as took "into account" these three factors that were listed in FDASIA: (1) the "severity" of the disease; (2) the "rarity" of the disease; and (3) the "lack of alternative treatments." (See, e.g., Medical Review at top of p. 59).[2]

---

[1] This is the reason for the commentators scoring clinical evidence on the actual clinical benefit as -1 on a scale of 0 to 3. The scale was set up under the assumption that, at worst, there would be an absence of any clinical evidence of benefit, or if clinical evidence, then not even any "lean" in favor of the investigational treatment, which then would have been rated as "0."

[2] In addition to Dr. Porcalla's medical review reaching this conclusion, every other review unanimously supported a recommendation for approval. For instance, the statistical review by Dr. Lit Higgins concluded: "The efficacy in terms of a surrogate endpoint, sputum culture conversion, was supported by the pivotal study

2. FERRIPROX (deferiprone)

FDA approved Ferriprox on October 14, 2011 as "an iron chelator . . . for the treatment of patients with transfusional iron overload due to thalassemia syndrome when current chelator therapy is inadequate." Ferriprox was approved on the basis of its showing on an unvalidated surrogate, serum ferritin.

Part 1. Factors Weighing into FDA Determination

a. Severity of the Condition

Persons with certain inherited anemias, especially sickle cell anemia and thalassemia, require frequent red blood cell (RBC) transfusions because they are unable to manufacture hemoglobin. Each unit of packed RBCs contains 200 mg of iron, which is an extreme excess of iron as compared with the dietary intake of 1 mg of iron necessary to maintain normal total body iron stores in healthy individuals. Without a way for the body to excrete excess iron, persons receiving these regular transfusions of RBCs build up massive iron overload which leads to morbidity and often eventually death due to cardiac damage. (Medical Review #1, Sept. 20, 2011, pp. 1-2).

b. Rarity of the Condition

FDA designated Ferriprox as an orphan drug on December 21, 2001.

c. Lack of Available Therapy

At the time of Ferriprox's approval, there were two other approved therapies for iron overload due to transfusions: Desferal (deferoxamine) and Exjade (deferasirox). Ferriprox was given fast track designation in January 2004, before Exjade was approved. Exjade, an orally active iron chelator, was approved in 2005. In January 2004, Desferal was the only available therapy and requires continuous infusion over many hours, every day.

C208 and supportive study C209. There was a significantly elevated mortality risk in the [Sirturo] group. This should be considered in an approval decision and use of this regimen." The reviews of the Cross-Discipline Team Leader, Dr. Navarro (December 21, 2012), the Deputy Division Director, Dr. Laessig (December 27, 2012) and the Office Director, Dr. Cox (December 28, 2012) all recognized the robust finding on the surrogate endpoint of sputum culture conversion and recommended approval despite serious consideration of the clinical safety results, especially the survival results in the pivotal study. This unanimity of support for a Subpart H approval decision within the entirety of the internal FDA expert review team was not always observed in the other 18 Subpart H precedents.

The sponsor first submitted its NDA seeking an indication for "all transfusion-dependent anemias for whom the use of other iron chelators has been considered inappropriate." A complete response letter was issued in November 2009 and a resubmission was made in April 2011 for essentially the same second-line use. However, the data submitted were almost exclusively from thalassemia patients and FDA's October 2011 approval is for "patients with transfusional iron overload due to thalassemia syndromes when direct chelator therapy is inadequate." For this specific use, there is a lack of available therapy.

d. External Expertise

FDA appears to have given consideration to two types of external expertise. First, FDA seems to have given some weight to the "expertise" of clinical practice that uses serum ferritin to monitor the patient's iron status. While serum ferritin is a non-specific endpoint for which FDA noted that "the relationship between the serum ferritin and clinical outcome is not well-established" (Medical Review #2, Sept. 16, 2011, p. 34), FDA nevertheless appears to give serum ferritin some weight because serum ferritin is "a commonly used parameter for following body iron burden in patients undergoing chronic red blood cell transfusions," (Medical Review #1 at p. 12), and because "in clinical practice, measurements of serum ferritin and [liver iron concentration] have been the generally accepted methods of evaluation of the efficacy of therapy in persons with iron overload." (Medical Review #3, November 20, 2009, p. 5).

Second, the Oncology Drugs Advisory Committee recommended Ferriprox for approval on September 14, 2011 by a vote of 10 to 2 for treating patients in whom current chelator therapy is inadequate.

Part 2. Understanding of the Disease Process

In this case, the pathophysiology by which iron overload leads to deposition of iron in tissues and leads to iron-catalyzed peroxidation of membrane lipids, which then leads to morbidity and death due to cardiac damage, is well-known. (Medical Review #1 at p. 1).

Part 3. Understanding of the Relationship Between the Effect on Serum Ferritin and Cardiotoxicity and Death

The mechanism of the drug's action is well-known, that is, binding to iron in a 3:1 complex which is excreted in the urine, and the reduction in iron in these persons is needed to avoid iron overload morbidities. (Medical Review #1 at p. 2). However, serum ferritin is non-specific and "changes in serum ferritin are difficult to interpret because serum ferritin is subject to variations induced by a number of mechanisms that are unrelated to total body iron." (Medical Review #4, Oct. 19, 2009, p. 15). Most of all, "[t]he relationship between the serum ferritin and clinical outcome is not well established." (Medical Review #1 at p. 34).

This part was scored a 2 on a scale of 0 to 3, mainly on the basis of the biologic plausibility that this drug, due to its mechanism, would reduce iron stores, notwithstanding the weakness of serum ferritin itself as a surrogate, due to its lack of specificity as a measure of iron stores. The non-specificity of serum ferritin and the lack of understanding of the relationship between the surrogate and outcomes led to a score of 2 instead of 3.[3]

Part 4. Clinical Evidence of Ferriprox's Effect on Serum Ferritin and on Outcome

It is of value here to note that FDA rejected the original NDA submitted in 2009 for Ferriprox because the "primary efficacy endpoint of the single major controlled trial . . . was the change in cardiac MRI T2* which was said to measure iron content within the heart. FDA stated that this endpoint was a surrogate endpoint and there were no data to support the incremental changes in the values as predictive of clinical benefit." (Medical Review #1 at p. 10) (emphasis added). Moreover, "secondary endpoints [of serum ferritin and liver iron concentration] also were not consistently corroborative of the primary endpoint [MRI T2*] results." (Medical Review #1 at p. 5). Overall, "the study did not find a significant correlation between change in cardiac MRI T2* and measures of cardiac function and there were no differences between treatments in change in liver iron concentration (LIC)." (Medical Review #1 at p. 2). The statistical review observed that "the patients in this study were not followed for clinical outcome and therefore, this study was not designed to obtain internal validation of MRI T2* change as a surrogate for any clinical outcome indicative of reduced cardiac iron."[4] (Statistical Review, March 24, 2009, p. 7).

"Although the data from this study provided statistically significant evidence . . . in MRI T2* . . . this study was not designed to and therefore, does not provide evidence that change in MRI T2* is reasonably likely to predict clinical benefit due to lack of long-term follow-up of these patients." (Statistical Review #2, Nov. 22, 2009, p. 3).

In response to FDA's rejection of the original NDA, the sponsor "conducted an analysis of a subpopulation of patients drawn from its previously conducted studies and defined as being inadequately treated with current chelator therapy." (Medical Review #1 at p. 10). In this analysis, approximately 50% met the primary efficacy endpoint of

---

[3] Others may score this differently, perhaps even only a "1" given the non-specificity of serum ferritin and lack of well-established relationship between surrogate and outcomes.

[4] Note that FDA states that this study could provide both evidence of the effect of the drug on an unvalidated surrogate and at the same time, in the same study, evidence of the effect of the drug on clinical outcome, thereby "validating" that surrogate.

having a 20% or greater decline in serum ferritin. Of additional importance, the sponsor-defined "success rate" in this same analysis was 42% for liver iron concentrate (LIC). (Medical Review #1 at pp. 7-8). FDA noted that "change in LIC using liver biopsy has generally been considered to be the standard measure of efficacy in response to iron chelation therapy." (Medical Review #4 at p. 15).

Overall, FDA first rejected the original NDA on grounds that the primary endpoint of the key pivotal study, MRI T2* changes, was not sufficiently correlated with any clinical outcome to warrant being the basis for even an accelerated approval, notwithstanding the disease being severe, rare, and without adequate therapy. However, FDA approved a second resubmission that was based on an analysis of a commonly used measure in clinical practice of patients with transfusion-related iron overload, serum ferritin, which itself was supported internally by a positive finding in the same population on liver iron concentration which is the "standard measure of efficacy in response iron chelator therapy."

FDA's actions on Ferriprox illustrate the fatal flaws in a clinical program attempting to rely upon a surrogate (MRI T2*), the factors to be considered and the clinical evidence that were found by FDA to be of sufficient merit to allow FDA, as a matter of its judgment, to conclude that serum ferritin is reasonably likely to predict clinical benefit, even without any clinical trial results on any cardiac outcomes, such as heart failure or mortality, and notwithstanding an FDA acknowledgement that serum ferritin is a non-specific measure. However, FDA's Subpart H approval here was based clinically on the corroboration of the serum ferritin results by the liver iron concentrate results and bolstered by the known mechanistic action of the drug (i.e., that by its mechanism of action the drug leads to iron excretion in the urine).

Overall, the clinical evidence of the surrogate was scored a full 4 out of a possible 4 due to the strength of evidence on serum ferritin which itself was buttressed by the clinical findings on LIC. However, since there was no clinical evidence on any ultimate clinical outcome, the score for clinical evidence of outcome benefit is zero.

3. **MAKENA (hydroxyprogesterone caproate)**

FDA's February 3, 2011 approval of Makena to reduce the risk of preterm birth (PTB) was based on a surrogate of reducing preterm birth as defined as those births occurring at less than 37 weeks of gestation. "Preterm birth <37 weeks gestation . . . was a surrogate[5] for pregnancy outcome (neonatal/infant morbidity and mortality)." (Medical Review, Feb. 3, 2011, p. 14).

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

The risks of miscarriage, stillbirths, and neonatal mortality are associated with delivery prior to full-term gestation, as well as neonatal morbidities and adverse maternal outcomes as well.

b. Rarity of the Condition

Makena was designated as an orphan drug on January 25, 2007.

c. Lack of Available Therapy

"Currently there is no drug product approved in the United States to reduce the risk of preterm birth; however, [the active ingredient in Makena] is compounded by pharmacists and is used widely for this indication in women at high risk." (Medical Review at p. 11). In 1956, FDA had approved an NDA for Delalutin, which had the same active ingredient as Makena, for treating pregnant women for "habitual and recurrent abortion, threatened abortion." (Medical Review at p. 12). In 2000, FDA withdrew the approval of Delalutin at the request of the NDA sponsor because it no longer marketed Delalutin. In a June 25, 2010 Federal Register notice, FDA announced its determination that Delalutin was not withdrawn from marketing for safety or efficacy reasons.

d. Use of External Expertise

With Makena, FDA relied upon two forms of external expertise and FDA reached its "informed judgment" that the surrogate endpoint of preterm birth less than 37 weeks was reasonably likely to predict clinical benefit, that is, pregnancy outcome or neonatal infant and maternal morbidity and mortality. These two forms of external advice are summarized in the Medical Review: (1) 2006 Advisory Committee; and (2) subsequent scientific papers published in the literature.

---

5 While FDA Medical and Statistical Reviews refer to PTB <37 weeks as a "surrogate," preterm birth is a clinical event and, therefore, in the terminology of the Draft Guidance, PTB <37 weeks is an "intermediate clinical endpoint."

i. "The surrogate endpoints of reductions of [preterm birth] at <35 and <32 weeks were thought by the Advisory Committee to predict a reduction in neonatal mortality and morbidity. At the time of the Advisory Committee meeting in 2006, the endpoint PTB at <37 weeks was not believed to be an adequate surrogate for neonatal outcome."[6] (Medical Review at p. 6).

ii. "The Applicant submitted a single phase 3 clinical trial which demonstrated a statistically strong ($p<.001$) reduction in the incidence of preterm births prior to 37 weeks gestation, the protocol pre-specified primary endpoint. There is recent evidence that 'late preterm births' (births between $34^{0/7}$ and $36^{6/7}$), which comprise 71.3% of all preterm births, are increasing, and suffer greater neonatal and childhood morbidity and mortality than previously thought [5 papers are cited that were published between the time of the 2006 Advisory Committee and the Medical Review]. These data indicate that 'preterm birth prior to 37 weeks' is a surrogate endpoint that is reasonably likely to predict clinical benefit." (Medical Review at p. 5).

Part 2. Understanding of the Disease Process

Here the disease process is complex and has multiple pathophysiologic pathways, and therefore, this mitigates against reliance upon any surrogate. The biological means by which the gestational process progresses to premature delivery is complex and multifaceted. Therefore, the surrogate endpoint of PTB <37 weeks is likely more analogous to the PSA example than the enzyme replacement example in the Draft Guidance (see Draft Guidance at p. 19, lines 634-648) in that PTB <37 weeks is not on the pathophysiological causal pathway and is not the biologic mechanism that causes the neonatal mortality and morbidity, even though, like PSA, it is correlated with increased risk.

Part 3. Understanding of the Relationship Between PTB and Pregnancy Outcomes

a. Epidemiological Evidence

The epidemiological evidence is strong with Makena. The 2006 Advisory Committee assessed the epidemiological evidence supporting the relationship between PTB and pregnancy outcomes and found that this evidence was strong enough to support the endpoints of PTB <32 weeks and PTB <35 weeks as surrogate endpoints but not PTB <37 weeks. However, additional evidence published subsequent to the 2006 Advisory Committee permitted the Medical Officer, Dr. Barbara Wesley, to conclude that PTB <37

---

6 "The Committee stated that a reduction of preterm birth <37 weeks was not an adequate surrogate (Yes: 5; No: 16) but that reductions in preterm birth <35 weeks (Yes: 13; No: 8) and <32 weeks (Yes: 20; No: 1) were adequate surrogates." (Medical Review #2, Jan. 23, 2009, p. 7).

weeks was also a reliable, consistent and acceptable surrogate endpoint.[7] (Medical Review at p. 5).

b. Effect of Drugs in the Same or Closely Related Pharmacologic Class to Affect Pregnancy Outcomes

Since there are no drugs in any pharmacologic class approved for reducing the risk of PTB, there are no analogous therapies here on which to draw support directly for reducing the risk of PTB. However, other progesterones including the active ingredient in Makena have been approved for aiding in assisted reproductive technologies and other conditions supporting the maintenance of pregnancy.

Part 4. Clinical Evidence of the Makena's Effect on PTB <37 Weeks and on Pregnancy Outcomes

a. PTB <37 Weeks

The surrogate of PTB <37 weeks was highly statistically significant ($p<0.001$).

b. Pregnancy Outcomes

"The proportion of babies with at least one event on the [secondary] composite index of neonatal morbidity and mortality was lower in the [Makena] group (11.9%, 35/295 infants) than in the vehicle group (17.2%, 26/151 infants) but the between-group differences was not statistically significant (nominal p-value of 0.1194)." (Medical

---

[7] It is also likely that the Advisory Committee was opining on PTB <32 weeks, PTB <35 weeks and PTB <37 weeks as validated surrogates which would have qualified Makena for traditional approval, not Subpart H approval. Outside of AIDS and cancer, FDA has not often asked Advisory Committees to opine on whether clinical evidence on a particular endpoint would qualify a therapy for Subpart H approval. For example, note that the August 5, 2013 Cardiorenal Advisory Committee, addressing the approvability of tolvaptan, a vasopressin V2 receptor antagonist, was not asked whether total kidney volume would qualify as an unvalidated surrogate that may support a Subpart H approval if the Advisory Committee found that total kidney volume is "reasonably likely to predict clinical benefit," which, in this case, clinical benefit would likely be end-stage renal disease and/or clinically meaningful outcomes such as significant worsening of renal function or kidney pain. However, there are exceptions outside of AIDS and cancer. For instance, the Oncology Drugs Advisory Committee (ODAC) was asked whether FAP was an adequate "unvalidated" surrogate, that is, to qualify Celebrex (Precedent #12) for Subpart H approval. But even this case was before ODAC, and while FAP is not cancer, the ultimate clinical benefit was prevention of colon cancer, so even this "exception" is not fully outside of AIDS and cancer.

Review #1 at p. 6). "Approximately 6.5% of the women in each treatment group experienced a fetal or neonatal deaths . . . The results . . . show that despite the treatment groups having about the same rate of fetal and neonatal deaths, the losses occur earlier among [Makena] women." (Statistical Review #2, Oct. 19, 2006, p. 20).

This impact on fetal or neonatal deaths was stated another way by the Medical Reviewer: "There was a trend toward an increased risk of miscarriage and stillbirths in the [Makena] treatment arm and a trend toward a decrease in neonatal death, with no overall net survival benefit." (Medical Review #1 at p. 7) (emphasis in original).

Overall, the secondary endpoint of a composite measure of neonatal morbidity/mortality leaned in favor of the Makena group while the separate analysis of neonatal mortality showed essentially no numerical difference and had a nominal p-value of 0.6887 (Medical Review #1 at p. 7). The clinical evidence for the ultimate clinical benefits in the single pivotal trial was not strong.

4. **PROMACTA (eltrombopag)**

FDA approved Promacta on November 20, 2008 on "short-term platelet count response" as a surrogate marker for longer platelet count responses (platelet counts are recognized as acceptable measures of clinical benefit for patients with ITP [idiopathic thrombocytopenic purpura])." (Medical Review #1, Nov. 4, 2008, p. 3). The two clinical trials of Promacta administered drugs over 6 weeks or less (this is the meaning of "short term" in the Reviewer's statement above). Had the Promacta trials studied and established the drug's effect on platelet counts out to 6 months, this approval would have been a traditional approval and not one under Subpart H.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

Chronic ITP is a serious medical condition. (Medical Review #1 at p. 3). The frequency of death from hemorrhage in patients with platelet counts below 30,000/mcl is estimated to be between 1.6 and 3.9% per patient year. (Medical Review #2, Sept. 12, 2008, p. 17).

b. Rarity of the Condition

FDA designated Promacta as an orphan drug on March 4, 2008.

c. Lack of Available Therapy

"[Promacta] approval would provide a meaningful therapeutic benefit to patients over existing treatments because of its minimal risk for immunogenicity (based upon [its] small molecule characteristics). The labeling for romiplostine, the only currently marketed TPO receptor agonist, includes information regarding the risks for immunogenicity. These risks are not applicable to [Promacta]." (Medical Review #1 at p. 3).

d. Use of External Expertise

In the medical and statistical reviews, the commentators found no evidence of any reliance on special government employees (SGEs), an Advisory Committee for Promacta, or specific published literature.

Part 2. Understanding of the Disease Process

"The clinical hallmark of the disease is an increased tendency to bleed." (Medical Review #2 at p. 17). Furthermore, the relationship of platelet count to bleeding is well-established: "Patients with platelet counts between 30,000/mcl and 10,000/mcl are generally considered treatment candidates due to slightly increased risk of spontaneous bleeding or increased risk of bleeding due to trauma." (Medical Review #2 at p. 17).

Part 3. Understanding of the Relationship Between the Drug's Effect on Short-Term Platelet Counts and Increased Risk of Bleeding

There was no epidemiological evidence cited in the FDA review documents to support the surrogate - which is "short term" (that is, six weeks) increase in platelet count - as reasonably likely to predict long-term, chronic increase in platelet count - which is generally established in six month trials or generally on increased risk of bleeding. While there was no evidence to support the use of this surrogate, there was a therapy approved from the same pharmacologic class but based on an endpoint of six-month duration. Earlier in 2008 (the year FDA approved Promacta), FDA had approved romiplostim, a biological product that is a member of the same pharmacologic class - thrompoietin (TPO) receptor agonists - and this approval for the same indication (that is, to treat ITP) was a traditional approval based on two clinical trials, each of six-months duration.

Part 4. Clinical Evidence of Promacta's Shorter-Term (Surrogate) Effect and Long-Term Effect on Platelets and/or Bleeding

Both Promacta pivotal studies showed a robust short-term (surrogate) effect on platelets ($p<0.001$) (Statistical Review, Apr. 29, 2008, pp. 19 and 27).

As for clinical evidence that FDA had at the time of the approval that Promacta's "short-term" (six weeks) impact on platelet counts would predict either clinical benefit of long-term impact on platelet counts or on bleeding, there was mixed evidence.

As supportive evidence that the platelets produced by Promacta behaved in a physiologically "normal" way, the Sponsor had conducted "an exploratory clinical study that demonstrated [that Promacta] prompted platelet count increases in healthy subjects. These drug-stimulated platelets had *in vitro* platelet function characteristics typical of platelets. Hence, this study supported the generally accepted use of platelet counts as an 'accepted' measure of clinical benefit for clinical studies of TPO receptor agonists among patients with chronic ITP." (Medical Review #1 at pp. 2-3).

As Promacta was only administered for six weeks (or less) in the two pivotal trials, there is no clinical evidence as to the impact long-term on platelet counts if Promacta was administered chronically (for which a trial of six-months duration would have been relied upon). Furthermore, of some concern, "discontinuation of [Promacta] at the end of the study resulted in an unacceptable amount of serious hemorrhage." (Medical Review #1 at p. 3). Also, the statistical reviewer observed that within two weeks after the subjects on drug were off treatment, there was a return to placebo levels of platelet counts. (Statistical Review at pp. 27-28).

As for bleeding events, there was a numerical lean in favor of Promacta, but in neither trial was this statistically significant with p-values of 0.121 and 0.088 for the between-group difference on bleeding events in the two pivotal trials. (Statistical Review at pp. 8-9).

5. EXJADE (deferasirox)

The FDA approval of Exjade for treating "chronic iron overload due to blood transfusions" on November 2, 2005 was based on a surrogate endpoint of improvement in liver iron concentration (LIC).

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

"Chronic iron overload due to requisite blood transfusion is a serious and life-threatening condition." (Medical Review #1, Nov. 2, 2005, p. 2).

b. Rarity of the Condition

Exjade was granted orphan drug designation on November 21, 2002.

c. Lack of Available Therapy

At the time FDA was reviewing the Exjade NDA, the Medical Team Leader, Dr. Dwaine Rieves, stated: "Deferoxamine, the only available therapy for this condition, presents unique compliance and infectious risks due to the need for prolonged administration of the drug. [Exjade] is an orally administered drug that provides a meaningful therapeutic benefit over the existing therapy." (Medical Review #1 at p. 2).

d. External Expertise

FDA sought the advice of the Blood Products Advisory Committee (BPAC) and at its September 29, 2005 meeting, the BPAC found that "the applicant [had] provided substantial evidence of the effectiveness of [Exjade] in the reduction of liver iron concentration, an outcome indicative of a clinical benefit . . . The sponsor's major clinical evidence of [Exjade] effectiveness . . . is based upon alterations in liver iron content, an endpoint the BPAC discussants regarded as a measure of clinical benefit. In this context, the endpoint is not regarded as a surrogate endpoint rather as an endpoint other than survival or irreversible morbidity[8], as cited in the Subpart H regulations." (Medical Review #1 at p. 2).

---

[8] An "intermediate clinical endpoint" (rather than a surrogate) is the term used in the Draft Guidelines for this kind of endpoint; however, during the later FDA approval of Ferriprox, the FDA Reviewers refer to both serum ferritin and LIC as "surrogates," and in an earlier medical review of Exjade, FDA refers to LIC in this pivotal trial as a "surrogate" (see Medical Review #2 at p. 38). Therefore, this analysis will refer to LIC as a surrogate and not as an intermediate clinical benefit.

Part 2. Understanding of the Disease Process

See Item 2 under Ferriprox.

Part 3. Understanding of the Relationship Between LIC and Cardiac Outcomes, Including Mortality

"Although accepted by the Division as a clinically meaningful endpoint, the primary endpoint [of LIC] is technically a surrogate endpoint since it does not necessarily address clinically significant morbidity or mortality. The main mortality on β-thallasemia is due to cardiac dysfunction whose etiology in β-thallasemia is probably multifactorial. Nonetheless, most of the literature in β-thallasemia has used LIC as a marker for morbidity for other organ involvement and as a surrogate for mortality. There is some information, however, that LIC does not completely correlate to the extent of cardiac hemosiderosis, the primary cause of mortality. Obviously, repetitive biopsy of the myocardium to measure iron concentration in the heart is not acceptable." (Medical Review #2, Oct. 10, 2005, p. 38).

As for understanding the relationship between drugs in the same pharmacologic class as LIC, the single pivotal trial for Exjade was a noninferiority study design which used as its active comparator, deferoxamine, and therefore, FDA had evidence from a within-study comparison of the only other member of the same or closely related class on the surrogate endpoint of LIC.

Part 4. Clinical Evidence of Exjade's Effect on LIC and/or Cardiac Outcomes Including Mortality

FDA, in its review of this NDA, noted that LIC as "the primary endpoint is acceptable and it was agreed to by the Division in the Special Protocol Assessment. It should be remembered, however, the LIC is a surrogate marker and that the effects of Exjade on morbidity/mortality, which are the truly important clinical endpoints, are not likely to be demonstrated in this short trial." (Medical Review #2 at p. 31).

Rather than bolster LIC results by seeing trends on irreversible morbidity and mortality in this "short" trial, FDA looked to find support from other critical surrogate markers such as serum ferritin.[9]

As for LIC, the protocol had specified that "non-inferiority of [Exjade] to [deferoxamine] was to be established if the two sided 95% confidence interval of the difference in success rate between the two groups was above -15%. The basis for the

---

[9] The authors must inform the reader that this trial was a year-long trial, and, therefore, by many would not be considered "short;" however, even a year long study is too "short" to see effects on mortality and irreversible morbidity.

choice of this [non-inferiority] margin was unclear in the submission. Notably, FDA had questioned the meaningfulness of this margin during the study's protocol review.[10] (Medical Review #1 at p. 4).

The primary efficacy result was a point estimate difference of -13.5%, with a lower 95% confidence interval of -21.6% (or, in other words, the margin defining success of the trial was not met). About this, FDA concluded: "Given that the original basis of the non-inferiority margin was poorly substantiated, little clinical meaningfulness could be assigned to failure to achieve the primary endpoint. The primary endpoint data did establish that both [Exjade and deferoxamine] lowered LIC over a 12 month period of time, a time period during which subjects would have been expected to have increases in LIC due to continuing blood transfusions. This observation provides evidence of a treatment effect for [Exjade]." (Medical Review #1 at p. 5).

With respect to serum ferritin, FDA concluded that "[s]erum ferritin values declined in a dose-related manner for subjects receiving [Exjade], a pattern similar to that for subjects receiving [deferoxamine]." (Medical Review #1 at p. 5).

---

[10] Query, though, how FDA nevertheless had accepted the design of this pivotal study under an SPA.

6. LEVAQUIN (levofloxacin)

FDA approved Levaquin for post-exposure prevention of inhalational anthrax on November 11, 2004. Much of what the Agency had learned from its Subpart H approval of Cipro for inhalational anthrax in August 30, 2000 was used to create a draft guidance. "FDA Draft Guidance for Industry: Inhalational Anthrax (Post-Exposure) - Developing Antimicrobial Drugs" ("Anthrax Draft Guidance") (March 2002). FDA then relied on its Anthrax Draft Guidance when it approved Levaquin in 2004. (Statistical Review, Nov. 15, 2007, p. 1).

As for Cipro, there was a two part or "compound" surrogate for this approval in that FDA concluded: (1) that "[m]ortality due to anthrax for animals that received a 30 day regimen of oral Levaquin beginning 24 hrs post exposure was significantly lower (1/10), compared to the placebo group (9/10) [p=0.0011]," and (2) "mean plasma concentrations of Levaquin associated with a statistically significant improvement in survival over placebo in rhesus monkey model of inhalational anthrax are reached or exceeded in adult . . . [human] patients receiving the recommended oral and intravenous dosage regimens." (Levaquin Package Insert).

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity and Rarity of the Condition

"Mortality for established [inhalational anthrax] even after treatment was 80-100% in the 20th century." (Anthrax Draft Guidance at p. 3). In addition, "inhalational anthrax is extremely rare. There have been only approximately 20 cases in the United States in the past 100 years . . . For these two reasons, the rarity of disease and the extremely high mortality rate, a clinical study is not feasible." (Cipro Statistical Review, Aug. 16, 2000, p. 1).

b. Rarity of the Condition

Although the prevalence of inhalational anthrax is sufficiently low, the Sponsor did not seek orphan drug designation.

c. Lack of Available Therapy

At the time of Levaquin's approval, Cipro was indicated specifically for post-exposure prophylaxis for disease caused by inhaled *B. anthracis*, and, although doxycycline and penicillin G procraine products were not specifically indicated for post-exposure prophylaxis for disease caused by inhaled *B. anthracis*, FDA "had published a notice in the Federal Register (66 Fed. Reg. 55679) that clarified the dosing regiments for [those drugs] in the management of patients with inhalational anthrax." (Anthrax Draft Guidance at p. 5).

d. External Expertise

Although no advisory committee was convened specifically for Levaquin, FDA had sought "input from the Anti-Infective Advisory Committee [and determined that] the use of the rhesus (macaque) monkey disease and treatment model for inhalational anthrax (post-exposure) provides convincing evidence of efficacy for regulatory evidence." (Anthrax Draft Guidance at p. 4).

Part 2. Understanding of the Disease Process

Before the approval of Cipro in 2000, four years before the approval of Levaquin, FDA had stated that "[t]he inhalational form of the disease, which affects the mediastinal lymph nodes, other organs of the reticuloendothelial system and the central nervous system, is considered the most likely clinical entity resulting from the intentional use of an aerosolized preparation of the spores of *B. anthracis*." (Cipro Medical Review, Aug. 31, 2000, p. 2).

Part 3. Understanding of the Relationship between the Monkey Data and Human Mortality; and Part 4. Clinical Evidence

FDA's draft guidance document on the development of treatment for post-inhalational anthrax exposure stated that, "a non-human primate model that models the drug disposition in humans [was] considered an adequate surrogate for human disease and objective endpoints such as mortality, time of death relative to antimicrobial use, pathology, and bacteremia in the macaque." (Statistical Review at p. 1).

Thus, two findings formed the basis of FDA's Subpart H approval of Levaquin for inhalational anthrax:

First, "[s]urvival was significantly better ($p=0.0011$, two-sided Fishers exact test) and time to death was significantly longer ($p<0.0001$, log rank test) [in macaques] in the levofloxacin group compared to the placebo group." (Statistical Review at p. 1). Also, Levaquin had a numerical advantage with 90% (9/10) of the macaques surviving, compared to 80% (8/10) in the ciprofloxacin group, and only 10% (1/10) in the placebo group. (Statistical Review at p. 1).

Second, as for comparative monkey/human exposure levels, the "mean plasma concentrations [and mean steady state AUC_{0-24}] associated with a statistically significant improvement in survival over placebo in the rhesus monkey model of inhalational anthrax are reached or exceeded in adult . . . patients receiving the recommended oral and intravenous dosage regimens." (Levaquin Package Insert).

As for understanding the relationship of drugs in the same or closely related class on the compound surrogate, see above discussion under 1.c. regarding other drugs including Cipro for anthrax.

Monkey survival data was one part of this unusual compound surrogate; see the discussion above. However, there were "complete pharmacokinetic data on the drug in human volunteers . . . and pharmacokinetic data in the rhesus monkey in the efficacy study of inhalational anthrax [is used] to demonstrate that the desired systemic exposure achieved in humans after the anticipated dosage regimen can actually be achieved and is effective in the animal model in preventing inhalational anthrax infection and consequent mortality." (Anthrax Draft Guidance at p. 10).

7. TYSABRI (natalizumab)

FDA approved Tysabri on November 23, 2004 for treating relapsing-remitting multiple sclerosis (RRMS), relying upon the reduction in MS relapse rates at one year as the surrogate endpoint. Applying the terms of the Draft Guidance, this would be an intermediate clinical endpoint that would be reasonably likely to predict the benefit at two years. All previous MS therapies were approved on the basis of two-year relapse rate reduction and "the clinical meaningfulness of a decrease in the relapse rate through only one year is uncertain." (Medical Review, Nov. 23, 2004, p. 6).

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

Relapsing-remitting multiple sclerosis is a serious, life-threatening condition.

b. Rarity of the Condition

While Tysabri was not designated an orphan drug for RRMS, the statutory threshold for qualifying as an orphan drug was, in part, set in the 1984 amendment to the Orphan Drug Act specifically to include all of multiple sclerosis as an orphan disease, not just the subset of RRMS. This was because, in considering how to amend the original 1983 Orphan Drug Act to make it less difficult to garner orphan drug designation, key Senators caucused with the National Organization for Rare Diseases (NORD) and mutually determined that the maximum number of Americans with a condition which would still qualify as an "orphan" would be 200,000. This number was chosen, specifically, to make sure that MS would be an "orphan" disease, and in 1984 there were just under 200,000 Americans diagnosed with MS. However, soon after FDA approved the first therapy for multiple sclerosis (Betaseron in August 1993, which was also the first non-AIDS Subpart H approval), the number of Americans diagnosed with multiple sclerosis dramatically increased. So, while Tysabri was never designated as an orphan drug for RRMS, the commentators, fully cognizant of the intent of the 1984 orphan drug amendment, view Tysabri as, nevertheless, falling within the "penumbra" of orphan drug status and score Tysabri a "1" on rarity.

c. Lack of Available Therapy

"Accelerated approval requires that the new drug provide evidence of the potential to address an unmet medical need. Many MS patients continue to have exacerbations while taking one of the available first-line MS therapies. None of the currently available therapies have proven efficacy when used as add-on therapy. [One of the two pivotal Tysabri studies] provides evidence that [Tysabri] is effective as add-on therapy for subjects who continue to have relapses while on a first-time therapy (Avonex). Therefore, [Tysabri] has the potential to address an unmet medical need." (Medical Review at p. 6).

d. External Expertise

FDA did not rely on an advisory committee during its initial review of Tysabri. However, Tysabri was withdrawn from the market by the manufacturer in February 2005 after three patients developed progressive multifocal leukoencephalopathy (PML). Subsequently, FDA convened an Advisory Committee to consider the reintroduction of Tysabri in March 2006. Furthermore, FDA had convened and considered the input from several earlier advisory committees on other multiple sclerosis therapies.

Part 2. Understanding of the Disease Process

"Multiple Sclerosis is a chronic, inflammatory, possibly autoimmune, demyelinating disease of the central nervous system." (Medical Review at p. 11). Note that the FDA review status that multiple sclerosis may be "possible autoimmune." Given that Tysabri's mechanism of an action is as an immunomodulator, having a more definitive view of the causative role of autoimmunity in the pathophysiology of this disease would have been more compelling.

Part 3. Understanding of the Relationship Between the One-Year Relapse Rate and Two-Year Relapse Rate

The effect of [Tysabri] on relapse rate in [the pivotal study on Tysabri's use as first-line therapy] was approximately twice the effect observed with current first-line drugs for this indication. Such comparisons of different agents across studies are problematic . . . However, the magnitude of [Tysabri's] effect is sufficient that the effect at one year is reasonably likely to predict a clinical benefit at two years." (Medical Review at p. 102).

As for understanding the relationship of drugs in the same or closely-related pharmacologic class on rate and extent of exacerbations at one year of treatment as predictive of their two year effectiveness, at the time of Tysabri's approval, there were four other approved immunomodulators approved for treatment of MS: Betaseron, Avonex, Rebif and Copaxone. While each of these was approved on the basis of two-year studies of impact on reducing rate and extent of MS exacerbations, their impacts after one year of therapy, while generally more modest than at the end of two years, were predictive of their two year results.

Part 4. Clinical Evidence on One-Year and Two-Year Relapse Rates

"For other MS products, FDA has required two-year data . . . A salutary effect on relapse rate at one year is not a validated surrogate for benefit at two years. However, the apparent treatment effect of [Tysabri] with respect to relapse rate at one year is unprecedented in the MS field, and its magnitude is reasonably likely to predict clinically meaningful effectiveness at two years. If, in fact, the benefit on clinical relapses is shown to be durable through two years, the product may be substantially more efficacious than

currently approved MS therapies. It is possible, however, that the magnitude of [Tysabri's] effect on relapse rate, when assessed through one year, may substantially overestimate [Tysabri's] benefit on relapse rate through two years . . . In particular, the treatment effect appears to wane with the development of [anti-Tysabri] antibodies, which may increase in time." (Medical Review at p. 53) (emphasis added).

8. LUVERIS (lutrepinalfo)

On October 8, 2004, FDA approved Luveris for stimulating follicular development in infertile hypogonadotropic hypogonadal women with profound LH deficiency (LH < 1.2). "The Division Director further concluded that in this orphan population of women with severe LH deficiency (LH <1.2), the surrogate endpoint of follicular development (as defined by the Sponsor) was reasonably likely to predict clinical benefit [with respect to pregnancy] . . . (Medical Review #1, Oct. 6, 2004, p. 2).

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

The inability to ovulate due to profound luteinizing hormone (LH) deficiency includes, among other serious consequences, the inability to become pregnant. "The Director believes that infertility in the context of hypogonadotropic hypogonadism and profound LH deficiency is a serious condition with very limited options for pregnancy." (Medical Review #2, Oct. 6, 2004, p. 7).

b. Rarity of the Condition

Luveris was granted orphan drug designation by FDA on October 7, 1994.

c. Lack of Available Therapy

"Luveris would be the only LH-alone product . . . on the U.S. market. There are no approved drug products that have the indication of treatment of infertility in women with hypogonadotropic hypogonadism." (Medical Review #3, Sept. 28, 2004, p. 17).

d. External Expertise

The Reproductive Health Advisory Committee considered Luveris on September 30, 2003. "After hearing presentations from experts in Reproductive Endocrinology on the subject of female hypogonadotropic hypogonadism . . . the Committee voted 15 to 0 that the Sponsor's data did not demonstrate efficacy for Luveris in ovulation induction when the primary endpoint was ovulation rate. The Committee voted 8 to 7 that the Sponsor's data demonstrated efficacy for Luveris in ovulation induction when the primary endpoint was follicular development. Finally, the Committee voted 11 to 3 . . . that the Sponsor's data demonstrated efficacy for Luveris for follicular development when the primary endpoint was follicular development." (Medical Review #1 at p. 2) (emphasis in original).

Part 2. Understanding of the Disease Process

FDA's medical review suggests that the disease process is complex and multifactorial: "the role of LH in hypogonadal female infertility patients is clouded by the spectrum of clinical disorders that cause hypogonadotropic hypogonadism with the

differing patterns of gonadotropin secretion may further confound clinical outcome results." (Medical Review #3 at p. 19).

Part 3. Understanding of the Relationship Between Follicular Development and Fertility

"The Division believed that although both follicular development and ovulation are surrogates for pregnancy (the clinically meaningful outcome), ovulation is more temporally proximate to pregnancy and therefore more appropriate as a surrogate." (Medical Review #2 at p. 5). Nevertheless, follicular development is on the causal pathway, as is ovulation. However, there was no epidemiological evidence cited in the FDA review documents linking follicular development to pregnancy.

As for understanding the relationship of drugs in the same or closely-related pharmacologic class on follicular development: "Recognition of the therapeutic potential of gonadotropins began in the 1950's with the extraction and purification of human menopausal gonadotropins (both follicle stimulating hormone and luteinizing hormone) from both human pituitaries and urine sources. Successful clinical pregnancies resulting from the use of these human-derived gonadotropins were first reported in the 1960's. In the 1990's cells that are capable of producing biologically-active LH in culture produced LH. This recombinant derived LH is from in vitro cultured cells." (Medical Review #3 at p. 17).

Part 4. Clinical Evidence of Luveris on Follicular Development and Fertility

"The primary efficacy parameter for both Studies 6905 and 6253 was follicular development as defined by three co-primary endpoints (follicle size as measured by ultrasound, pre-ovulatory serum estradiol levels and mid-luteal progesterone levels). The Sponsor's analysis demonstrated that in Study 6253, 75 IU of Luveris was numerically better than 25 IU of Luveris or placebo for follicular development in women with LH <1.2 IU/L." (Medical Review #1 at p. 3). "The Division's analysis of Study 6905 demonstrated . . . the placebo was as efficacious as 75 IU of Luveris. Therefore, in the opinion of the Division, Luveris was not demonstrated to be effective." (Medical Review #1 at p. 3).

Therefore, the Sponsor planned and conducted a third study, Study 21008, with follicular development as the Sponsor's prespecified primary endpoint, despite the Division's recommendations that ovulation rate be the primary endpoint. The Sponsor's "evaluable patient analysis of Study 21008 demonstrated that 67% of patients receiving 75 IU of Luveris achieved follicular development compared to 20% of patients receiving placebo." (Medical Review #1 at p. 4). "The Director [Dr. Shames] concluded that the results from Studies 21008 and 6253 provide substantial evidence that Luveris 75 IU, when administered concomitantly with FSH, induces follicular development in this population of infertile women. These studies, however, do not demonstrate a positive effect on clinical pregnancy, etc. Study 21415 evaluated titrable FSH dosing with the

dose of Luveris fixed at 75 IU and demonstrated a 36% clinical pregnancy rate after one cycle. While reassuring, this finding is not definitive because there was no placebo comparator group in Study 21415, and the finding has not been replicated in a second trial." (Medical Review #1 at p. 7). Study 21415 also reported follicular development rates of 63% "in all cycles combined." (Medical Review #3 at pp. 29-30). Therefore, in Study 21415, there was within-study clinical evidence both on follicular development, the surrogate, as well as on pregnancy, the ultimate clinical outcome.

9. FABRAZYME (agalsidose beta)

FDA approved Fabrazyme on April 23, 2003 to treat Fabry's disease. This approval was based on a surrogate endpoint of near-elimination of all accumulation of enzyme in renal capillary endothelium, one type of vascular endothelium.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of Condition and Lack of Available Therapy

"[W]ith age, the principal manifestations of concern in Fabry's disease are in the kidney, heart, and brain. Renal disease is manifested by proteinuria, hypertension, and progressive azotemia; the principal cause of death in Fabry's disease in the past was renal failure . . . The median age of death for homozygous males is 50 years." (Medical Review #1, Apr. 21, 2003, p. 4).

b. Rarity of the Condition

Fabrazyme was designated an orphan drug on January 19, 1988.

c. Lack of Available Therapy

"There is no specific treatment for Fabry's disease." (Medical Review #1 at p. 4).

d. Use of External Expertise

"Vessels (capillaries in this case) that are essentially near-normal in appearance that may well lead to an altered development of vascular occlusion, and thus to an alteration in expression of the clinical impairments of the disease. The [January 2003] Advisory Committee has also supported this assessment of the potential impact of near-absence of capillary accumulation, as well as concurring that the evidence submitted by [the Sponsor has] demonstrated this effect on capillary endothelium." (Medical Review #2, Apr. 23, 2003, p. 3).

Part 2. Understanding of the Disease Process

"The underlying basis of Fabry disease is well understood; it is an X-linked enzyme deficiency leading to a lipid storage disorder. Lipid storage occurs in a wide variety of cell types, and consequently there are a wide variety of signs and symptoms from different organ systems . . . However, [there] is widespread belief that a number of the organ injury manifestations are related to vascular injury. It is believed that while this may not be the sole pathologic process, progressive substrate accumulation within vascular walls will ultimately lead to local vessel occlusion, with organ impairment as a consequence." (Medical Review #2 at p. 3).

Part 3. Understanding of the Relationship Between Near-Elimination of Substrate in the Renal Capillary Endothelium and the Outcomes of Fabry's Disease Including Renal Failure and Mortality

"Vascular injury does appear to be an important mechanism of promoting the progressive organism impairment, and substrate accumulation within vascular walls is the basis for this. The exact (quantitative) relationship between the amount of substrate accumulation and the degree or rate of vascular ischemia is unknown and not addressed in any information presented by [the Sponsor]. It is unknown if reducing substrate accumulation by half might show vascular injury by half, or if there is a threshold effect, wherein some specific amount of accumulation will invariably lead to vascular occlusion and thus no change in the clinical expression of the disease. However, by focusing upon a near-elimination of all accumulation within a specific cell type [the Sponsor's] data appear to overcome these concerns." (Medical Review #2 at p. 3).

"Following FDA requests to [the Sponsor], additional data were submitted which demonstrated that while not all cell types show a marked decrease in substrate accumulation (e.g., renal podocytes, with a limited degree of reduction in substrate accumulation) there are a variety of cell types with moderate and several that show marked reduction in substrate accumulation." (Medical Review #2 at p. 1).

As for understanding the relationship of drugs in the same or closely related pharmacologic class on near-elimination of substrate in specific cell types and Fabry's disease, there were no other drugs approved at that time, and there was only one other drug with controlled clinical studies in Fabry's disease, Replagal.

Part 4. Clinical Evidence on Substrate Reduction in Certain Cell Types and Fabry's Disease Outcomes

a. Substrate Reduction

The primary endpoint in the 58 patient, placebo-controlled randomized trial was clearance (that is, elimination) of kidney intestinal capillary endothelium GL-3 inclusions (or substrate). While none of the 29 placebo subjects achieved a score of "zero" GL-3 inclusions over the 5 month duration of the trial, 20 of the 29 Fabrazyme subjects "cleared" all substrate ($p<0.001$) (Medical Review #1 at p. 30).

b. Clinical Outcomes

"The clinical trials failed to show clinical benefit on a wide range of tests of neurologic, renal, and cardiac function. This finding weakens confidence in the clinical importance of the reduction of kidney interstitial capillary endothelial cell GL-3 [enzyme substrate] that constituted the primary endpoint of the pivotal trial." (Medical Review #1 at p. 74).

In the pivotal study, there was only one secondary endpoint that assessed a clinical outcome, and that was pain. In the five ways in which pain was assessed, the placebo group outperformed the treated group in 4 of the 5 measures of pain. (Medical Review #1 at pp. 35-36). There were tertiary endpoints that assessed clinical outcomes and in eight of these, there were no numerical between-group differences, and in one measure of neuropathy, the placebo group fared somewhat better and in two measures (symptom-free days and episode-free days), the Fabrazyme group fared somewhat better. Of interest, renal function was assessed by Inulin-GFR and by serum cystatin-C, and on both of these measures of renal function, there were essentially no numerical differences between placebo and Fabrazyme groups. Among "other" endpoints, there were ophthalmic assessments, and "the ophthalmological findings, like the tertiary endpoints, did not show a clinical change effected by the product." (Medical Review #1 at pp. 39-42).

10. REMODULIN (trepostinil)

The May 21, 2002 approval of Remodulin for treating pulmonary hypertension (now referred to as pulmonary arterial hypertension or PAH) was based on an intermediate clinical endpoint of 6-minute walk (6MW) test, a measure of exercise capacity that is a clinical endpoint, but not the ultimate clinical outcome of this serious disease.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

PAH is a serious, life-threatening condition.

b. Rarity of the Condition

FDA designated Remodulin for PAH an orphan drug on June 4, 1997.

c. Lack of Available Therapy

The only other therapy approved before Remodulin was Flolan, whose labeling states that "8 of 40 patients receiving standard therapy alone died, whereas none of the 41 patients receiving Flolan died ($p=0.003$)." (Medical Review, Mar. 28, 2001, p. 55). This same Medical Review states also that Flolan's "use is difficult and inconvenient. The infusion of Flolan requires the insertion of an indwelling central catheter with the . . . subsequent risk of catheter infection . . . Any inadvertent interruption of the infusion is potentially life-threatening." (Medical Review at p. 55).

d. External Expertise

The Cardiovascular and Renal Drugs Advisory Committee, on August 9, 2001, voted 6 to 3 in favor of approving Remodulin.

Part 2. Understanding of the Disease

The pathophysiology of PAH is well-understood.

Part 3. Understanding of Relationship Between 6MW Results and Clinical Worsening of PAH

Exercise capacity as measured by the 6MW test was judged by FDA as reasonably likely to predict clinical benefit, which was determined to be clinical worsening of PAH symptoms. Confirmation of FDA's decision to rely upon the 6MW test results as predictive of clinical benefit was later seen in that this same measure, 6MW, was the basis for the approval of several subsequent PAH therapies, especially after this Sponsor's successful completion of its Phase 4 confirmatory trial established Remodulin's effect on preventing clinical worsening ($p<0.001$). The Sponsor's Phase 4

trial results on clinical worsening demonstrated the positive predictive value of the 6MW test results with Remodulin.

Part 4. Clinical Evidence on 6MW and on Clinical Worsening or Mortality

The primary endpoint of the pivotal trials was "change in [6 minute] walking distance from baseline at the end of week 12 . . . The database was to be considered demonstrating a benefit for [Remodulin] if either both studies where by themselves significant at the $p<0.049$ or if one study was significant ($P<0.049$) and the pooled studies had a p-value of less than 0.01 . . . Neither of the studies demonstrated a p-value of <0.049 ($p=0.06$ for both studies), although the pooled studies demonstrated an overall p-value of <0.01 ($p = 0.006$ for the pooled studies)." (Medical Review at p. 10). In the pivotal [Remodulin] studies, the drug demonstrated no mortality benefit. (Medical Review at p. 14).

11. CIPRO (ciprofloxacin hydrochloride)

On August 30, 2000, FDA approved a supplemental NDA for Cipro for prophylaxis after exposure to inhalational anthrax. There was a two-part or "compound" surrogate for this approval in that FDA concluded: (1) that Cipro reduced "the rate of death due to anthrax over control in the macaque monkey model," (Statistical Review, Aug. 16, 2000, p. 3), and (2) "that [Cipro] serum concentrations achievable in human populations reach or exceed those associated with improved survival in animals exposed to aerosol challenge with spores of *B. anthracis* [in that] serum concentrations in both human and animal populations consistently exceed the MIC_{90} of the causative organism."[11] (Medical Review, August 31, 2000,[12] p. 34).

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

"The mortality rate of inhalational anthrax is as high as 80-100% . . ." (Statistical Review, August 16, 2000, p. 1).

b. Rarity of Condition

"[I]nhalational anthrax is extremely rare. There have been only approximately 20 cases in the United States in the past 100 years . . . For these two reasons, the rarity of disease and the extremely high mortality rate, a clinical study is not feasible." (Statistical Review at p. 1). Although the prevalence of inhalation of anthrax is sufficiently low, the sponsor did not seek orphan drug designation.

c. Lack of Available Therapy

"There are drugs with currently approved labeling by FDA for disease associated with *B. anthracis.* Labels for penicillin, tetracycline, doxycycline, and minocycline products list *B. anthracis* among the organisms susceptible to these agents. None of these agents is indicated specifically for post-exposure prophylaxis for disease caused by inhaled *B. anthracis.*" (Medical Review at p. 2).

d. External Expertise

The Anti-Infective Drug Products Advisory Committee on July 28, 2000 unanimously voted "yes" to the question: "Do the data presented support the safety and

---

[11] Obviously, there was no requirement for a Phase 4 confirmatory study, and the commentators hope there is never any open-label uncontrolled anecdotal evidence obtained.

[12] The Medical Review was completed, signed and dated the day after the approval.

efficacy of [Cipro] for post-exposure prophylaxis of inhalational anthrax?" (Medical Review at p. 33).

Part 2. Understanding of the Disease Process

"The inhalational form of the disease, which affects the mediastinal lymph nodes, other organs of the reticuloendothelial system and the central nervous system, is considered the most likely clinical entity resulting from the intentional use of an aerosolized preparation of the spores of *B. anthracis*." (Medical Review at p. 2).

Part 3. Understanding of the Relationship Between the Monkey Studies and Human Mortality; and Part 4. Clinical Evidence

First, "the p-value comparing the death rate of [Cipro] to that of control is highly significant (p=0.0011) showing that the treatment with [Cipro] significantly reduces the rate of death due to anthrax over control in the macaque monkey model." (Statistical Review at p. 3).

Second, as for comparative monkey/human exposure levels, the data "demonstrates that [Cipro] peak and trough serum concentrations achieved in the Rhesus monkey are reached or exceeded in human populations receiving the doses recommended for the post-exposure inhalational anthrax. Peak and trough concentrations reported in both monkey and human populations are shown to consistently exceed 0.06 mcg/ml, the value of the MIC_{90} for *B. anthracis*." (Medical Review at p. 10).

As for understanding the relationship of drugs in the same or closely related class on the compound surrogate, see above discussion under 1.c. regarding other drugs approved for anthrax, but note that none had evidence that assessed their utility specifically against post-exposure inhalational anthrax.

Monkey survival data was one part of this unusual compared surrogate; see the discussion above. Moreover, the Medical Reviewer stated: "There have been no prospective studies performed that link clinical outcome to drug exposure for infection with *B. anthracis*. However, in general, when there is a demonstrated relationship between plasma concentrations of drug and response, pharmacokinetic data may be used as one way to relate dose and possible outcome." (Medical Review at p. 14).

12. CELEBREX (celecoxib)

FDA's December 23, 1999 approval of a supplemental NDA for Celebrex to reduce the risk of colorectal cancer in patients with familial adenomatous polyposis (FAP) was based on a surrogate endpoint which was reduction in colorectal polyps.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

"The average life expectancy for patients with untreated FAP has been estimated to be 42 years." (Medical Review, Dec. 22, 1999, p. 25).

b. Rarity of the Condition

"The frequency of the FAP gene has been estimated on the basis of disease prevalence to be 1 in 5,000 to 1 in 7,500." (Medical Review at p. 22). Although the prevalence of FAP is sufficiently low, the Sponsor did not seek orphan drug designation.

c. Lack of Available Therapy

"Surgical therapy is the only acceptable option for patients with FAP after colonic polyps have been detected." (Medical Review at p. 26).

d. Use of External Expertise

Here are the recommendations of the Oncologic Drugs Advisory Committee that met on December 14, 1999:

i. Do you believe that a reduction in colorectal polyps count in FAP patients in focal areas of some magnitude is "reasonably likely" to predict benefit?

Yes: 13 No: 0 Abstain: 2

ii. Do you believe that the observed reduction (about 25% at 6 months) is likely to predict benefit in FAP patients?

Yes: 12 No: 0 Abstain: 3

iii. Do you recommend approval of Celebrex under the accelerated approval rule for treatment of FAP?

Yes: 14 No: 0 Abstain: 1

(Medical Review at pp. 76-77).

Part 2. Understanding of the Disease Process on Polyp Counts on Colon Cancer

"FAP is characterized by the presence of hundreds to thousands of colorectal adenomatous polyps and the inevitable development of colon cancer . . . The disease results from germ line mutations of the APC gene . . . The APC gene is thus believed to be a tumor suppressor gene." (Medical Review at pp. 22-23). "A significant body of evidence suggests that cellular expression of COX-2 is prominent in several types of tumors, including colon . . . as well as pre-cancerous changes such as Barrett's esophagus, the adenomatous polyp and actinic keratosis." (Medical Review at p. 15).

Part 3. Understanding of the Relationship Between Reducing Polyp Counts and Colon Cancer

"Celebrex was evaluated in two models of colon cancer. The Min mouse model represents a genetic model of human FAP . . . Adenomas and adenocarcinomas of the colon can be chemically induced in rats by administration of azoxymethane." Celebrex was shown to prevent or inhibit colorectal tumor development in both of these animal models. (Medical Review at pp. 16-17).

As for understanding the relationship of drugs in the same or a closely-related class on FAP polyp counts, "studies have shown that Sulindac, one of the non-selective NSAIDs, induces apoptosis . . . Recent study of COX-2 inhibitors showed that inhibition of COX-2 produced sequential increases in arachidonic acid and ceramide, the latter a potent stimulant of apoptosis. Furthermore, *in vitro* evidence exists that angiogenesis is regulated by COX-2 expression in colon cancer cells. Therefore, another mechanism by which tumor growth may be inhibited by COX-2 inhibitor is through blockade of angiogenesis and tumor vascularization." (Medical Review at pp. 15-16).

Part 4. Clinical Evidence on Polyp Counts and on Colon Cancer

"A single, randomized, double-blind, placebo-controlled study has been submitted. A total of 83 patients received treatment with either placebo, Celebrex 100mg BID, or Celebrex 400mg BID for 6 months (with a 1:2:2 randomization) . . . The mean reduction in colorectal polyps count was 28% on the Celebrex 400mg BID arm, 15% on the Celebrex 100mg BID arm and 5% on placebo. Only treatment with Celebrex 400mg BID was associated with a statistically superior mean reduction in polyp counts, with $p=0.003$." (Medical Review at pp. 1-2). In a six-month study there were, as expected, no cases of colon cancer in any arm of the trial.

13. SYNERCID (dalfopristin/quinupristin)

The FDA approval of Synercid on September 21, 1999 was for treating patients with vancomycin-resistant *Enterococcus faecium* (VREF) and was based on a surrogate showing of clearance of the VREF bacteremia.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

"The mortality rates in both [pivotal] studies [were] approximately 50%." (Statistical Review, Mar. 5, 1998, p. 17).

b. Rarity of the Condition

The Sponsor has no intention of developing Synercid for this use, but a "rise in the United States in both the number of nosocomial infections due to *E. faecium* and in the proportion of strains of this pathogen found to be vancomycin-resistant, led to increasing requests for the emergency use of Synercid." (Medical Review, Aug. 21, 1998, p. 2). Synercid appears not to have been granted orphan drug designation. Given the Sponsor's reluctance to submit an NDA for this use, the Sponsor likely never had applied for designation, even though the condition was rare.

c. Lack of Available Therapy

Those patients who enrolled in the two pivotal trials were only those "infected with VREF who did not have any other therapeutic option." (Statistical Review at p. 2).

d. External Expertise

On February 19, 1998, the Anti-Infective Drugs Advisory Committee voted 9 to 1 in favor of approval of Synercid for VREF.

Part 2. Understanding of the Disease Process

The understanding of the pathophysiology of infections with vancomycin-resistant strains of *Enterococcus faecium* is well-known.

Part 3. Understanding of the Relationship Between Clearance of the VREF Bacteremia and Mortality (and Other IDSA/FDA Guideline Clinically Meaningful Endpoints)

"The VREF literature is clear that VREF bacteremia . . . should be treated and that clearance of VREF from the bloodstream can be seen as beneficial to the patient . . . There is consensus that bacteremia should be treated. Thus, while clearance of bacteremia is not a clinical benefit by itself, it can be seen as likely to predict clinical benefit. Thus, it is proposed that the clearance of VREF bacteremia be viewed as a

surrogate endpoint likely to predict clinical resolution of infection." (Medical Review at p. 32).

Part 4. Clinical Evidence on VREF Bacteremia Clearance and Mortality

FDA concluded that the four emergency use VREF studies did not provide evidence of an improvement in mortality or resolution of infection due to a host of issues. None of these four studies had a concurrent control and, while FDA had advised that the lack of concurrent control would be acceptable because it would be unethical to include a placebo arm, FDA had stipulated that the studies either: (1) had to show a "dramatic improvement in overall mortality as compared to a historical perspective" (Medical Review at p. 30) and these studies did not (these four studies had mortality rates of 48.8%, 49.5%, 53.8% and 54.0% compared to the VREF literature reporting "all-cause" mortality rates in the range of 30% to 70%) (Medical Review at p. 18); or (2) had to have a historical control and this was not established (Medical Review at pp. 18-19).

While two of the four studies, according to the FDA Medical Reviewer, established clearance of VREF bacteremia, only 18% of the patients in these emergency use studies were "evaluable" due primarily to missing data, and there was a low response rate as well. (Medical Review at pp. 19-20, 29-32). In addition, "in the unevaluable patients who died on therapy but with negative blood cultures, there is 'apparent' clearance of the organism." (Medical Review at p. 32).

14. REMICADE (infliximab)

The August 24, 1998 FDA approval of Remicade to treat patients with Crohn's disease was based on an intermediate clinical endpoint of a clinical response defined as a reduction in the Crohn's Disease Activity Index (CDAI) of at least 70 points at the 4-week evaluation.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

"The prognosis for Crohn's disease is generally unfavorable . . . The mortality rate increases with the duration of disease and most likely ranges from 5% to 10%. Most deaths occur from peritonitis and sepsis." (Medical Review, July 10, 1998, p. 4).

b. Rarity of the Condition

"In . . . the United States, the prevalence is estimated at 20 to 40 per 100,000." (Medical Review at p. 3). Remicade was designated as an orphan drug on November 14, 1985.

c. Lack of Available Therapy

The FDA Medical Review surveys all the therapies being used and at the time, no robustly effective therapies were available. "Because its cause is unknown, medical management of the disease is largely empirical and is designed to reduce inflammation." (Medical Review at p. 5).

d. External Expertise

On May 28, 1998, the Anti-Infective and Gastrointestinal Drug Advisory Committees voted unanimously in favor of approval for both: treatment of patients with moderate-severe inflammatory disease refractory to conventional therapy, and treatment of patients with fistulizing Crohn's disease for the reduction in the number of draining enterocutaneous fistula(s).

Part 2. Understanding of the Disease Process

"Crohn's disease most likely represents a heterogeneous group of disorders. After much effort that has focused on the identification of a specific pathogenic cause, it is being recognized that disease manifestations could result from a combination of any, or all of, a number of factors." (Medical Review at p. 2).

Part 3. Understanding of the Predictive Potential of a 70 Point Change in CDAI at Week 4 on Crohn's Disease

"Pathologic review of biopsy . . . often can aid in . . . measurement of extent and severity of disease. Pathologically, Crohn's disease is described as a transmural disease with focal or microscopic skip areas of inflammation in the lamina propria. The degree of inflammation in the most heavily involved area often is an accurate assessment of the severity of disease . . . Disease activity indices are used to objectively measure the activity of disease for judgment of response in clinical trials. The [CDAI] was developed . . . [in] 1979 . . . to objectively assess response to therapy . . . Although imperfect and cumbersome, e.g., requirement of recording of symptoms for 7 days and for hematocrits, the CDAI remains the most commonly [used] index." (Medical Review at p. 4).

As for understanding the relationship between drugs in the same pharmacologic class, Remicade is a chimeric monoclonal antibody to Tumor Necrosis Factor (TNF). As such, Remicade was the first of this kind in a new class of immunomodulatory drugs. Other immunomodulatory drugs, including azathioprine, mercaptopurene, cyclosporine, and methotrexate were accepted for use for long-term treatment of some Crohn's patients. "The mechanism of action of these drugs may involve inhibition of lymphocyte function, primarily that of T cells." (Medical Review at p. 5). As such, they have a different mechanism of action than Remicade.

Part 4. Clinical Evidence on CDAI and on Long-Term Clinical Benefit

Study T16, a placebo-controlled, dose-ranging (n=108) study, "was designed as a Phase 2 trial to determine an effective dose in the acute treatment of patients with active Crohn's disease not responding to immunosuppressant therapy and to explore maintenance therapy with a single dose in patients who responded initially. This clinical trial became the pivotal trial for licensure of [Remicade] for this indication." (Medical Review at p. 10). "65.1% of the [Remicade] treated patients achieved a clinical response ($\geq$ 70-point reduction from baseline in the CDAI) at the week 4 evaluation compared to 16.7% of the placebo patients ($p<0.001$) . . . There was no apparent relationship between [Remicade dose] [5mg/kg, 10mg/kg, 20mg/kg] and the proportion of patients responding; the highest clinical response was observed in the 5mg/kg dose group (81.5%; $p<0.001$ vs placebo)." (Medical Review at p. 19).

In the Medical Review's Summary Conclusions on the Review of the Safety and Efficacy Data, the Medical Reviewer stated that "[t]he Sponsor has presented phase 2 clinical data results to support licensing of a potent, novel immunomodulating agent for the management of patients with Crohn's disease, a chronic debilitating disease . . . The number of patients with moderate to severe disease who have received the proposed dose of 5mg/kg . . . is very low (n=28) and no patients have received chronic retreatment with 5mg/kg every 8 weeks as proposed in the original submission. The effects of a single dose [last] approximately 12-16 weeks, compatible with the half-life of the compound.

For patients with fistula, although the majority of patients experienced stoppage of drainage in two weeks, there are no data on internal healing of the fistula canal. Once [Remicade] was stopped the effect of therapy was lost. In summary, there are inadequate data to support the long-term benefit of [Remicade] in patients with either fistulizing or moderate/severe disease." (Medical Review at p. 81).

From the conclusions of the Medical and Statistical Reviews, there appear to have been some concerns among FDA Reviewers as to the appropriateness of the short-term (CDAI improvement after 4 weeks) surrogate endpoint as being adequate to predict long-term benefit in a chronic disease. The conclusion of the Statistician on Study T16 in moderate to severe Crohn's disease patients was redacted from the publicly available version of the Statistician's Review. However, there was a second Phase 2 study in patients with Crohn's disease with fistula, Study T20, which is referred to in the conclusions of the Medical Review. From the information in the Statistician's Review of Study T16 that was made publicly available, it would seem that the Statistician's conclusions with respect to Study T20 may have closely paralleled those for Study T16. With respect to Study T20, here are the Statistician's conclusions: "Although the differences in response rates between the placebo group and the [Remicade]-treated groups were statistically significant, questions remain about the durability of response. Patients received doses at weeks 2, 4, and 6, but this dosing strategy should be thought of as one-time dosing. After 6 months of follow-up, the drug effect had disappeared and the proportion of responding patients in the placebo arm was similar to the proportions in the treatment arms. The data suggest, therefore, that although this agent has an initial beneficial effect on Crohn's disease, a single set of doses is unlikely to provide durable benefit in this chronic disease. There are no data to assess chronic use of [Remicade] for this indication. There is no information regarding the formation of neutralizing antibodies (HACA) with repeated dosing and how this may affect the efficacy of this product. There is also no safety data to allay concerns of a possible increase in malignancies or serious infections. The Agency should carefully weigh the observed early benefits seen with this product against the paucity of information regarding the safety and efficacy of repeated use for this chronic indication." (Statistical Review, Aug. 5, 1998, p. 13).

15. PRIFTIN (rifapentine)

On June 22, 1998, FDA approved Priftin for treating pulmonary tuberculosis (TB) and this approval was based on a surrogate of a 6-month relapse rate as contrasted with the standard 2-year relapse rate information for a traditional approval.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

"[TB] is the leading infectious cause of morbidity and mortality worldwide." (Medical Review, June 19, 1998, p. 5).

b. Rarity of the Condition

"In 1990, there were 25,701 new cases of TB reported in the [U.S.]" (Medical Review at p. 5). Priftin was designated as an orphan drug on June 9, 1995.

c. Lack of Available Therapy

"During development of rifapentine for TB, the applicant was encouraged to submit 6 month follow-up data from one study, under the accelerated approval regulations (21 CFR 314 Subpart H). There is a need for new anti-tuberculosis medications, and for medications which will potentially increase the adherence to dosing thereby decreasing the potential for the development of resistant organisms. It was anticipated that rifapentine would be such an agent. Six-month relapse data would serve as a surrogate for two-year relapse data predictive of long term clinical benefit." (Medical Review at p. 8). FDA had previously approved rifampin for use in treating TB.

d. Use of External Expertise

At the Anti-Viral Advisory Committee Hearing on May 5, 1998, "the committee voted to recommend approval of [Priftin] for the treatment of pulmonary tuberculosis, with only one dissenting vote." (Medical Review at p. 61).

Part 2. Understanding of the Issues

In this case, the pathophysiology of TB is well-understood.

Part 3. Understanding of the Predictive Potential of a Six-Month Relapse Rate on Two-Year Relapse Rate and on Mortality

The Medical Review stated that: "It is expected that the majority of relapses will occur by 6 months of follow-up, however, the 'gold standard' is 2-year relapse rate." (Medical Review at p. 19). However, the pattern of relapses for [Priftin] does not appear to reflect the same showing of relapses in the latter half of six-month follow-up that was seen for rifampin in the pivotal study. See discussion of results under Section 4.

Part 4. Clinical Evidence on Six-Month and Two-Year Relapse Rates

The single pivotal trial was an open-label, randomized, two-arm parallel, rifampin-controlled trial with 570 patients in the modified ITT analysis. "The primary efficacy endpoint for this accelerated approval review was treatment outcome at the end of 12 months (6 months of active treatment + 6 months of follow-up). This was a binary variable with success defined as achieving a negative sputum culture during active treatment and sustaining it to the end of [6] months of follow-up." (Statistical Review, July 27, 1998, p. 3).

"There is essential equivalence for [negative sputum culture] rates at the end of [the 6-month active treatment] between the rifampin [83% negative sputum cultures] and [Priftin] [88%] arms." (Medical Review at p. 39). However, "[t]here is a statistically significant difference between the treatment arms for relapse . . . The risk is 5% for rifampin . . . and 11% for [Priftin]." (Medical Review at p. 40). The Statistical and Medical Reviews agree that while 10 of the 11 relapses on rifampin occurred within the first 6 months of follow-up, 7 relapses occurred in the [Priftin] arm at time points between 6 and 12 months of follow-up. (Note: While the endpoint was at 6 months of follow-up, almost all subjects had had 12 months of follow-up, so FDA analyzed the 12 months of follow-up data as well and noted that the Priftin arm continued to experience sizable numbers of relapses beyond the first 6 months of follow-up, which was much different than the pattern of relapses observed for rifampin).

Despite the above discrepancy between the rifampin and Priftin arms in relapse rate beyond 6 months, the FDA reviewers seemed (as well as the Advisory Committee members) to believe that this may reflect lack of optimized dosing of Priftin, rather than a lack of confidence in the prognostic surrogate of 6-month relapse rate predicting 2-year relapse rate, and eventually, survival. However, at the time of approval there appear to be no clinical evidence of Priftin on 2-year relapse rate or on mortality.

16. SULFAMYLON (mafenide acetate)

FDA approved Sulfamylon on June 5, 1998, "to control bacterial infection when used under moist dressings over meshed autografts on excised burn wounds." The approval was based on an intermediate clinical endpoint of evidence derived from patients who were burned over up to 20% of their total body surface area (TBSA) with a Phase 4 commitment to conduct a confirmatory trial in patients with 20% to 60% TBSA thermal injuries.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition

The Medical Review commenting on the results of the single pivotal trial (done exclusively in children) observed the following: "It is remarkable that so many of these severely burned children survived to leave the hospital . . . It is not unexpected that survival rates fall as TBSA burned increases." (Medical Review, Sept. 23, 1997, p. 17). Large [TBSA] burns are serious and life-threatening." (Medical Review at p. 49).

b. Rarity of the Condition

The number of persons in the country in need of such care is small, thankfully, very small. FDA designated Sulfamylon as an orphan drug for this use for two different sponsors at separate times: on August 29, 1985 and on July 18, 1990. (Medical Review at p. 3).

c. Lack of Available Therapy

"There is no existing approved treatment for these burn patients who require excision and meshed autografts." (Medical Review at p. 50).

d. External Expertise

"Sulfamylon [was] discussed by the FDA Anti-Infective Drug Products Advisory Committee [on July 24, 1996]. The Committee concluded that since topical antimicrobial solutions had evolved to a standard of care [(SOC)] over the last 20 years, a placebo-controlled study would be unethical." (Medical Review at p. 3).

Part 2. Understanding of the Disease Process

"There is adequate evidence available in the literature to establish that wounds, including burn wounds, may be expected to progress satisfactorily if the microbial load present is reduced to less than 10^5 organisms per gram of tissue . . . it may be said that if a topical antimicrobial is successful in maintaining low bacterial levels on a newly placed skin graft until the graft is adequately vascularized, the antimicrobial has contributed to take of the graft." (Medical Review at p. 42).

Part 3. Understanding of the Relationship Between the Treatment Failures in Those with <20% TBSA Burned and Treatment Failures in Those with >20% TBSA Burned

"The applicants have been reluctant to use a vehicle control on the grounds that failure to treat a burn patient with a [TBSA] burn of larger than 10-20% would be unethical." (Medical Review at p. 4). This was supported by the deliberations of the Advisory Committee. Therefore, while the single pivotal trial enrolled all patients with burns, regardless of how extensively the body was burned, there was "no protocol-specified assignment of patients to treatment with [either Sulfamylon or standard of care (SOC)]. This was a medical decision, made by the attending physician . . . The reviewers separated the results into patient groups by TBSA burned. All patients who had burns covering more than 40% TBSA were treated with [Sulfamylon] . . . It is impossible to assess the effect of [Sulfamylon] in this group. In the 20-40% TBSA burn group, there were a few patients who received [SOC] but . . . the contribution of [Sulfamylon] is difficult to quantify. However, there [were] sufficient [SOC] patients in the 0-20% TBSA burn group to permit comparison of the two treatment regimens." (Medical Review at p. 48).

As for understanding the relationship between drugs in the same pharmacologic class as Sulfamylon, "Sulfamylon for 5% Topical Solution" is the drug product that was the subject of this NDA. However, "Sulfamylon cream is currently approved for use in the treatment of second and third degree burns and the proposed indication for the Sulfamylon 5% solution is related. (Medical Review at p. 49). "Because of the pain caused by the cream, burn physicians began to make a 5% solution using mafenide acetate power in the mid-1970s . . . and the 5% solution has become the standard of use in some burn units for maintaining skin grafts in the period between graft placement and graft take." (Medical Review at p. 4).

Part 4. Clinical Evidence on Those with <20% and Those with >20% TBSA Burned

The single pivotal efficacy study was an unblinded, retrospective, non-randomized, parallel group study with an active control of Standard of Care (SOC) and was conducted at a single site and with a single investigator: Dr. Glenn Warden at Shriner's Burn Institute in Cincinnati, Ohio.

In this study, among the 229 procedures in persons with less than 20% TBSA burned, there were 19 (19%) who were "treatment failures" in those treated with Sulfamylon compared to 33 (26%) who failed on SOC. However, those treated with Sulfamylon had more serious burns, that is, third-degree burns (6.5% vs. 3.3% SOC), a higher percentage of the body surface area burned (10.6% vs. 7.0% SOC), and fewer with only less serious burns, that is, those with second-degree burns only (4.4% vs. 17.3% SOC).

"In her review, the reviewing Statistician, Dr. Yulan Li, reached the following conclusion: Based on the Cincinnati study, the applicant has demonstrated that the use of [Sulfamylon] is associated with the decreasing of treatment failure in the subgroup of patients with 0-20% TBSA burns after separately adjusting for etiology and degree of burn. However, it is unknown whether . . . treatment failure reflects the benefit of [Sulfamylon] due to non-random treatment assignment and investigator knowledge of treatment at the time treatment failure was assessed." (Medical Review at p. 6).

While there appears to be no disagreement in any FDA review as to the intermediate clinical endpoint of effect in those with less than 20% TBSA burned as "reasonably likely to predict benefit" in those with burns over more than 20% TBSA; there were concerns expressed, especially by the Statistician, as to the strength of the efficacy evidence for the findings in those with less than 20% TBSA burned.[13]

---

[13] While scored as a "1," the strength of clinical evidence on the surrogate here with Sulfamylon could reasonably be scored as either "1" or "zero," and the same may be said of the strength of clinical evidence for the surrogate in Synercid, Precedent #13.

17. PROAMATINE (midodrine hydrochloride)

FDA approved Proamatine for treating "symptomatic orthostatic hypotension" on September 6, 1996 on the basis of "increases in 1-minute standing systolic blood pressure, a surrogate marker likely to correspond to a clinical benefit" (as stated in FDA-approved labeling).[14]

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Condition and Lack of Alternative Therapy

Although the review documents for Proamatine are not publicly available on FDA's website, the Agency's Subpart H approval of Proamatine has to mean that FDA assessed the condition as rather serious and lacking available therapy.

b. Rarity of the Condition

Proamatine was designated as an orphan drug on June 21, 1985.

c. External Expertise

There is no evidence from documents currently available, including approved labeling and trade press, whether FDA sought the advice of an Advisory Committee. Therefore, we scored this as a "zero."

Part 2. Understanding of the Disease

For FDA to have approved Proamatine on the basis of a change in 1-minute systolic blood pressure suggests that FDA must have considered that there was a sound understanding of the pathophysiology of the disease.

Part 3. Understanding of the Relationship Between Change in 1-Minute Systolic Blood Pressure and the Ability to Perform Life Activities

Since there are no other drugs in any class approved for this condition, FDA could not have relied upon their effects on this disease. However, many drugs are approved on changes in blood pressure as a validated surrogate based upon both robust epidemiology and multiple interventions affecting serious cardiovascular outcomes such as MACE, and FDA may have relied upon this strong association for support of the power of a change in 1-minute systolic blood pressure in this disease to predict clinical benefit in this disease.

---

[14] All of the formation in this analysis is drawn from the FDA approved labeling, as no Medical or Statistical Reviews from FDA were publicly available.

Part 4. Clinical Evidence on 1-Minute Systolic Blood Pressure and Clinical Outcome

"Midodrine has been studied in 3 principal controlled trials, one of 3-weeks duration and two of 1-to-2 days duration. All studies were randomized, double-blind and parallel-design trials in patients with orthostatic hypertension of any etiology and supine-to-standing fall of systolic blood pressure of at least 15 mmHg accompanied by at least moderate dizziness/lightheadedness . . . In the 3-week study in 170 patients . . . , the midodrine-treated patients . . . had significantly higher (by about 20 mmHg) 1-minute standing systolic pressure 1 hour after dosing . . . for all 3 weeks. After week 1, midodrine-treated patients had small improvements in dizziness/lightheadedness/unsteadiness scores and global evaluations, but these effects were made difficult to interpret by a high early drop-out rate (about 25% vs. 5% on placebo). Supine and sitting blood pressure rose 16/8 and 20/10 mmHg, respectively, on average. In the 2-day study, after open-label midodrine, known midodrine responders received midodrine 10 mg or placebo at 0, 3, and 6 hours. One-minute standing systolic blood pressures were increased 1 hour after each dose by about 15 mmHg and 3 hours after each dose by about 12 mmHg; 3-minute standing pressures were increased also at 1, but not 3, hours after dosing. There were increases in standing time seen intermittently 1 hour after dosing, but not at 3 hours. In the 1-day, dose-response trial, single does of 0, 2.5, 10, and 20 mg of midodrine were given to 25 patients. The 10- and 20-mg doses produced an increase in standing 1-minute systolic pressure of about 30 mmHg at 1 hour; the increase was sustained in part for 2 hours after 10 mg and 4 hours after 20 mg. Supine systolic pressure was =200 mmHg in 22% of patients on 10 mg and 45% of patients on 20 mg; elevated pressures often lasted 6 hours or more." (Midodrine Package Insert).

18. BIAXIN (clarithromycin)

FDA approved Biaxin on December 23, 1993 for treating disseminated mycobacterial infections due to mycobacterium avium complex (MAC) on the basis of a showing of Biaxin's effect on the surrogate of decreases in MAC bacteria.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity of the Disease, Rarity, and Lack of Alternative Therapy

The pivotal studies were conducted in persons with CDC-defined AIDS and CD_4 counts <100 cells/μL, and median survival time in the one trial that was randomized and blinded was 249 days and 215 days for the two dose groups reported in the approved labeling.[15]

While Biaxin was not designated as an orphan drug for this use, this condition was not prevalent and the absence of orphan drug status is likely due to the FDA approval of Biaxin for many other prevalent diseases (such that orphan drug exclusivity would have had substantially diminished, if any, value).

b. External Expertise

On May 11, 1993, the Antiviral Drugs Advisory Committee provided insight on the approvability of Biaxin for treatment of MAC.[16]

Part 2. Understanding of the Disease

The pathophysiology of MAC in immune-compromised AIDS patients was likely understood relatively well for the extent of time that the condition had been known.

Part 3. Understanding of the Relationship Between Reducing MAC Bacteremia and Clinical Outcomes

The general axiomatic principles of infectious disease likely guided and illuminated FDA's interpretation of the prognostic value of reducing MAC bacteremia on achieving negative cultures and clinical benefit. Other antibiotic regimens had shown some value as well.

Part 4. Clinical Evidence on Reducing MAC Bacteremia and Clinical Outcomes Including Mortality

---

15 There were no FDA medical or statistical reviews publicly available and nearly all information is from the FDA approved labeling.

16 Based on public documents currently available, it is unclear what the outcome of this Advisory Committee was.

Of the 3 studies conducted from May 1991 to March 1992, Study 500 was the only one to be blinded and randomized (dose comparison trial of 3 different doses of Biaxin). Study 500 showed a reduction in MAC bacteremia with the lowest dose having the smallest decrease in colony-forming units (CFUs). There was seemingly no survival benefit, as the FDA-approved labeling reported that: "The median survival times for these [Biaxin] dosages were similar to recent historical controls with MAC when treated with combination therapies." However, there was some evidence of improvement in other signs and symptoms of MAC infection including night sweats, fever, and weight loss.

19. BETASERON (interferon beta-1b)

FDA approved Betaseron as the first therapy to treat multiple sclerosis (MS) on July 23, 1993 on the basis of a showing on both rate and extent of exacerbations and on improvement in MRI-measured lesion area.

Part 1. Regulatory Factors Weighing into FDA Determination

a. Severity, Rarity, and Lack of Available Therapy

MS is a serious disease for which, prior to Betaseron, there was no FDA approved treatment. Betaseron was designated as an orphan drug on November 17, 1988.

b. External Expertise

The FDA Peripheral and Central Nervous System Advisory Committee on March 19, 1993 voted 7-2 to recommend approval of Betaseron.

Part 2. Understanding of the Disease

The pathophysiology of multiple sclerosis was known to a fair degree at the time of the conduct of the pivotal trial which permitted the Sponsor, in collaboration with the lead FDA CBER official, Dr. Woodcock, and her office, to have general agreement on co-primary endpoints of clinical utility related to exacerbations, as well as somatic measures of the putatively key causal biologic marker, MRI lesion volume.

Part 3. Understanding of the Relationship Between MRI Lesion Volume and Multiple Sclerosis

"It was also clear that the Committee as a whole placed great weight on the MRI findings in their deliberations. Specifically, although the clinical benefit, as measured by the proportion of exacerbation-free patients and exacerbation frequency, was considered real and of value clinically, the Committee considered the size of the treatment effect to be relatively small.

However, it was obvious that great emphasis was placed on the MRI findings. Specifically, the Committee appeared convinced by the firm's presentation that the drug had an important effect on the underlying pathology as measured by total lesion area as seen on MRI. The statistically significant decrease in the total lesion area in the high dose group as compared to placebo patients over the course of the study that the sponsor claimed was demonstrated was interpreted by the Committee, in my view, as powerful support for the conclusion that the drug was having an important effect on the underlying disease process. While the Committee stopped short of declaring that the data **proved** the drug had an effect on the progression of the disease, I believe it is fair to characterize their view with a quote, made at the meeting, by Dr. McFarland, who said at one point, that, while the sponsor had not proved that the drug had an effect on the course of the

disease, 'I would be amazed if it didn't change the course of disease.' A number of Committee members explicitly referred to Dr. McFarland's comments in this regard when explaining their votes." (Memo of Dr. Katz, May 28, 1993, pp. 356-357) (emphasis in original).

"That is, it appears clear that the Committee felt that the MRI results not only were consistent with the clinical benefit observed (i.e., the changes seen corresponded to the exacerbation rate data at a given point in time), but that they could be relied upon to accurately 'predict' patients' future courses. In other words, the MRI data were considered, for all intents and purposes, as a surrogate marker for disease." (Memo of Dr. Katz at p. 359)

"If the lesions detected on MRI are taken to be a better index of the 'activity' of the pathologic process than are clinical manifestations of MS, (a not unreasonable possibility given the knowledge that lesions detected on MRI may be unaccompanied by clinical signs/symptoms when they occur in so-called 'silent' regions of the CNS) and if the rate of clinical progression of MS (in the sense of increasing physical disability) is a positive function of the activity of that pathologic process, it follows logically that any drug suppressing this 'activity' 'must'[17] have some beneficial effect on the progression of MS (as manifest by increasing physical disability). Although the clinical evidence collected[18] in Study TB01-35(6/8)86 does not provide convincing affirmative support for

---

17 "Must" appears in quotations as a reminder of prior occasions in the history of therapeutics where perfectly logical extrapolations based on beliefs about the pathophysiology of a disease and the postulated mechanism of a drug's action have led experts to reach totally incorrect conclusions about the promise of a particular drug (e.g., CAST: the suppression of ventricular ectopy 'must' save lives.) [Footnote is part of quotation.]

18 In their report of the study, the sponsor asserts that the correlation between EDSS disability scores and MRI lesion areas detected at both baseline (r=0.169) and at the end of year two (r=0.2) establishes that MRI 'burden' predicts disability (EDSS score). Although these statements are correct in a statistical sense, the correlation does not tell us what we really seek to learn: whether a treatment reducing the extent of MRI area increase over time will reduce the extent of clinical worsening, as judged by EDSS, over the same interval or in a future one. [Footnote is part of quotation.]

this hypothesis, that does not necessarily undercut its appeal or its psychological impact on those asked to render an opinion about the 'therapeutic potential' of Betaseron.

During the PCNS meeting, the sponsor's representatives, several members of the Committee and, in particular, Dr. Henry McFarland, who was attending the meeting as the Agency's expert consultant on neuro-imaging and MS, espoused the hypothesis just described. Although virtually all proponents of this hypothesis acknowledged that the link between MRI lesion frequency/intensity/area and subsequent outcome (progression in level of physical disability) in MS was not proven, almost all affirmed that they would be very surprised if the link was not eventually demonstrated. Thus, for many experts, the number and area of lesions detected on MRI are tantamount to a 'surrogate' endpoint that predicts disease progression in MS." (Memo of Dr. Leber, May 28, 1993, pp. 340-341) (emphasis in original).

"In the Betaseron data there is a second kind of replication, the MRI results, which are more or less persuasive, depending on one's beliefs. At a minimum, as Dr. Leber says, these data are an independent measurement that supports the clinical finding, a kind of 'within-study' replication. At best, they are evidence of an effect far more important than the modest effect on exacerbations. We certainly are not qualified to choose between these interpretations, but our advisors seem to believe the latter, even though all would agree that, strictly, the correlation of improved clinical outcome and improved MRI has not been made.

It would be possible, we believe, to grant approval under the accelerated approval regulations, which allow this procedure where a surrogate or clinical, but non-ultimate endpoint is the basis for approval. (Memo of Dr. Temple to Dr. Woodcock, June 3, 1993, pp. 329-330) (emphasis in original)

Part 4. Clinical Evidence on MRI Lesion Volume and on Reduction in Exacerbations of MS

"The trial was designated as a randomized, double-blind, and placebo-controlled study to evaluate the safety and efficacy of Betaseron in the treatment of patients with relapsing-remitting MS . . . The protocols proposed that the primary efficacy evaluations will be based on reduction in frequency of exacerbations per subject and proportion of exacerbation-free subjects." (Statistical Review, March 1, 1993, p.1).

"The proportions of exacerbation-free subjects in the three arms of the study are given in Table 1. If we consider all reported exacerbations, 18 of the 112 placebo patients (16.1%) and 36 of the 115 45 mIU Betaseron patients (31.3%) were exacerbation-free. This difference was significant at $p=0.008$." (Statistical Review at p. 3).

"The second primary endpoint, prospectively specified in the protocol, was the frequency of exacerbation per subject . . . If we consider the outcomes in all six

categories of exacerbations (i.e., 0, 1, 2, 3, 4, and 5+) then the probability of better response on Betaseron therapy is 63%. It is significantly different (p=0.0004) from 50%." (Statistical Review at pp. 5-6).

As for the MRI lesion volume results, depending upon the analysis used by the FDA reviewer, Dr. Jay Siegel, the p-value for the comparison between Betaseron and placebo arms ranges from a p-value of 0.03 to a p-value of 0.001. (Memo of Dr. Siegel, June 24, 1993, p. 1)

ORIGINAL ARTICLE

A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis

Talmadge E. King, Jr., M.D., Williamson Z. Bradford, M.D., Ph.D., Socorro Castro-Bernardini, M.D., Elizabeth A. Fagan, M.D., Ian Glaspole, M.B., B.S., Ph.D., Marilyn K. Glassberg, M.D., Eduard Gorina, M.D., Peter M. Hopkins, M.D., David Kardatzke, Ph.D., Lisa Lancaster, M.D., David J. Lederer, M.D., Steven D. Nathan, M.D., Carlos A. Pereira, M.D., Steven A. Sahn, M.D., Robert Sussman, M.D., Jeffrey J. Swigris, D.O., and Paul W. Noble, M.D., for the ASCEND Study Group*

ABSTRACT

BACKGROUND

In two of three phase 3 trials, pirfenidone, an oral antifibrotic therapy, reduced disease progression, as measured by the decline in forced vital capacity (FVC) or vital capacity, in patients with idiopathic pulmonary fibrosis; in the third trial, this end point was not achieved. We sought to confirm the beneficial effect of pirfenidone on disease progression in such patients.

METHODS

In this phase 3 study, we randomly assigned 555 patients with idiopathic pulmonary fibrosis to receive either oral pirfenidone (2403 mg per day) or placebo for 52 weeks. The primary end point was the change in FVC or death at week 52. Secondary end points were the 6-minute walk distance, progression-free survival, dyspnea, and death from any cause or from idiopathic pulmonary fibrosis.

RESULTS

In the pirfenidone group, as compared with the placebo group, there was a relative reduction of 47.9% in the proportion of patients who had an absolute decline of 10 percentage points or more in the percentage of the predicted FVC or who died; there was also a relative increase of 132.5% in the proportion of patients with no decline in FVC ($P<0.001$). Pirfenidone reduced the decline in the 6-minute walk distance ($P=0.04$) and improved progression-free survival ($P<0.001$). There was no significant between-group difference in dyspnea scores ($P=0.16$) or in rates of death from any cause ($P=0.10$) or from idiopathic pulmonary fibrosis ($P=0.23$). However, in a prespecified pooled analysis incorporating results from two previous phase 3 trials, the between-group difference favoring pirfenidone was significant for death from any cause ($P=0.01$) and from idiopathic pulmonary fibrosis ($P=0.006$). Gastrointestinal and skin-related adverse events were more common in the pirfenidone group than in the placebo group but rarely led to treatment discontinuation.

CONCLUSIONS

Pirfenidone, as compared with placebo, reduced disease progression, as reflected by lung function, exercise tolerance, and progression-free survival, in patients with idiopathic pulmonary fibrosis. Treatment was associated with an acceptable side-effect profile and fewer deaths. (Funded by InterMune; ASCEND ClinicalTrials.gov number, NCT01366209.)

From the University of California, San Francisco, San Francisco (T.E.K.), InterMune, Brisbane (W.Z.B., E.A.F., E.G., D.K.), and Cedars–Sinai Medical Center, Los Angeles (P.W.N.) — all in California; Neumocare, Clínica San Borja, Lima, Peru (S.C.-B.); Alfred Hospital, Melbourne, VIC (I.G.), and Prince Charles Hospital, Brisbane, QLD (P.M.H.) — both in Australia; University of Miami Miller School of Medicine, Miami (M.K.G.); Vanderbilt University Medical Center, Nashville (L.L.); Columbia University Medical Center, New York (D.J.L.); Inova Fairfax Hospital, Falls Church, VA (S.D.N.); Paulista School of Medicine, Federal University of São Paulo, São Paulo (C.A.P.); Medical University of South Carolina, Charleston (S.A.S.); Atlantic Health System–Overlook Medical Center, Summit, NJ (R.S.); and National Jewish Health, Denver (J.J.S.). Address reprint requests to Dr. King at the Department of Medicine, University of California, San Francisco, 505 Parnassus Ave., Box 0120, San Francisco, CA 94143-0120, or at tking@medicine.ucsf.edu.

*A complete list of members of the Assessment of Pirfenidone to Confirm Efficacy and Safety in Idiopathic Pulmonary Fibrosis (ASCEND) study group is provided in the Supplementary Appendix, available at NEJM.org.

This article was published on May 18, 2014, at NEJM.org.

DOI: 10.1056/NEJMoa1402582

Idiopathic pulmonary fibrosis is a chronic, progressive, and fatal lung disease that is characterized by irreversible loss of lung function.[1] Although periods of transient clinical stability may be observed, continued progression of the disease is inevitable.[2] The prognosis is poor, with a 5-year survival rate that is similar to the rates for several cancers.[3-6]

Pirfenidone is an oral antifibrotic therapy that has been evaluated for the treatment of idiopathic pulmonary fibrosis in three phase 3, randomized, controlled trials. One of these trials was conducted in Japan and involved 275 patients. It was followed by two multinational studies, Clinical Studies Assessing Pirfenidone in Idiopathic Pulmonary Fibrosis: Research of Efficacy and Safety Outcomes (CAPACITY studies 004 and 006), that were conducted in the United States, Europe, and Australia and involved 779 patients.[7,8] In the Japanese trial, pirfenidone reduced the decline in vital capacity at week 52 and improved progression-free survival. In the multinational trials, the primary end point of change from baseline to week 72 in the percentage of the predicted forced vital capacity (FVC) was met in study 004 but not in study 006, prompting U.S. regulatory authorities to request an additional trial to support the approval of pirfenidone.

In the Assessment of Pirfenidone to Confirm Efficacy and Safety in Idiopathic Pulmonary Fibrosis (ASCEND) study, a randomized, double-blind, placebo-controlled trial, we aimed to confirm the effect of pirfenidone on disease progression in patients with idiopathic pulmonary fibrosis. Our design modifications with respect to the CAPACITY trial included the implementation of centralized procedures for diagnosis, spirometry, and adjudication of deaths; a minor modification of eligibility criteria to allow enrollment of patients with an increased risk of disease progression; and a standard 1-year study period.

METHODS

STUDY SITES AND PATIENTS

The study was conducted at 127 sites in 9 countries (11 sites in Australia, 6 in Brazil, 2 in Croatia, 5 in Israel, 5 in Mexico, 2 in New Zealand, 8 in Peru, 1 in Singapore, and 87 in the United States). Eligible patients were between the ages of 40 and 80 years and had received a centrally confirmed diagnosis of idiopathic pulmonary fibrosis. The diagnostic criteria, based on published consensus guidelines, were findings on high-resolution computed tomography (HRCT) that indicated either definite or possible usual interstitial pneumonia; the latter was confirmed on surgical lung biopsy.[1] Other criteria for enrollment included a range of 50 to 90% of the predicted FVC, a range of 30 to 90% of the predicted carbon monoxide diffusing capacity, a ratio of the forced expiratory volume in 1 second (FEV_1) to the FVC of 0.80 or more, and a 6-minute walk distance of 150 m or more. (A comprehensive list of inclusion and exclusion criteria is provided in the Supplementary Appendix, available with the full text of this article at NEJM.org.) All patients provided written informed consent.

STUDY DESIGN AND ASSESSMENTS

Eligible patients were randomly assigned to receive oral pirfenidone (at a dose of 2403 mg per day) or placebo for 52 weeks. The study drug was administered with food in three equally divided doses, and the dose was gradually increased to the full dose over a 2-week period. Randomization codes were generated by computer with the use of a permuted-block design, and the study drug was assigned by means of an interactive voice-response system. Concomitant treatment with any investigational therapy was prohibited. Selected concomitant medications that are used for the treatment of idiopathic pulmonary fibrosis were permitted if they were used for another indication, provided that there was no clinically acceptable alternative.

Physical examination and clinical laboratory assessments were performed at baseline and at weeks 2, 4, 8, 13, 26, 39, and 52. Pulmonary function, exercise tolerance, and dyspnea were assessed at baseline and at weeks 13, 26, 39, and 52. Central reviewers at Biomedical Systems, who were unaware of study-group assignments, evaluated all FVC results for adequacy and repeatability, according to the criteria of the American Thoracic Society.[9] A data and safety monitoring committee reviewed safety and efficacy data throughout the trial.

The study protocol was approved by the institutional review board or ethics committee at each participating center. The protocol and statistical analysis plan are available at NEJM.org.

STUDY OVERSIGHT

The study sponsor (InterMune) and the steering committee cochairs were primarily responsible

for the design of the study. All authors participated in the conduct of the study, analysis of data, and reporting of the results. A writing committee comprising the first and last authors, the study medical monitor, and a medical writer (who was paid by the study sponsor) prepared the first draft of the manuscript. All authors vouch for the accuracy and completeness of the report and for the fidelity of the report to the protocol; all the authors critically reviewed the manuscript and approved the final draft. All the authors had full access to data, and no limits were placed on the content of the report.

STATISTICAL ANALYSIS

The primary efficacy end point was the change from baseline to week 52 in the percentage of the predicted FVC in the intention-to-treat population. The test statistic for the primary efficacy analysis was a ranked analysis of covariance (ANCOVA), with the average standardized rank change in the percentage of the predicted FVC as the outcome variable and the standardized rank baseline value as a covariate. The primary efficacy analysis was tested with the use of a final two-tailed P value of 0.0498, which was adjusted for two planned interim analyses. The magnitude of the treatment effect was estimated by comparing the distribution of patients in the pirfenidone group with those in the placebo group across two thresholds of change at week 52: an absolute decline of 10 percentage points in the percentage of the predicted FVC or death, or no decline in the percentage of the predicted FVC. Supportive analyses to assess the robustness of the effect on FVC were also conducted.

Two key secondary end points and three additional secondary end points were prespecified. The key secondary end points, which were analyzed with the use of the Hochberg procedure for multiple comparisons,[10] were the change from baseline to week 52 in the 6-minute walk distance and progression-free survival. Progression-free survival was defined as the time to the first occurrence of any one of the following: a confirmed decrease of 10 percentage points or more in the percentage of the predicted FVC, a confirmed decrease of 50 m or more in the 6-minute walk distance, or death. Additional secondary end points included change in dyspnea, which was measured with the use of the University of California San Diego Shortness of Breath Questionnaire (UCSD SOBQ), with scores ranging from 0 to 120 and higher scores indicating worse dyspnea (minimally important difference, 5 to 11 points) (Fig. S4 in the Supplementary Appendix); the rate of death from any cause; and the rate of death from idiopathic pulmonary fibrosis during the period from baseline to 28 days after the last dose of the study drug.

In accordance with the prespecified statistical analysis plan, rates of death from any cause and death from idiopathic pulmonary fibrosis were analyzed in the ASCEND study population and in the pooled population from the ASCEND trial and the two CAPACITY trials; the latter analysis was performed for the purpose of increasing the statistical power and deriving a more stable estimate of the treatment effect. For the pooled analysis, CAPACITY results were censored at day 365 so that the follow-up time would be the same for all three studies. The primary cause of death and its relation to idiopathic pulmonary fibrosis were assessed in a blinded fashion by an independent mortality assessment committee in the ASCEND trial and by the site investigators in the CAPACITY trials (Tables S1 and S2 in the Supplementary Appendix).

All efficacy analyses were conducted in the intention-to-treat population with the use of SAS software, version 9.2 (SAS Institute). For the ranked ANCOVA analyses, missing values owing to death were assigned the worst ranks, with early deaths ranked worse than later deaths. In analyses of mean change, missing values owing to death were assigned the worst possible outcome (e.g., FVC=0). Missing values for reasons other than death were imputed as the average value for the three patients with the smallest sum of squared differences at each visit. For time-to-event analyses, pirfenidone was compared with placebo with the use of a log-rank test; hazard ratios were based on the Cox proportional-hazards model.

Adverse events were coded according to preferred terms in the *Medical Dictionary for Regulatory Activities*, version 11.0. Safety outcomes are reported as events that occurred in the period from baseline to 28 days after the last dose of the study drug.

RESULTS

STUDY PATIENTS

From July 2011 through January 2013, a total of 555 patients were enrolled; 278 were assigned to receive pirfenidone, and 277 were assigned to re-

ceive placebo. Demographic and baseline characteristics are summarized in Table 1. There were no significant imbalances in clinically relevant baseline characteristics between the two study groups. The majority of patients were male (79.9% and 76.9% in the pirfenidone and placebo groups, respectively), white (91.7% and 90.6%, respectively), and 65 years of age or older (73.7% and 68.2%, respectively). The mean (±SD) baseline FVC was 67.8±11.2% of the predicted value in the pirfenidone group and 68.6±10.9% of the predicted value in the placebo group.

A total of 522 patients (94.1%) completed the study: 261 patients (93.9%) in the pirfenidone group and 261 patients (94.2%) in the placebo group (Fig. 1). Study treatment was discontinued prematurely in 55 patients (19.8%) in the pirfenidone group and in 39 patients (14.1%) in the placebo group. Adherence to the study treatment was high; 237 patients (85.3%) and 256 (92.4%) patients in the pirfenidone and placebo groups, respectively, received at least 80% of the prescribed doses of the assigned study drug.

Table 1. Characteristics of the Patients at Baseline.*

Characteristic	Pirfenidone (N=278)	Placebo (N=277)
Age — yr	68.4±6.7	67.8±7.3
Male sex — no. (%)	222 (79.9)	213 (76.9)
U.S. enrollment — no. (%)	187 (67.3)	184 (66.4)
Former smoker — no. (%)	184 (66.2)	169 (61.0)
Lung physiological features		
FVC — % of predicted value	67.8±11.2	68.6±10.9
FEV_1:FVC	0.84±0.03	0.84±0.04
Carbon monoxide diffusing capacity — % of predicted value	43.7±10.5	44.2±12.5
Dyspnea score†	34.0±21.9	36.6±21.7
Distance on 6-min walk test — m	415.0±98.5	420.7±98.1
Use of supplemental oxygen — no. (%)	78 (28.1)	76 (27.4)
Time since diagnosis — yr	1.7±1.1	1.7±1.1
Diagnostic finding on high-resolution computed tomography — no. (%)		
Definite pattern of usual interstitial pneumonia	266 (95.7)	262 (94.6)
Possible pattern of usual interstitial pneumonia‡	12 (4.3)	15 (5.4)
Surgical lung biopsy — no. (%)	86 (30.9)	79 (28.5)

* Plus–minus values are means ±SD. There were no significant differences between the two groups in any of the baseline characteristics shown. FEV_1 denotes forced expiratory volume in one second, and FVC forced vital capacity.

† Dyspnea was evaluated with the use of the University of California, San Diego, Shortness of Breath Questionnaire, scores on which range from 0 to 120, with higher scores indicating worse dyspnea; the minimally important difference is 5 to 11 points.

‡ The diagnosis was subsequently confirmed on surgical lung biopsy indicating a histologic pattern of usual interstitial pneumonia.

PRIMARY EFFICACY ANALYSIS

In the ranked ANCOVA analysis, treatment with pirfenidone resulted in a significant between-group difference in the primary end point, the change from baseline to week 52 in the percentage of the predicted FVC (P<0.001). At week 52, the proportion of patients who had a decline of 10 percentage points or more in the percentage of the predicted FVC or who had died was reduced by 47.9% in the pirfenidone group as compared with the placebo group (46 patients [16.5%] vs. 88 patients [31.8%]) (Fig. 2A), and the proportion of patients with no decline in the percentage of the predicted FVC was increased by 132.5% in the pirfenidone group (63 patients [22.7%] vs. 27 patients [9.7%]) (Fig. S1 in the Supplementary Appendix).

The treatment effect was evident by week 13 and increased throughout the duration of the trial. Supportive analyses of the primary end point yielded similar results. The mean decline from baseline in FVC was 235 ml in the pirfenidone group and 428 ml in the placebo group (absolute difference, 193 ml; relative difference, 45.1%; P<0.001) (Fig. 2B). The linear slope of decline in FVC at week 52 was −122 ml in the pirfenidone group and −262 ml in the placebo group (absolute difference, 140 ml; relative difference, 53.5%; P<0.001) (Fig. S2 in the Supplementary Appendix).

PRESPECIFIED SECONDARY EFFICACY ANALYSES

Pirfenidone resulted in a significant between-group difference in the change from baseline to week 52 in the 6-minute walk distance (P=0.04). At week 52, a decrease of 50 m or more in the 6-minute walk distance or death occurred in 72 patients (25.9%) in the pirfenidone group and in 99 patients (35.7%) in the placebo group, for a relative reduction of 27.5% in the pirfenidone group (Fig. 2C, and Fig. S3 in the Supplementary Appendix).

Pirfenidone, as compared with placebo, reduced the relative risk of death or disease progression by 43% (hazard ratio in the pirfenidone group, 0.57; 95% confidence interval [CI], 0.43 to 0.77; P<0.001) (Fig. 2D). For each component

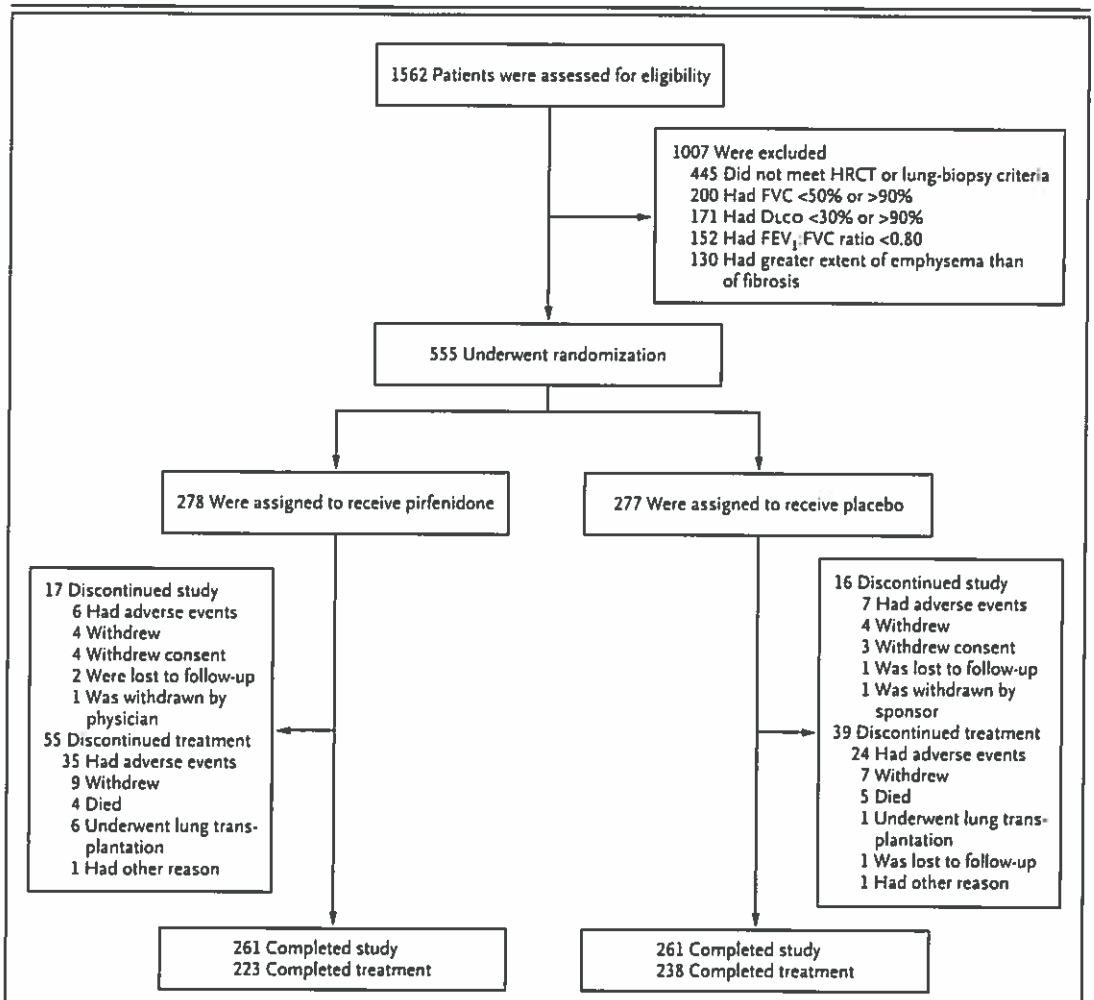


Figure 1. Enrollment and Outcomes.

The listed reasons for exclusion from the study were reported in at least 10% of the patients who underwent screening. Patients could have more than one reason for exclusion. The numbers of patients who withdrew from the study do not include patients who died or underwent lung transplantation. Patients who discontinued the study treatment were included in the analysis of data for patients who completed the study. DLCO denotes carbon monoxide diffusing capacity, FEV_1 forced expiratory volume in one second, FVC forced vital capacity, and HRCT high-resolution computed tomography.

of the composite end point, fewer patients in the pirfenidone group than in the placebo group had a qualifying event, including death (3.6% vs. 5.1%), a confirmed absolute decrease of 10 percentage points or more in the percentage of the predicted FVC (6.5% vs. 17.7%), and a confirmed decrease of 50 m or more in the 6-minute walk distance (16.5% vs. 19.5%).

Analysis of UCSD SOBQ scores showed no significant between-group difference in dyspnea at week 52. The end point of an increase of 20 points or more (indicating worsening) on the dyspnea score or death occurred in 81 patients (29.1%) in the pirfenidone group and in 100 patients (36.1%) in the placebo group (absolute difference, 7.0 percentage points; relative reduction, 19.3%; P=0.16) (Fig. S4 in the Supplementary Appendix).

MORTALITY OUTCOMES

Analysis of all-cause mortality showed fewer deaths in the pirfenidone group than in the placebo group, although the difference was not significant. Eleven patients (4.0%) in the pirfenidone group died during the study, as compared

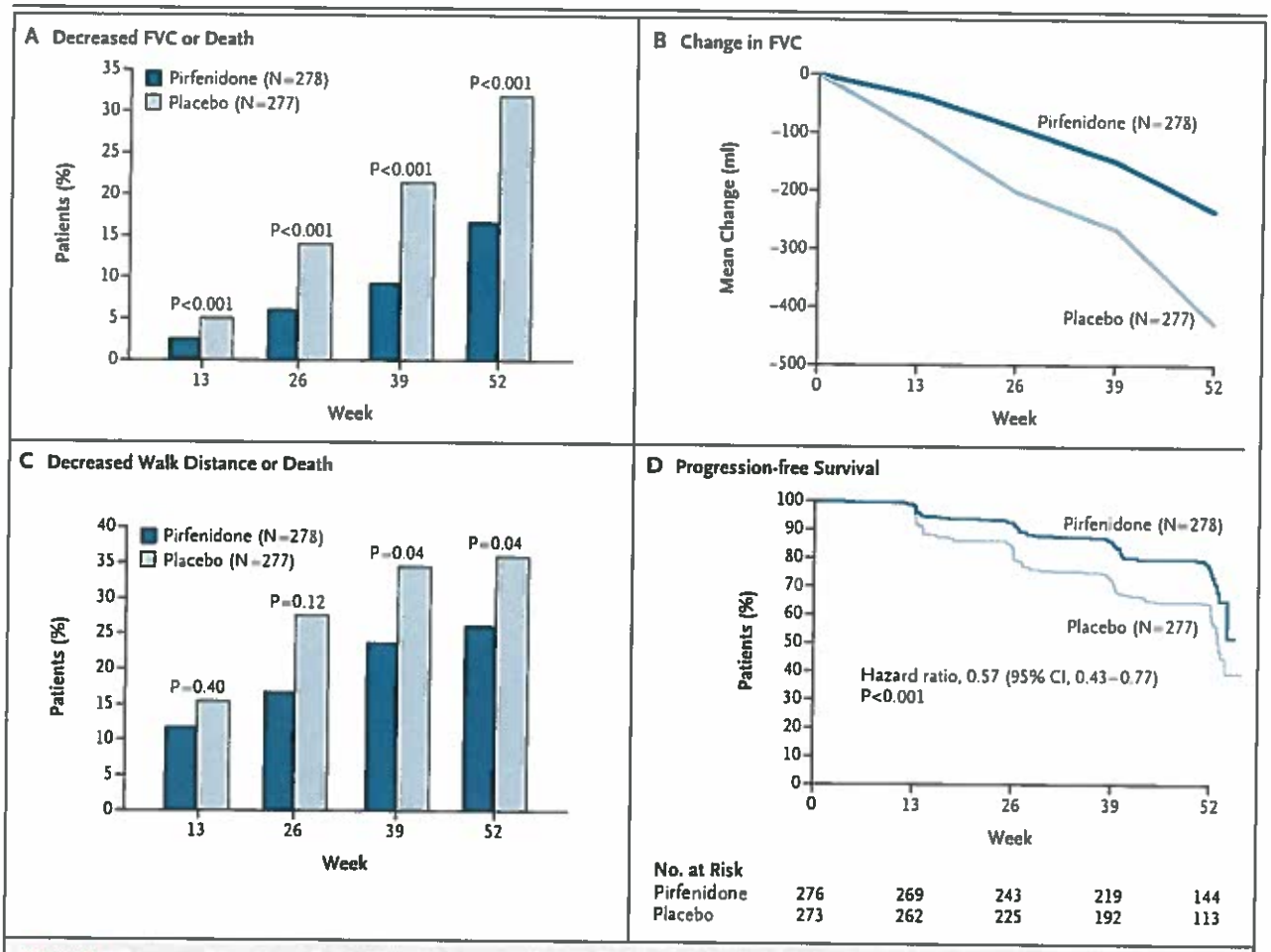


Figure 2. Primary and Key Secondary Efficacy Outcomes during the 52-Week Study Period.

Panel A shows the proportion of patients who had a decreased percentage of the predicted FVC (defined as a decline of at least 10 percentage points from baseline) or who died. Panel B shows the mean change from baseline in FVC. Panel C shows the proportion of patients who had a decreased walk distance (defined as a decline of 50 m or more in the distance walked in 6 minutes) or who died. P values shown in Panels A, B, and C were calculated with the use of ranked analysis of covariance. Panel D shows the Kaplan–Meier distribution for the probability of progression-free survival. The P value was calculated with the use of the log-rank test.

with 20 patients (7.2%) in the placebo group (hazard ratio, 0.55; 95% CI, 0.26 to 1.15; P=0.10). Deaths from idiopathic pulmonary fibrosis occurred in 3 patients (1.1%) and 7 patients (2.5%) in the pirfenidone and placebo groups, respectively (hazard ratio, 0.44; 95% CI, 0.11 to 1.72; P=0.23).

In the prespecified analysis of all-cause mortality in the pooled population of 1247 patients (555 from the ASCEND study and 692 from the CAPACITY studies), pirfenidone reduced the risk of death at 1 year by 48%, as compared with placebo (hazard ratio, 0.52; 95% CI, 0.31 to 0.87; P=0.01) (Table 2). In addition, in the pooled population, the risk of death from idiopathic pulmonary fibrosis at 1 year was reduced by 68% in the pirfenidone group, as compared with the placebo group (hazard ratio, 0.32; 95% CI, 0.14 to 0.76; P=0.006). (Additional mortality results are provided in Tables S3, S4, and S5 in the Supplementary Appendix.)

ADVERSE EVENTS

Adverse events that occurred during the study period are summarized in Table 3. Gastrointestinal and skin-related events were more common in the pirfenidone group than in the placebo group; these events were generally mild to moderate in severity, reversible, and without clinically signifi-

Table 2. Mortality in the ASCEND and CAPACITY Trials.*

Variable	Pirfenidone	Placebo	Hazard Ratio (95% CI)†	P Value‡
ASCEND trial				
No. of patients	278	277		
Death — no. (%)				
From any cause	11 (4.0)	20 (7.2)	0.55 (0.26–1.15)	0.10
Related to idiopathic pulmonary fibrosis§	3 (1.1)	7 (2.5)	0.44 (0.11–1.72)	0.23
Pooled data from ASCEND and CAPACITY trials				
No. of patients	623	624		
Death — no. (%)				
From any cause	22 (3.5)	42 (6.7)	0.52 (0.31–0.87)	0.01
Related to idiopathic pulmonary fibrosis§	7 (1.1)	22 (3.5)	0.32 (0.14–0.76)	0.006

* Data from the two CAPACITY studies[8] were censored at 1 year to standardize the follow-up for the three studies.
† Hazard ratios are for the pirfenidone group, as compared with the placebo group, and were calculated with the use of the Cox proportional-hazards model.
‡ P values were calculated with the use of the log-rank test.
§ Death related to idiopathic pulmonary fibrosis was defined as death that occurred during the period from randomization to 28 days after the last dose of the study drug. This category was evaluated in a blinded fashion by an independent mortality-assessment committee in the ASCEND trial and by clinical investigators in the CAPACITY trials.

cant sequelae. Grade 3 gastrointestinal adverse events were reported in 15 patients (5.4%) in the pirfenidone group and 4 patients (1.4%) in the placebo group. Grade 3 skin-related adverse events were reported in 5 patients (1.8%) in the pirfenidone group and 1 patient (0.4%) in the placebo group. No patients in either group had a grade 4 gastrointestinal or skin-related event. Cough, worsening of idiopathic pulmonary fibrosis, and dyspnea occurred more frequently in the placebo group. There were fewer deaths in the pirfenidone group than in the placebo group (8 [2.9%] vs. 15 [5.4%] between baseline and 28 days after the last dose of a study drug).

The relative difference between treatment groups in the overall incidence of serious adverse events is less clear. If worsening of idiopathic pulmonary fibrosis is counted as an adverse event (as specified in the protocol), there were 55 patients (19.8%) in the pirfenidone group and 69 patients (24.9%) in the placebo group who had a serious adverse event. The most common serious adverse event was worsening of idiopathic pulmonary fibrosis, which was reported in 7 patients (2.5%) in the pirfenidone group and in 27 patients (9.7%) in the placebo group. However, since worsening of idiopathic pulmonary fibrosis is a study outcome, it is reasonable to exclude patients with worsening fibrosis in the analysis of serious adverse events. With such patients excluded, serious adverse events occurred in 52 patients (18.7%) in the pirfenidone group and 56 patients (20.2%) in the placebo group.

Elevations in the level of alanine or aspartate aminotransferase (values that were three or more times the upper limit of the normal range) occurred in eight patients (2.9%) in the pirfenidone group and two patients (0.7%) in the placebo group, including one patient in the pirfenidone group who had a concurrent elevation in the total bilirubin level that was more than two times the upper limit of the normal range. All aminotransferase elevations were reversible and without clinically significant consequences.

Adverse events led to discontinuation of study treatment in 40 patients (14.4%) in the pirfenidone group and 30 patients (10.8%) in the placebo group. The most common adverse event resulting in treatment discontinuation was a worsening of idiopathic pulmonary fibrosis in 3 patients (1.1%) in the pirfenidone group and in 15 patients (5.4%) in the placebo group. The only other adverse events leading to treatment discontinuation in at least 1% of the patients in the pirfenidone group were elevated hepatic en-

Table 3. Adverse Events.*

Adverse Event	Pirfenidone (N=278)	Placebo (N=277)
	no. of patients (%)	
Cough	70 (25.2)	82 (29.6)
Nausea	100 (36.0)	37 (13.4)
Headache	72 (25.9)	64 (23.1)
Diarrhea	62 (22.3)	60 (21.7)
Upper respiratory tract infection	61 (21.9)	56 (20.2)
Fatigue	58 (20.9)	48 (17.3)
Rash	78 (28.1)	24 (8.7)
Dyspnea	41 (14.7)	49 (17.7)
Dizziness	49 (17.6)	36 (13.0)
Idiopathic pulmonary fibrosis†	26 (9.4)	50 (18.1)
Bronchitis	39 (14.0)	36 (13.0)
Constipation	32 (11.5)	38 (13.7)
Back pain	30 (10.8)	37 (13.4)
Dyspepsia	49 (17.6)	17 (6.1)
Nasopharyngitis	33 (11.9)	30 (10.8)
Anorexia	44 (15.8)	18 (6.5)
Vomiting	36 (12.9)	24 (8.7)
Decrease in weight	35 (12.6)	22 (7.9)
Gastroesophageal reflux	33 (11.9)	18 (6.5)
Insomnia	31 (11.2)	18 (6.5)

* Listed are all adverse events that were reported in at least 10% of patients in either study group. Preferred terms in the *Medical Dictionary for Regulatory Activities*, version 11.0, were used for documentation of adverse events.

† Since idiopathic pulmonary fibrosis was a criterion for enrollment, this category of adverse events refers to worsening of disease.

zyme levels, pneumonia, rash, and decreased weight in 3 patients (1.1%) each.

DISCUSSION

In this phase 3 study comparing pirfenidone with placebo in patients with idiopathic pulmonary fibrosis, treatment with pirfenidone for 52 weeks significantly reduced disease progression, as measured by changes in FVC, the 6-minute walk distance, and progression-free survival. The treatment effect on FVC emerged early and increased during the course of the trial, resulting in an approximate halving in the rate of decline at 1 year. The highly significant finding with respect to the primary end point was supported by the favorable effect on rates of death from any cause and from idiopathic pulmonary fibrosis.

Treatment with pirfenidone was generally safe and had an acceptable side-effect profile, findings that are consistent with those in previous studies.[7,8,11,12] Gastrointestinal and skin-related adverse events were more common in the pirfenidone group than in the placebo group; these events were generally mild to moderate in severity and led to treatment discontinuation in 2.2% and 2.9% of patients, respectively, in the pirfenidone group and 1.1% and 0.4% of those, respectively, in the placebo group. There were fewer serious adverse events and deaths in the pirfenidone group than in the placebo group. Clinically significant elevations in aminotransferase levels occurred more frequently in the pirfenidone group; however, these elevations occurred in less than 3% of patients, were reversible, and did not have clinically significant consequences.

The results of this study confirm and extend the findings of the two CAPACITY trials (studies 004 and 006),[8] each of which was smaller and of longer duration than the ASCEND trial. An important observation in the CAPACITY 006 trial was the attenuated rate of decline in FVC in the placebo group, as compared with that in the CAPACITY 004 study and another multinational trial.[13] In our study, we modified certain aspects of the CAPACITY study design, including increasing the sample size and requiring central confirmation of the diagnosis. We also modified selected eligibility criteria in order to enroll patients at higher risk for disease progression. Thus, we excluded patients with major airflow limitation (ratio of FEV_1 to FVC, <0.80) and reduced the minimum baseline carbon monoxide diffusing capacity from 35% to 30% of the predicted value. The latter modification meant that 22% of the patients in our study had a baseline carbon monoxide diffusing capacity of less than 35% of the predicted value. Despite these and other minor design modifications, the baseline characteristics of the patients in the ASCEND study were strikingly similar to those in the CAPACITY studies, and the magnitude of the treatment effect at 1 year was generally consistent in these three studies and the Japanese phase 3 trial.

Our findings are strengthened by the high rates of study completion and treatment adherence and the consistent magnitude of treatment

effect across the primary and secondary end points. In addition, both FVC and 6-minute walk distance are reliable, valid, and responsive measures of disease status and independent predictors of the risk of death among patients with idiopathic pulmonary fibrosis.[14-24] Finally, the thresholds of change that were selected for the categorical analyses of FVC and 6-minute walk distance are well above the estimated minimal clinically important difference for each measure.[14,15,24-26]

The mortality analyses were prespecified to be conducted in both the ASCEND population and in the pooled population from the ASCEND and CAPACITY trials because of the low rate of death among patients who are typically enrolled in clinical trials of idiopathic pulmonary fibrosis and because of the need for a larger sample to obtain precise estimates of the treatment effect.[27] The magnitude of the treatment effect on mortality was large and internally consistent across analyses and populations — an important clinical finding. In addition, the effect size was generally consistent with the observed effect on measures of disease progression, providing further support for the use of these measures in subsequent clinical trials.

The results of our study should be interpreted in the context of certain limitations. First, we enrolled patients with mild-to-moderate physiological impairment; the degree to which our findings can be generalized to a population of patients with advanced disease is therefore uncertain. Second, we required central confirmation of the diagnosis of idiopathic pulmonary fibrosis on the basis of criteria from recent diagnostic guidelines.[1] However, the general similarity in outcomes at 1 year between our study and the CAPACITY studies — in which the site investigator determined the diagnosis — militates against any limitation that this requirement might impose on the generalizability of our results.

In conclusion, we found that pirfenidone as compared with placebo reduced disease progression in patients with idiopathic pulmonary fibrosis. Treatment was generally safe, had an acceptable side-effect profile, and was associated with fewer deaths.

Supported by InterMune.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Kenneth Glasscock for medical writing and editorial assistance and Gary Koch, Ph.D., for statistical consultation; Robert S. Fishman, M.D., and Jonathan A. Leff, M.D., who contributed to data analysis, as well as the members of the CAPACITY program steering committee (Carlo Albera, M.D., Ulrich Costabel, M.D., Roland M. du Bois, M.D., and Dominique Valeyre, M.D.); and the patients, their family members, and participating staff at all the study centers.

REFERENCES

1. Raghu G, Collard HR, Egan JJ, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 2011;183:788-824.
2. Martinez FJ, Safrin S, Weycker D, et al. The clinical course of patients with idiopathic pulmonary fibrosis. Ann Intern Med 2005;142:963-7.
3. Nicholson AG, Colby TV, du Bois RM, Hansell DM, Wells AU. The prognostic significance of the histologic pattern of interstitial pneumonia in patients presenting with the clinical entity of cryptogenic fibrosing alveolitis. Am J Respir Crit Care Med 2000;162:2213-7.
4. Daniil ZD, Gilchrist FC, Nicholson AG, et al. A histologic pattern of nonspecific interstitial pneumonia is associated with a better prognosis than usual interstitial pneumonia in patients with cryptogenic fibrosing alveolitis. Am J Respir Crit Care Med 1999;160:899-905.
5. Olson AL, Swigris JJ, Lezotte DC, Norris JM, Wilson CG, Brown KK. Mortality from pulmonary fibrosis increased in the United States from 1992 to 2003. Am J Respir Crit Care Med 2007;176:277-84.
6. Surveillance, Epidemiology, and End Results Program (SEER). SEER cancer statistics review, 2003–2009 (http://seer.cancer.gov).
7. Taniguchi H, Ebina M, Kondoh Y, et al. Pirfenidone in idiopathic pulmonary fibrosis. Eur Respir J 2010;35:821-9.
8. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. Lancet 2011;377:1760-9.
9. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. Eur Respir J 2005;26:319-38.
10. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988;75:800-2.
11. Azuma A, Nukiwa T, Tsuboi E, et al. Double-blind, placebo-controlled trial of pirfenidone in patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2005;171:1040-7.
12. Valeyre D, Albera C, Bradford WZ, et al. Comprehensive assessment of the long-term safety of pirfenidone in patients with idiopathic pulmonary fibrosis. Respirology (in press).
13. King TE Jr, Albera C, Bradford WZ, et al. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. Lancet 2009;374:222-8.
14. du Bois RM, Weycker D, Albera C, et al. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. Am J Respir Crit Care Med 2011;184:1382-9.
15. du Bois RM, Weycker D, Albera C, et al. Six-minute-walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. Am J Respir Crit Care Med 2011;183:1231-7.
16. Collard HR, King TE Jr, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2003;168:538-42.
17. Flaherty KR, Mumford JA, Murray S,

et al. Prognostic implications of physiologic and radiographic changes in idiopathic interstitial pneumonia. Am J Respir Crit Care Med 2003;168:543-8.
18. Jegal Y, Kim DS, Shim TS, et al. Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. Am J Respir Crit Care Med 2005;171:639-44.
19. King TE Jr, Safrin S, Starko KM, et al. Analyses of efficacy end points in a controlled trial of interferon-gamma1b for idiopathic pulmonary fibrosis. Chest 2005;127:171-7.
20. Zappala CJ, Latsi PI, Nicholson AG, et al. Marginal decline in forced vital capacity is associated with a poor outcome in idiopathic pulmonary fibrosis. Eur Respir J 2009;35:830-6.
21. Richeldi L, Ryerson CJ, Lee JS, et al. Relative versus absolute change in forced vital capacity in idiopathic pulmonary fibrosis. Thorax 2012;67:407-11.
22. du Bois RM, Weycker D, Albera C, et al. Ascertainment of individual risk of mortality for patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2011;184:459-66.
23. du Bois RM, Albera C, Bradford WZ, et al. 6-Minute walk test distance is an independent predictor of mortality in patients with idiopathic pulmonary fibrosis. Eur Respir J 2014;43:1421-9.
24. Caminati A, Bianchi A, Cassandro R, Mirenda MR, Harari S. Walking distance on 6-MWT is a prognostic factor in idiopathic pulmonary fibrosis. Respir Med 2009;103:117-23.
25. Holland AE, Hill CJ, Conron M, Munro P, McDonald CF. Small changes in six-minute walk distance are important in diffuse parenchymal lung disease. Respir Med 2009;103:1430-5.
26. Swigris JJ, Wamboldt FS, Behr J, et al. The 6 minute walk in idiopathic pulmonary fibrosis: longitudinal changes and minimum important difference. Thorax 2010;65:173-7.
27. King TE Jr, Albera C, Bradford WZ, et al. All-cause mortality rate in patients with idiopathic pulmonary fibrosis. Implications for the design and execution of clinical trials. Am J Respir Crit Care Med 2014;189:825-31.

MEMORANDUM

DEPARTMENT OF HEALTH AND HUMAN SERVICES
PUBLIC HEALTH SERVICE
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH

DATE: December 27, 2004

FROM: Director, Office of Drug Evaluation I, HFD-101

SUBJECT: Iloprost, NDA 21-779, Cotherix, Inc.

TO: File and HFD-110

I. Introduction and Effectiveness Analysis

This NDA has been reviewed critically by Drs. Gordon and Karkowsky, both of whom recommend approval, albeit with some reservation from Dr. Karkowsky about the target population [primary vs. secondary (mostly post-pulmonary embolism) pulmonary hypertension]. The principal effectiveness issues are:

1. Reliance on a single study
2. Whether to indicate the drug for pulmonary hypertension (PHT) generally or only for primary pulmonary hypertension

1. Reliance on a single study

Under FDAMA, we are permitted to rely on a single study plus "confirmatory evidence" (never really defined). In general, based on the FDA guidance (Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products), the single study should be convincing statistically and it helps if there is internal consistency (e.g., in the present case, similar effects on NYHA classification and walking distance). The role of the effectiveness of related therapy is considered only briefly in the guidance but has been explicitly used in the approval (based on studies with non-extreme statistical tests) of two angiotensin II blockers to delay renal functional deterioration in type II diabetics (each study supported the other) and less explicitly (but nonetheless pretty clearly) in approving ACEI's for the treatment of CHF, relying on single studies with p-values between 0.05 and 0.01 with the backgroup of multiple drugs in the class showing favorable effect.

In the present case, the strongest external support comes from the closely related prostacyclin analogues epoprostanol (Flolan), delivered through a central venous line, and treprostanil (Remodulin), given through an indwelling subcutaneous catheter, and approved without a clear effect on exercise but an effect on a combined breathlessness - exercise endpoint. There is also a second small iloprost study that generally favors iloprost over placebo but had numerous problems (single blind, changing definitions, etc.) and was not considered seriously.

Results of Study ME 97218

ME 97218 was a 12-week RCT with 201 randomized patients (101 iloprost, 100 placebo), stratified by primary vs. secondary PHT and by NYHA class (III vs. IV). The endpoint was a novel one (most Rx for PHT was approved based on the 6 minute walk as the primary endpoint), the "responder rate," with responders defined as patients with:

>10% increase in walking distance

≥1 grade increase in NYHA class

no deterioration (death, worse hypotension, worse R-sided CHF, ≥30% worse walking distance, cardiogenic hepatic or renal deterioration, new need for I.V. meds, CI<1.31 L/min/m^2; CVP>22, SVO_2<45% on nasal O_2

Walking distance was a secondary endpoint.

Results:

	Iloprost	Control	
Responder	17/101 (21%)	5/102 (5%)	p=0.007
Walking Distance at 12 week	+22 meter	-3 meter	p=0.032
peak			
trough	+15 meter	0	

Considering the components of the primary endpoint (from Dr. Karkowsy).

	Iloprost	Control
Walk increase >10%	38/101 (38%)	26/102 (25%)
Change in NYHA >1	25/101 (25%)	13/102 (13%)
Deterioration	6/101 (5%)	15/102 (15%)
(No deterioration)	95/101 (95%)	87/102 (85%)

This shows considerable consistency across these (probably highly correlated) components of the endpoint.

2. Primary vs. Secondary PHT

Drugs for PHT approved to date have studied largely primary PHT (including, however, PHT following scleroderma, etc.), not PHT following pulmonary emboli. Although the present study of iloprost clearly had as a primary endpoint the entire population of both primary and secondary PHT (and showed a highly significant result for the whole group), results were not the same in the two etiologic strata.

	Primary		Secondary	
	Iloprost	Placebo	Iloprost	Placebo
Overall Resp	11/53 (21%)	3/55 (5%)	6/48 (13%)	2/47 (4%)
Components				
Walk >10%	26/53 (49%)	17/55 (31%)	12/48 (25%)	9/47 (19%)
NYHA >1	13/53 (25%)	4/55 (7%)	12/48 (25%)	9/47 (19%)
Overall WD	42	-2	2	8

One certainly cannot conclude that iloprost does not work in secondary PHT but there is a question as to whether there are adequate data to conclude that it does.

The sponsor has urged that the indication be for PHT 1) because the combined group was the primary endpoint (and neither subgroup was proven to show an effect alone, 2) trends did favor iloprost, 3) when walking distance included zero values for patients who died, results are stronger [Note, the figure in labeling showing a 36 m difference at 12 weeks is based on this analysis; I do not agree with this post-facto analysis] in both subgroups.

I conclude that:

1. The data based on a single principal study are convincing and provide substantial evidence of effectiveness of iloprost.

2. The claim should be limited to the primary PHT; the results in secondary PHT can be noted in clinical trials and the primary endpoint identified but that section should note that there are too few data to conclude that effectiveness has been demonstrated.

3. The labeling figure of walking distance should be replaced by one that does not attribute zero walking to people who died.

II. Safety

Safety has been well discussed by Drs. Gordon and Karkowsky. No deaths appeared drug-related. Iloprost clearly can cause hypotension and even syncope, predictable from its vasodilatory properties. In the 129 inhalation patients there were 10 reports of syncope and 10 of hypotension, vs. 6 each in the placebo group, with 3 withdrawals (one each) for syncope, hypotension, and vasodilation. The 6 syncope events reported as serious (MOR, pages 31-2) are unimpressive, often occurring well after the inhalation, and attributable to (1) second degree AV block (treated with a pacemaker), (1) "vasovagal" episodes (with an event more than 6 hours after medication, (1) at the end of an ETT, (1) associated with stair-climbing, (1) associated with probable Iloprost-induced R heart decompensation, and (1) probably resulting from hyperventilation (confirmed in provocative test).

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Robert Temple
12/28/04 03:54:12 PM
MEDICAL OFFICER

MEMORANDUM

DEPARTMENT OF HEALTH & HUMAN SERVICES
Public Health Service
Food and Drug Administration
Center for Drug Evaluation and Research

DATE: December 23, 2004

FROM: Abraham Karkowsky, M.D., Ph.D.; Group Leader Division of Cardio-Renal Drug, Products HFD-110

TO: Robert Temple, M.D.; Office Director ODE-I

SUBJECT: Approvability of Iloprost (Ventavis®) inhalation for the treatment of pulmonary hypertension (NDA 21-779, Cotherix Inc).

This memo is in support of the approvable recommendation of Iloprost, administered by inhalation, for use to provide symptomatic benefit, limited to patients with primary pulmonary hypertension. The nature of this benefit is a composite of a 10% increase in walk distance, an improvement in NYHA class and without any of the pre-specified criteria defining a worsening of status. It is likely that patients will have benefit for at least 30 minutes after an inhalation treatment, as reflected in an increase in walk-distance during the clinical trial. Benefit at the interdosing interval appears less than at the 30–minute post inhalation time point.

Source Materials:

The following reviews and sources of information were consulted for the purposes of constructing this memo.

- Medical officer review by Dr. Maryann Gordon, M.D., dated 12 November 2004.
- Pharmacology review by Dr. James Willard Ph.D., dated 14 December 2004.
- CMC reviews by Dr. M.D. Cooper Ph.D. and Dr. W.C. Timmer Ph.D. dated 3 and 17 December 2004.
- Clinical pharmacology and biopharmaceutic review by Dr. Robert O. Kumi, Ph.D., dated 1 December 2004.
- Statistical review of efficacy by Dr. Valeria Freidlin, Ph.D., dated 28 October 2004.
- DMETS review from D. P. Toyer, PharmD., dated 15 December 2004.
- Clinical inspection summary by Mary I. Mease dated 13 December 2004.
- DSRCS review of patient labeling by Jeanne Best, M.S.N, R.N., P.N.P. dated 16 December 2004.
- Microbiology reviews by James L. McVey dated 9, 15 and 21 December 2004.

- Proprietary name review; DMETS consult by Scott Dallas R.PH. dated 28 October 2004.
- Statistical review of carcinogenicity by Jasmine Choi, M.S., dated 15 November 2004.
- DDMAC draft label review by Catherine Gray Pharm.D., and Lance McLeroy Pharm.D., dated 18 November 2004.
- The sponsor's submission of 30 June 2004.

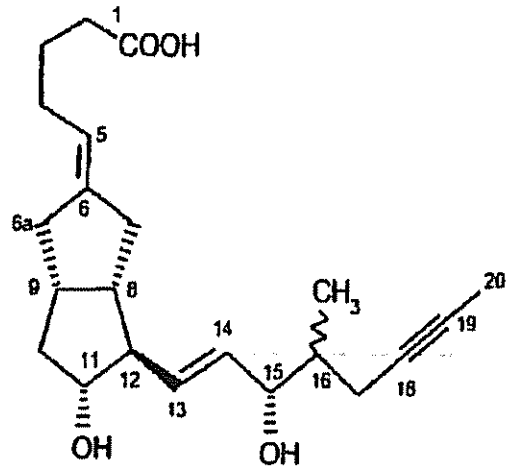
Appears This Way
On Original

Chemistry:

Iloprost is a diastereoisomeric mixture whose IUPAC and USAN name as well as its structure are shown below.

Chemical (IUPAC) Name: 5-[(E)-(1S, 5S, 6R, 7R)-7-Hydroxy-6-[(E)-(3S, 4RS)-3-hydroxy-4-methyl-1-octen-6-ynyl]-bicyclo[3.3.0]oct-3-ylidene]-pentanoic acid

Chemical (USAN) Name: (*E*)-(3a*S*,4*R*,5*R*,6a*S*)-Hexahydro-5-hydroxy-4-[(*E*)-(3*S*,4*RS*)-3-hydroxy-4-methyl-1-octen-6-ynyl]-$\Delta^{2(3H)A}$-pentalenevaleric acid



Note: The numbering system in the above structure does not correspond to the IUPAC or USAN chemical names, but to the prostacyclin numbering system.

Chemical Formula:	$C_{22}H_{32}O_4$
Molc. Wt.:	360.49
CAS No.:	78919-13-8

Iloprost contains 6 optically active sites; five of which are fixed. The sixth asymmetric site, the 4-position methyl group (labeled as carbon # 15 in the above diagram, as represented by a wavy line), as a consequence of the synthetic process, is not fixed relative to the other optically-active centers. The to-be marketed product consequently, contains two-distinct chemical entities in a ratio of 53:47 of 4R:4S Iloprost. These entities have different pharmacologic properties and are chemically distinct and theoretically readily separable.

As Drs. Cooper and Timmer note, current FDA policy[1] is to treat such diastereoisomers as separate chemical entities unless they spontaneously interconvert (not apparently the case here).

The sponsor argued that large-scale separation of the enantiomers would be a difficult and complicated process. In addition, neither the kinetic or dynamic properties of the two diastereomers suggest an apparent hazard in their concurrent use. After intravenous administration to dogs or rats, both diastereomers demonstrate rapid and similar clearance. Both isomers have activity in rats and human platelets ex-vivo assays in preventing either ADP or collagen induced platelet aggregation. Both diastereomers also had vasodilatory activity of rabbit mesenteric artery and both decreased blood pressure in anesthetized rat. The potency of the 4S isomer is greater than the 4R isomer. The toxicology of the either the 4R or 4S isomers does not indicate that the less active isomer is substantially more toxic and consequently, its presence does not reflect a substantial hazard to the patient. The effect of the two diastereoisomers on the main pharmacologic properties of Iloprost, that is, vasodilation and platelet inhibition are proportional, with the 4R isomer approximately 1/10 to 1/20 as active in both activities. Lastly, the large safety database of approximately 3,000 treated patients exposed to the diastereomeric mixture already exists and the adverse events profile does not strongly suggest the existence of adverse events other than events reflecting an extension of the known pharmacologic activity of either isomer.

In sum what is currently known about the animal toxicology, animal kinetics and available safety data in humans coupled with the accepted assertion, that separation of the diastereoisomers mixture is a complex process, the requirement to isolate and restudy a single isomer would only delay the approval of this drug. Despite the stated agency's policy, approval of this diasterisomeric mixture appears warranted.

Adequate responses to all chemistry deficiencies have been received. The submission is approvable from the chemistry vantage point.

Delivery Systems:

During the pivotal phase 3 trial, Iloprost was administered with the HaloLite nebulizer, manufactured by Profile Therapeutics. This model of nebulizer is not available in the United States. The ProDose inhaler is a modification of the HaloLite model and currently available. The differences between the two-nebulizers are a more durable compressor unit and an improved patient nebulizer interface with a programmable dose-control disc for the ProDose nebulizer. The ProDose received 510(k) approval predicated on the HaloLite device.

The microchip disc that was initially calibrated to yield the equivalent total dose of 2.5 mcg per treatment (250 mcL) as in the clinical study, in actuality administered 3.8 mcg ± 14.8%. As a consequence, the ProDose disc was changed to one calibrated to deliver a lower volume (150 mcL). The 5.0 mcg dose was reprogrammed with a chip to

[1] "FDA Policy Statement for the Development of New Stereoisomeric Drugs"

deliver 450 mcL (as opposed to 500 mcL). Upon reprogramming the disc, the amount of delivery approximated that as delivered by the HaloLite device.

The rate of delivery of drug does not appear to be uniform during a single inhalation treatment. During the initial anticipated 2.5 mcg dose the real delivered dose was closer to 2.8 mcg. During the second 2.5 mcg portion of the dose only 2.3 mcg was delivered. The reason for the non-uniformity is unclear. Since there is excess of drug in each ampoule (each ampoule contains 20 mcg of Iloprost), enough for more than a single inhalation, the patient need to be advised that the attempt to obtain more than one treatment per ampoule would not reliably deliver to an effective dose.

Environmental Assessment Exclusion.

The chemistry reviewer accepted the sponsor's assertion that the concentration of the active drug and its not metabolites will not exceed 1 ppb in the environment, with no otherwise extraordinary circumstances suggesting that either the drug or metabolites would adversely alter the environment. A waiver of the environmental assessment is appropriate.

Inspections.

EES report was received on 13 December 2004. The results of the inspections were acceptable.

Microbiology.

"Sterile" Iloprost was recommended as approvable per microbiology.

Pharmacology.

Iloprost is an analog of PGI2, and belongs to the same class as two currently approved treatments for pulmonary hypertension; treprostinil and Flolan. Inhibition of binding to receptors other than PGI2 was not observed except to histamine and purinergic P2 receptors at a concentration of 10 uM. The inhibition curve, at lower concentrations of Iloprost was not studied. The10 uM concentrations far exceed the concentration anticipated at the site of action[2]. The inhibition of binding at 10 uM was limited to the 4S-isomer. The 4R-isomer and the mixture of isomers (Iloprost) did not apparently inhibit binding to the histamine and purinergic receptors.

Activity for both vasodilation and platelet inhibition, at least as measured in animal models, resides with both the 4S and 4R isomers. In general the 4R diastereoisomer was generally 10 to 20-fold less potent than the 4S isomer.

It is unclear if the dilation by Iloprost is homogenous across all vascular beds. In conscious rats that were infused concentrations of 0.1 mcg/kg/min, blood flow was

[2] Consider a dose of 5 mcg administered over 5 minutes, or a rate of approximately 1mcg/min. Assuming a cardiac output of 2 L/min, the concentration in the resulting blood flow from the lung to the myocardium and into the arterial system would be 0.5 mcg/L. The MW of Iloprost is 360, the concentrations would correspond to approximately 1.4 nM, or approximately 4 orders of magnitude lower than the single concentration that inhibited either purinergic or histaminic receptors..

increased significantly as measured by labeled micro spheres to spleen, stomach, and small intestine; it was less significantly increased in skin, colon and lung. There was apparently no increase in blood flow to heart and muscle.

The results of the genotoxicity and mutagenicity studies, with one exception, were benign. In the chromosomal aberration assay with Chinese Hamster Lung Cells, there was a mildly positive response. Dr. Willard notes in his review, that these cells have both surface receptors for prostanoids as well the mechanism for translocation of these receptors into the cell nucleus. The relevance of the finding of chromosomal abnormalities to cells without the capabilities to bind and translocate prostanoids into the nucleus is therefore, unclear. No carcinogenicity effect was observed in mice and rats orally-treated with Iloprost.

With respect to reproductive toxicology, in rabbits or Sprague-Dawley rats, Iloprost at oral doses less than those found to be maternal toxic, showed no significant effects on either dam or fetus. At maternally toxic doses (by the oral route) the number of non-viable fetuses was increased. In Han-Wistar rats intravenous doses of 1 mg/kg to the dam were embryolethal in approximately 1/3 of the litters. In Han-Wistar rats Iloprost infused at approximately 1/10 the embryolethal dose, skeletal and digit abnormalities were observed.

The above observations were included in the labeling as edited by the pharmacologist.

Biopharmaceutics.

ADME

Upon inhalation, Iloprost, a mixture of both diastereoisomers rapidly appears in plasma. None of the assays employed in human studies separated the two (4S from 4R isomers). Peak levels of Iloprost based on 12 PAH patients was 157 ± 64 pg/mL. The half-life of the combined diastereoisomeric mixture in humans is 7.9 ± 3.2 minutes. In dogs and rats there does not appear to be interconversion of the two diastereoisomers.

In rats after oral administration, metabolism of the Iloprost diastereoisomers is by ß-oxidation. The metabolism is not substantially dependent of CYP-450 enzymes. The major metabolites of Iloprost are tetranor-Iloprost and tetranor derivatives (glucuronides). A mass balance study was performed by the sponsor in humans (n=8) with tritium labeled Iloprost administered either by the intravenous (2 ng/kg/min x 4 hours) or orally at two different doses (0.1 and 0.48 mcg/kg). Blood was collected for through 24 hours. Urine and feces were collected for up to 1 week. Collection of radioactivity in urine was > 95% complete by approximately 14 hours and 2 days in feces. The total dose recovered was approximately 80% of the radioactivity; with 68% collected from urine and 12% from feces.

Special populations.

Hepatic impairment.

There were no studies performed in hepatically impaired patients with inhaled Iloprost. However, after intravenous infusion of Iloprost at a dose of 1 ng/kg/min in a small number of subjects with Child Pugh class A, B and C (1, 5 and 2 subjects, respectively), CP_{ss} was increased by 50% to 120 % in the various classes of liver dysfunction. $T_{1/2}$ however was not convincingly increased.

It is unclear if there is a safety price that is a consequence of the higher peak concentrations. It is unclear if peak serum concentration is correlated with the drug's benefit, given that the concentration at the site of action (the pulmonary vasculature) is unlikely to be reflective of serum concentrations at steady state. The appropriate recommendation for this population is unclear. The uncertainty of the appropriate recommendation for this population should be incorporated into labeling.

Renal Impairment.

There were no studies performed in patients with renal impairment with inhaled Iloprost. However, after an intravenous infusion of Iloprost at a dose of 1 ng/kg/min to subjects with impaired renal function but not on dialysis (n=7) or who routinely require dialysis (n=8). Peak concentration among those who generally require dialysis was approximately three-fold higher than those not requiring dialysis. Clearance was rapid and by two hours post infusion there was little Iloprost measurable in either group.

It is unclear if there is a safety price that is a consequence of the higher peak concentrations. It is unclear if peak serum concentration is correlated with the drug's benefit, given that the concentration at the site of action (the pulmonary vasculature) is unlikely to be reflective of serum concentrations at steady state. The appropriate recommendation for this population is unclear. The uncertainty of the appropriate recommendation for this population should be incorporated into labeling.

Clinical Efficacy.

The current database for the approval of inhalation Iloprost for the treatment of pulmonary hypertension is dependent on a single, placebo-controlled, double-blind study (study #ME97218). A second smaller study (study #ME98998) was flawed in that the dose used differed from study ME97218. In addition, results were reclassified and modified after the blind was broken. The smaller study adds little to the decision for approval.

Safety of Iloprost is supported by the two extension studies of the placebo-controlled studies of Iloprost by inhalation. In addition, there is some experience, although of limited utility, with Iloprost administered either as an intravenous or oral formulation.

With respect to efficacy, study ME97218 was a placebo-controlled study in patients with pulmonary hypertension. Patients were stratified at baseline based on the origin of pulmonary hypertension (primary versus secondary) and NYHA classification at baseline (NYHA III versus IV). Only a single dosing regimen was used. Patients received as the first inhalation 2.5 mcg over 4.5 minutes. If the initial dose was tolerated,

subsequent doses were 5.0 mcg over 9 minutes. The initial regimen was for six inhalations, no more frequent than every two hours. The number of inhalations could be increased to a total of nine daily.

The primary metric of the study was a combined endpoint comparing the number of responders among those treated with Iloprost to placebo-treated subjects. A responder was one who had a greater than 10% increase in baseline walk distance and had at least one grade improvement in their NYHA classification at the 12-week visit and who did not deteriorate during the course of the study. Deterioration was defined as either death or by the occurrence of two or more of the following criteria:

- Refractory systolic arterial hypotension of > 85 mm Hg.
- Worsening right heart failure (cardiac edema, ascites or pleural effusion), despite adequate background therapy
- Rapidly progressive cardiogenic hepatic failure.
- Rapidly progressive cardiogenic renal failure.
- A decrease in walking distance by ≥ 30% from baseline.
- New and new need for intravenous medication (e.g., catecholamines or diuretics).
- Cardiac index < 1.3 l/min/m^2.
- CVP > 22 mm Hg (via indwelling catheter) despite adequate diuretic therapy.
- SVO2 < 45% despite nasal O_2 therapy (right heart catheterization).

Secondary endpoints were not pre-ordered and included: exercise capacity, NYHA class, dyspnea index, hemodynamic parameters and gas exchange, deterioration of pulmonary hypertension, mortality and quality of life.

Of the 235 patients who were screened, 203 were randomized; 101 to Iloprost inhalation and 102 to placebo. The etiology of the pulmonary hypertension was idiopathic in 108 (108/203= 53%) and secondary forms in the other patients. Among the 95 patients classified as having secondary pulmonary hypertension 57 (57/95=60%) had as their etiology of pulmonary hypertension thromboembolic events. This population is not subsumed in the INDICATION by for either of the prostanoids currently approved to treat pulmonary hypertension. Thirty-nine percent (35/90) had as their etiology some form of collagen vascular disease (systemic sclerosis, CREST, SLE, and overlap syndrome). The etiology of the secondary pulmonary hypertension in the other patients included: post partum, familial, previous appetite suppressant use, and other causes.

With respect to the demographics of those enrolled, the average age was approximately 52 years, approximately 2/3 of those enrolled were female and approximately 3% were other than Caucasian. With respect to concomitant medications, approximately 80% were taking anticoagulants, 66% diuretics, 44% calcium antagonists, 25% ACE antagonists and 44% were on long-term O_2 therapy.

Dropouts were more frequent in the placebo than Iloprost inhalation group. There were four versus 1 death in the placebo and Iloprost groups, respectively.

There were 17/101 (16.9%) responders in the Iloprost group and 5/102 (4.9%) responders in the placebo-treated group. In considering the two stratified subgroups, there were 11/53 responders in the primary pulmonary hypertension group treated with Iloprost and 3/55 among those treated with placebo. There were 6/46 among those with secondary pulmonary hypertension who were responders on Iloprost and 2/47 treated with placebo. The components of the primary end point are included in the table below. In addition, I have included the walk distance both at 30-minutes post inhalation and at pre-inhalation. The pre-inhalation time point was at least 2 hours after the last treatment.

Although there was an overall effect on the composite end point, the small number for each of the stratified groups is not entirely informative. Walking distance at either 30-minutes post inhalation or at least two hours from the previous treatment, limited to those with data available at week 12 (this excludes the deaths and dropouts), however, did not appear to indicate a benefit for those with secondary pulmonary hypertension and who were treated with Iloprost. Since there is inadequate information,

Table 1:Primary endpoint and individual components of the composite as well as walking distance at 30 minutes post-inhalation and at least 2 hours after an inhalation for study ME97218.

	Iloprost	Control
Overall (Responders/ nonresponders) %	17/101 (17%)	5/102 (5%)
PPH (responders/ nonresponders) %	11/53 (21%)	3/55 (5%)
Secondary PH(responders/ nonresponders) %	6/48 (13%)	2/47(4%)
NYHA Class III (responders/ nonresponders) %	10/60 (17%)	4/60(7%)
NYHA IV (responders/nonresponders) %	7/41(17%)	1/42 (2%)
Components of Response criteria		
Walk distance increased by > 10% (responders/nonresponders) %: Overall	38/101(38%)	26/102 (25%)
PPH (responders/nonresponders) %	26/53 (49%)	17/55 (31%)
Secondary PH (responders/ nonresponders) %	12/48 (25%)	9/47 (19%)
NYHA Class III (responders/nonresponders) %	25/60 (42%)	17/60 (28%)
Class IV (responders/nonresponders) %	13/41 (32%)	9/42 21%)
Change in NYHA Class > 1 grade: Overall	25/101 (25%)	13/102 (13%)
PPH	13/53 (25%)	4/55 (7%)
Secondary PH	12/48 (25%)	9/47 (19%)
NYHA Class III	15/60 (25%)	6/60 (10%)
NYHA Class IV	12/48 (25%)	7/42 (17%)
No deterioration by above listed criteria	95/101 (95%)	87/102 (87%)
PPH	49/53 (92%)	46/55 (84%)
Secondary PH	46/48 (96 %)	41/47 (87%)
NYHA Class III	56/60 (93%)	54/60 (90%)
Class IV	39/41 (95%)	33/42 79%)
Overall walking distance at 30 minutes (change in meters ± SD [median]	22.2 ± 71 [20]	-3.2 ± 74 [0]
PPH	42 ± 73 [31]	-2 ±89 [10]
Secondary PH	2 ± 57 [12]	8 ± 47 [0]
NYHA Class III	17 ± 64 [21]	-5 ± 80 [7]
NYHA Class IV	32 ± 75 [20]	17 ± 57 [2]
Overall walking distance at trough (change in meters ± SD), available at week 12	14.6±68[16]	0.2 ± 67[0.5]
PPH	28 ± 76 [32]	1 ± 75 [10]
Secondary PH	-0.2 ± 54 [7]	10 ± 48 [5]
NYHA III	8 ± 66 [13]	-0.6 ± 69 [5] 3
NYHA IV	24 ± 69 [19]	16 ± 54 [12]

from other sources or other similar drugs for this predominantly thromboembolic population, the labeling should limit the approval to those with primary disease.

Hemodynamic measurements were performed for those who were available at the 12-week time point at trough, however, it is not clear if trough represents the measurements after the overnight period when Iloprost was not inhaled or reflects the measurement performed at least, two-hour time point after a last inhalation treatment. Although there was a suggestion of a decrease in PVR, the effect was not statistically significant. After the inhalation of either Iloprost or placebo, there was a substantial further decline in PVR but the two treatments did not substantially differ in this effect (data not shown here).

Table 2:Hemodynamic parameters at trough measurement at week 12, change from baseline

	Iloprost		Control		p- value*
PVR (dyn.Sec.cm^{-5})	N=76	-9.2 + 275	N=77	96.2 + 323	0.07
mPAP mm Hg	N=93	-0.2 +7.3	N=82	-0.1 ± 6.9	0.96
CO l/min	N=91	0.1 ± 0.9	N=80	-0.2 ± 0.8	0.32
SVO_2 (%)	N=72	-1.1 ± 7.6	N=63	-3.2 ± 6.7	0.43

* ANCOVA for treatment term without baseline adjustment (derived from sponsor's Table TT51).

Clinical Safety.

Safety has been reviewed by Dr. Gordon. There are three databases which contribute to the understanding of the safety profile of Iloprost. The most pertinent of these is the modest database among those randomized in the PAH clinical studies. This database consists of 262 patients exposed to either Iloprost inhalation or placebo in controlled studies and 123 patients who subsequently were enrolled in a long-term extension study. Of these patients, 80 were treated for ≥ 1 year and 64 for ≥ 24 months. This database reflects the safety in the target population.

Two additional databases are also pertinent to defining the safety of Iloprost. Iloprost has been previously administered as an intravenous infusion or by the oral route. Systemic exposure during an intravenous infusion assures exposure to both diastereoisomers. With respect to oral Iloprost, there were over 2,000 patients who received Iloprost by this route. Since bioavailability of Iloprost is low (approximately 16%) compared to the intravenous exposure and the precise composition of the diastereoisomers after an oral dose is uncertain. The oral safety database, although useful reflects a greater degree of uncertainty.

Inhalation database.

Deaths.

There were 2/129 deaths in the Iloprost treated patient and 5/133 in the control group. There were an additional 15/123 patients that died during the open-label extension portion of the study. The two deaths in the controlled studies and 13 of the deaths during the long-term extension were related to progression of disease. The two remaining deaths consisted of one patient who died of colon cancer and one who apparently drowned.

Serious adverse events.

The serious adverse events listed during the controlled portion of the study (in more than 1 patient and more frequent in the Iloprost group) are listed below.

Table 3: Serious adverse events in the placebo-controlled studies (ME98998 and ME97218)(> 1% and more frequent in the Iloprost –treated patients):

	Iloprost (N=129)	Placebo/control (N=133)
Overall	29 (23%)	30 (23%)
CHF	6 (5%)	11 (8%)
Syncope	6 (5%)	0
Aggravation reaction	4 (3%)	5 (4%)
Pneumonia	2 (2%)	0
Laboratory test abnormal	2 (2%)	0
Dyspnea	2 (2%)	2 (2%)

During the open-label extension the most common serious adverse events were not dissimilar from those noted during the placebo-controlled exposure of patients. Events occurring in > 2 subjects are listed below.

Table 4:Adverse events in (> 2%) during either placebo-controlled or the open-label long term extension studies

	Any Iloprost in studies with > 1 dose (N=215)+
Body as a whole	36 (17%)
Aggravation reaction	11 95%)
Death	7 (3%)++
Surgery	7 (3%)
No drug reaction	5 (2%)
Asthenia	3 (1%)
Infection	3 (1%)
Cardiovascular System	34 (8%)
Congestive heart failure	17 (8%)
Syncope	9 (4%)
Respiratory system	13 (6%)
Dyspnea	4 (2%)
Pneumonia	3 (1%)
Metabolic and nutritional	9 (4%)
Peripheral edema	4 (2%)
Edema	3 (1%)

+ The database consists of 28 patients treated with Ilopost during the controlled portion of ME 98008, Plus 26 control patients who completed the study plus 4 who terminated early but received long term Iloprost. In addition there were 101patientstreated with Iloprost during the double-blind portion of study ME97218 and 58 patients treated with placebo who received open-label Iloprost.
++ Not all deaths were classified as an adverse event

Labs.

As Dr. Gordon notes, no patient discontinued Iloprost during the double-blind portion of the study as a consequence of a lab abnormality. Three patients on Iloprost had

Elevated LFTs (> 3x of AST, ALT or Alk Phos) during the controlled portion of the study. Two of these patients had baseline elevations, the third had a transient increase which was labeled as a heparin allergy. The value returned to normal levels at the 4-week follow-up.

There were four Iloprost and nine placebo patients with abnormal (> 1.5 x ULN creatinine values) during the double-blind portion of the studies. One patient had a pre-mortal increase in creatinine, reflective of overall poor perfusion. Two subjects had baseline elevations with no significant increases over baseline. One patient a 56 year old female Caucasian had a worsening of creatinine from 150 uM/L at baseline to 195 uM/L at 12weeks. No explanation was supplied for this patient's increase in creatinine.

There were seven patients with abnormalities in either platelets or hemoglobin below the lower limit of normal. All patients had similar abnormalities at baseline (5 with low platelets and 2 with low hemoglobin).

ECG.

A definitive QT study supports the lack of effect of Iloprost inhalation on repolarization. Study C-200-004 was a parallel 4 arm study that enrolled 161 normal volunteers. One group received a single dose of moxifloxacin (400 mg), one group received 2.5 mcg by inhalation every 2 hours: The third group received ascending doses of Iloprost, as tolerated starting with 5 mcg and increasing to 7.5, 10, 12.5, 15, and 20 mcg) every two hours. The fourth group received placebo.

ECGs were performed at baseline and between inhalations (at midpoint and just previous to next inhalation) and after the last dose at 5, 15, 60 minutes, 4, 8, and 15.5 hours after the final inhalation. In the ascending dose group, dose escalation was limited in 13 patients by adverse event. The most frequent of these was chest pain (5 patients), nausea (2 patients), headache (3 patients), tachycardia, dizziness, atrial flutter (1 patient each). Repolarization, as assessed by QT, QTcb, QTcf or QTcI for moxifloxacin was prolonged but not for either the fixed low-dose Iloprost inhalation or the ascending dose Iloprost inhalation group. Since there does not appear to be any long-lasting accumulating metabolites, the results of this study indicate no effect of Iloprost inhalation of repolarization, with a substantial safety margin.

Safety from intravenous studies.

A second database that defines the safety of Iloprost consists of those patients who received intravenous Iloprost. This database consisted of 12 placebo-controlled studies of at least two weeks duration and exposed 764 and 709 patients to Iloprost and placebo, respectively. The population was composed of patients with peripheral atherosclerotic occlusive disease 425/764 (56%); atherosclerotic peripheral vascular disease, with ischemic ulcers 154 /764 (20%); TAO 74/764, (9.7%); diabetic patients with ulcerated/necrotic ulcers critical limb ischemia (56/764) 7.3%; and critical limb ischemia 53/764 (6.9%). The dose for all these studies ranged from 1.5- 4 ng/kg/min for a six hour infusion period 6-7 days per week (32.4 - 86.4 mcg/day assuming a 60-kg person). The duration of treatment ranged from 2-4 weeks.

During the double-blind intravenous studies there were five deaths four in the Iloprost and one in placebo-treated patients. In the subsequent 30-day post-treatment period there were 8 Iloprost and 12 placebo patients who died.

The adverse events leading to withdrawal (in more than two Iloprost patients) during intravenous placebo-controlled studies is shown below. The most common events leading to discontinuation were headache and hypotension.

Table 5: Intravenous Iloprost database: adverse events leading to discontinuation (on > 2 Iloprost patients).

	Iloprost (n=764)	Placebo (n=709)
Nervous system	12 (2%)	2 (< 1%)
Headache	8 (1%)	1 (1%)
Cardiovascular System	16 (2%)	9 (1%)
Hypotension	4 (1%)	3 (< 0.5%)
Digestive system	7 (1%)	3 (< 0.5%)
Vomiting	4 (1%)	1 (< 0.1%)

Laboratory abnormalities for those treated with intravenous Iloprost were not submitted.

Safety from oral Iloprost studies.

The third database consists of 3161 patients in 12-randomized in placebo-controlled studies of > 2 weeks duration. Of these patients, 2033 were treated with oral Iloprost and 1128 with placebo. The studies evaluated the use of Iloprost to treat peripheral vascular disease (n= 1341/2033); Raynaud's syndrome (n= 314/2033); thromboangiitis obliterans (216/2033); rheumatoid arthritis (138/2033); and multiple sclerosis (24/2033). The doses in these studies ranged from 50 - 200 mcg BID. The main difficulty with the interpretation of the oral data with respect to safety is that the bioavailability of oral formulations of Iloprost are low (approximately 16%). Adequate information as to whether the more active of the two diastereomers is preferentially cleared is poorly documented.

For the oral population the mean $\pm$ SD duration of treatment was 15.9 $\pm$ 15.6 weeks (median 8 weeks) and the mean $\pm$ SD daily dose was 173.5 $\pm$ 96 mcg (median 148 mcg). The corresponding duration for the placebo group is not stated. A greater fraction of the oral Iloprost patients than placebo patients did not complete the duration of study (38 versus 25%).

Serious Adverse events (in greater than 1% of either population) are shown below.

Table 6: Serious adverse events (> 1%) incidence in patients treated with oral Iloprost.

	Iloprost (n=2033)	Placebo (n=1128)
Overall	367 (18%)	218 (19%)
Body as a whole	208 (10%)	116 (10%)
Pain in extremity	72 (4%)	48 (4%)
Aggravation reaction	62 (3%)	32 (3%)
Surgery	53 (2%)	30 (3%)
Infection	37 (2%)	15 (1%)
Cardiovascular System	126 (6%)	64 (6%)
Peripheral gangrene	28 (1%)	10 (1%)
Angina pectoris	22 (1%)	5 (< 0.5%)
Digestive System	34 (2%)	19 (2%)
Nervous system	33 (2%)	10 (1%)
Respiratory system	32 (2%)	25 (2%)
Skin and Appendages	31 (2%)	30 (3%)
Metabolic and nutritional disorders	28 (1%)	15 (1%)

Common causes for discontinuation more frequent in the Iloprost than placebo group were: headache (9% versus 1%), dizziness (1.1 versus 0.4%), vasodilatation (4% versus 0%), nausea (7% versus 2%), diarrhea (2.2 versus 0.4%) vomiting (2.2 versus 0.4%). The sum of both the serious and adverse events leading to discontinuation reflect the vasodilatory and gastrointestinal effect of prostanoids; suggesting systemic exposure to active Iloprost diastereoisomers when the mixture is administered orally.

After oral administration there were small differences in laboratory abnormalities. In particular there were 3 subjects with > 5 x ULN in SGOT in the Iloprost group and none in the placebo group. The sponsor notes none of these patients had elevated bilirubin (> 2 mg/dL)

DSI

A single study site was inspected, [ ] and the site was deemed acceptable.

Pediatrics:

Because pulmonary hypertension is an orphan indication, Iloprost was granted a waiver from performing pediatric studies.

Financial Disclosure:

As per Dr. Gordon's review, no financial arrangements were entered into between the sponsor and investigators that could impact on the outcome of the study.

Trade name:

DMETS originally expressed concern about the use of the TRADENAME Ventavis based on orthographic similarities and the possibility of confusion with

Ventolin. Based on reassurance by the sponsor that the distribution of Ventavis will be limited to specialty pharmacies ,which only stock medications for restricted distribution, such as Flolan and Treprostinil and do not normally stock common medications like Ventolin, the likelihood of medication errors is diminished. DMETS accepts the use of Ventavis as a trade name as long as the distribution is limited to such specialty pharmacies.

Additional DMETS comments concerning the proposed packaging of Ventavis are listed at the end of this memo.

Conclusions and Comments:

Approvability of a diastereomeric mixture:

The rationale for the approvability of a diastereomeric mixture was described under Chemistry.

Number of studies:

Only a single study supports approval of the use of Iloprost by inhalation. Approval relies on this study coupled with the benefit observed for Flolan and the suggestion of benefit from treprostinil, who are members of the same class of drugs. Because of the limited data, I have suggested that a conservative approach be taken with respect to limiting the labeling claims.

Population:

The majority of the effect on the primary end-point in the clinical study can be attributed to a beneficial effect in those patients with primary disease. The secondary pulmonary hypertension population that was studied in the single pivotal study had a minimal benefit in considering the primary end point or in considering walk-distance at either pre-dose or post inhalation. Since this population consists predominantly of patients with thromboembolic disease and since no previous prostanoid has been approved for this population, there is insufficient reason to recommend this treatment for the secondary pulmonary hypertension population.

Dose regimen:

Only one dose regimen was studied. An initial dose was 2.5 mcg by nebulization via a HaloLite or its successor ProDose nebulizer, over 4.5 minutes. If the single dose was tolerated the dose was increased to 5 mcg/ treatment over approximately 9 minutes with 6-9 of such inhalations per day. Iloprost was not studied in conjunction with other therapies for pulmonary arterial hypertension. Should the patient's condition deteriorate, there is no information as to whether other medications can be used with Iloprost or whether higher doses or more frequent treatments of Iloprost would be useful. The label should recommend consideration of alternate therapies should the patient's condition deteriorate.

Choice of Inhalers:

The pivotal clinical study (Study # ME97218) employed the HaloLite nebulized. The ProDose nebulizer is predicated on the operating characteristics of the HaloLite

nebulizer and is available in this country. Although some modifications to the microchip disc were required to assure the dose was the same as administered during the clinical trial with the HaloLite nebulizer, the performance characteristics of the ProDose nebulizer appear acceptable.

Other Instructions for dosing:

The ampoule that will be distributed by the sponsor contains 20mcg of Iloprost, which is far greater than needed for a single inhalation treatment (5 mcg). The delivery of Iloprost is not uniform during the time of the single inhalation. Far greater amounts of Iloprost are delivered (approximately 2.8 mcg) during the first portion of an inhalation than is delivered during the latter portion of an inhalation (2.3 mcg). Reliability of delivery of a second inhalation treatment from a single ampoule has not been tested for reproducible delivery of Iloprost. The use of residual Iloprost in the well of the nebulizer at the end of each dose, should therefore, be proscribed by the label.

Interdosing interval:

Based on serum levels, the sum of Iloprost diastereoisomers decrease rapidly after a single inhalation treatment (presumably these levels are reflective of Iloprost concentrations in the pulmonary vasculature). Whether there will remain adequate effects at the interdosing interval is uncertain. In the absence of data that would allow use of Iloprost to be incorporated into a treatment regimen with other drugs, the label should indicate both that timing of dosing should be commensurate with the anticipated need for additional symptom relief, such as when exercise is planned. No recommendation can be made about the concurrent use of Iloprost with other treatments for pulmonary hypertension.

Based on what is known from clinical trials, the minimal time between doses of Iloprost should be two-hours. The maximal number of daily doses should be limited to six - nine per day. The dose of Iloprost per inhalation treatment should be limited to less than 5 mcg, with a total daily dose of < 45 mcg/day.

The benefit of Iloprost at 30 minutes post dose is clearly evident for walking distance and for the composite definition of responder, the primary metric of the study. At the interdosing interval there appears to be a diminishment of benefit and whether there is residual benefit is unclear.

Description of Benefit:

The benefit to a patient based on the single study would suggest that the expectation should be similar to the composite endpoint; a composite of an increase in 10% over baseline walk-distance, an improvement in NYHA classification without the components classified as deterioration.

Withdrawal effects:

Iloprost is administered asymmetrically, with dosing no more frequent than every two hours and a maximum of none daily doses. Patients usually do not have inhalations during the overnight period when they sleep. Although trough measurements of

hemodynamics and walk-distance did not show a rebound effect, it is unclear if the trough is after an overnight fast or after the two hour inter-treatment interval. Whether there is some consequence of withdrawal is unclear.

DMETS Comments:
DMETS comment concerning about additional modifications to the container label and carton labeling follow:

a. Delete use of the terminal zero on the carton labeling (i. e., Contents) and the ProDose Nebulizer Disc (i. e., 5 mg size). b. We recommend reorganizing the information in the net quantity box to read as follows:

b. We recommend reorganizing the information in the net quantity box to read as follows:

NDC10148-101-01
Ventavis (Iloprost
Inhalation Solution
20 mcg/2 ml
100 Single-Use ampules
Discard Any Unused Portion
Rx Only

NDC 10148- 101- 01 Ventavis (Iloprost) Inhalation Solution 20 mcg/ 2 mL 100 Singe-Use Ampules Discard Any Unused Portion Rx Only

c. DMETS notes that the sponsor has submitted a label that will be placed on the ProDose nebulizer disc for our comment and review. We note that the terminal zero should be deleted on the 5 mcg dose. However, DMETS cannot comment whether this is an appropriate label to use with this device.

Appears This Way
On Original

Appears This Way
On Original

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Abraham Karkowsky
12/23/04 01:03:31 PM
MEDICAL OFFICER

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Food and Drug Administration
Rockville, MD 20857

NDA 21-894

Prestwick Pharmaceuticals, Inc.
Attention: Benjamin Lewis, Ph.D.
Senior Director, Regulatory Affairs
1825 K Street N.W., Suite 1475
Washington, DC 20006

Dear Dr. Lewis:

Please refer to your new drug application (NDA) dated September 23, 2005, received September 26, 2005, under section 505(b) of the Federal Food, Drug, and Cosmetic Act for Xenazine (tetrabenazine) Tablets 12.5mg and 25mg.

We acknowledge receipt of your submissions dated:

18-Oct-2005	09-Dec-2005	14-Dec-2005	15-Dec-2005
19-Dec-2005	23-Dec-2005	23-Dec-2005	23-Dec-2005
18-Jan-2006	27-Jan-2006	06-Feb-2006	21-Feb-2006
21-Feb-2006	01-Mar-2006	06-Mar-2006	

We also acknowledge receipt of your submissions dated:

1-Mar-2006	6-Mar-2006	10-Mar-2006	14-Mar-2006
15-Mar-2006	16-Mar-2006		

These latter submissions were not reviewed for this action. You may incorporate these submissions by specific reference as part of your response to the deficiencies cited in this letter.

We completed our review of this application, as amended, and it is approvable. Before the application may be approved, however, it will be necessary for you to respond to the following issues:

CLINICAL

We believe that you have provided substantial evidence of effectiveness for Xenazine as a treatment for chorea in patients with Huntington's Disease (HD).

Specifically, the results of Study 004 are clearly and robustly consistent with this conclusion. Not only is the p-value for the primary contrast extremely small ($p < 0.0001$), but the results clearly favor drug over placebo in 14 of the 15 study sites. In addition, other analyses of the data in this study also document the robustness of this finding. Specifically, we note that upon drug withdrawal at Week 12, patients' chorea scores returned to baseline levels by Week 13, confirming the drug effect seen over the previous 12 weeks. In addition, exploratory analyses document that the responses of patients during the first 11 weeks of Study 007, the open-label extension to Study 004, during which all patients were re-titrated, were essentially identical to the responses seen in the drug treated patients during the titration period in Study 004. This effect in Study 007 was seen in both patients who had previously received active treatment in Study 004 as well as in those who had previously

received placebo. A similar effect was seen for patients enrolled in Study 006, the open-label extension to Study 005. That is, although patients (after their participation in Study 005) were placed back on their best dose in Study 006 (as opposed to being re-titrated, as the patients in Study 007 were), their responses over the first 12 weeks in Study 006 were also essentially identical to those of the drug treated patients in Study 004. Further, although patients were not randomized to fixed dose in Study 004, PK/PD analyses strongly suggest a dose response relationship in that study.

The drug effect seems to be present regardless of the baseline degree of severity of the chorea.

We recognize that the results of the analyses of Study 005 do not meet the usual test for being considered "positive" (p=0.078). However, we note your observation that patients in Group 2 were not treated in compliance with the protocol (that is, placebo was inadvertently substituted for active drug on the morning of Day 3), and we agree that the protocol-specified prospective analysis is therefore inappropriate. We believe that the comparison of Group 1 to Group 3 on Day 3 is an appropriate post hoc analysis under these circumstances, because it is consistent with the rationale for your prospective analysis (that is, it compares patients off drug [Group 1] with patients continuing on treatment [Group 3]). Although the results of this analysis do not achieve nominal statistical significance (p=0.11), the estimate of the treatment effect is essentially identical to that seen in Study 004 (mean between treatment difference of about 3.5 points). In this case, we believe that the absence of statistical significance for this comparison is related to the extremely small sample size (12 patients in Group 1 and only 6 patients in Group 3).

We believe, given the results described above, that the findings establish the effectiveness of Xenazine as a treatment for the chorea of HD, under FDAMA's provision that substantial evidence can consist of the results of a single adequate and well-controlled investigation plus confirmatory evidence. We believe that the statistically strong result of Study 004, its marked internal consistency, as well as the results of Study 005, provide the necessary confirmatory evidence required by this provision of the Act.

Despite the documented effect on chorea, there remain troubling questions about the utility and ultimate approvability, of this application.

In particular, we note that there was a consistent tendency for the results of the analyses of multiple secondary outcomes to favor placebo in Study 004. Specifically, the between-treatment comparisons on the Cognitive Assessment (UHDRS Part 2), the Behavioral Assessment (UHDRS Part 3), the Functional Assessment (UHDRS Part 4), the Independence Scale (UHDRS Part 5), the Functional Capacity (UHDRS Part 6) all numerically favored placebo, and the comparisons on the Cognitive Assessment (UHDRS Part 2) and the Functional Assessment (UHDRS Part 4) actually achieved nominal statistical significance in favor of placebo (p=0.025 and p=0.018, respectively). We also note that there were no patient – rated measures of overall benefit in Study 004. These results, taken together, raise serious questions, not only about the overall utility of Xenazine's effect on chorea, but also, of course, about Xenazine's capacity to cause harm in these patients. We acknowledge that the (negative) effects seen on these secondary measures appear to be numerically small, but we do not have a good understanding of the effects on patient functioning of these sorts of changes. We also do not have data on the consequences of long-term treatment with Xenazine. If overall patient functioning continues to worsen (in the face of reasonable control of the chorea) as a result of chronic treatment, we are not confident that such deterioration could easily be detected clinically (because detailed neuropsychiatric testing may be necessary to detect it). In such a case, clinical deterioration may continue unnoticed; when it does become manifest, the patient's clinical condition would very probably be attributed to progression of the underlying HD.

Beyond the question of these specific ways in which treatment with Xenazine may harm patients, we are concerned with Xenazine's capacity to cause other, serious, adverse events.

In particular, among the numerous adverse events seen in association with the use of tetrabenazine, we note parkinsonism, akathisia, depression, and dysphagia (with associated aspiration pneumonia). Although we acknowledge that the incidence of some of these events in Study 004 is not significantly different from placebo

(e.g., parkinsonism, dysphagia) the incidence of others is substantially greater in the drug-treated patients than in the placebo patients (e.g., depression: 15% vs 0; akathisia: 9% vs 0). Further, it is not clear that other events coded differently from akathisia do not, in fact, represent the same phenomenon (e.g., agitation, anxiety, irritability). All of these events are consistent with the pharmacologic effects of the drug, and the incidence of these events increases with increasing duration of use. We acknowledge, of course, that the long-term safety data were collected in an open-label, uncontrolled setting, and also that these can themselves be manifestations of progressive HD. For these reasons a definitive conclusion about causality clearly can not be made at this time. Nonetheless, we are concerned that these events may be drug-related.

We are particularly concerned about the ability of practitioners to readily identify these events and consider the possibility that they may be drug -related. We would agree that, should these events occur relatively acutely after treatment initiation (or dose increase), the prescriber might consider them drug related (and take the appropriate action). However, to the extent that they might be drug-related, but occur slowly over time, it is less likely that they will be considered potentially drug-related and more likely to be considered related to disease progression. In such a scenario, the possibility that the specific symptom might reach a severe stage (with the possibility that it may become irreversible), or result in a serious outcome even if reversible (e.g., depression leading to suicide), is raised. (In the case of parkinsonism, an article in the literature (Satou T et al. Exp Toxic Pathol 53:303-308, 2001) suggests that there is irreversible damage to the substantia nigra pars compacta in Wistar rats following 7 daily i.p. doses of tetrabenazine.)

Also, in regard to dysphagia specifically, we note the disturbing finding that Dr. Jankovic did not systematically record episodes of dysphagia in many of his patients because he considered it to be a symptom of progression of the underlying HD. Because his experience represents a large portion of the clinical experience submitted in this application, we are concerned that the incidence of dysphagia (which can have devastating clinical consequences) may be significantly underestimated.

For all of these reasons, then, we are not sure Xenazine can be used safely, even with labeling that describes, as accurately as possible, the known risks of its use. Because we are unable to reach a definitive conclusion about the ultimate approvability of the application at this time, we plan to discuss your NDA at a public meeting of the Peripheral and Central Nervous Systems Advisory Committee (PCNSAC). We will attempt to arrange this meeting as soon as possible.

CMC

1. [redacted]

2. [redacted]

3. Approval from a CMC standpoint will be contingent on the overall recommendation on establishment from the Office of Compliance.

NON-CLINICAL

Prior to approval, you will need to address the following nonclinical issues:

1. There is a lack of adequate in vivo metabolism data in the animal species used in the definitive nonclinical studies. There is a similar lack of metabolism data in humans. You need to provide additional data identifying and quantitating the major circulating metabolites in animals and humans. These data are needed in order to determine the relevance (and adequacy) of the nonclinical studies to an assessment of human risk. In particular, there is concern that the potential toxicity of the major circulating drug-related material in humans (peak 16) may not have been adequately assessed in animals.

2. The 26-week oral toxicity study is the only definitive toxicity study conducted in rats. Therefore, it is particularly important that you provide the data from this study in a complete and accurate manner. The following deficiencies were identified in the report of the study:

    a. The reporting of clinical signs is incomplete. For example, several instances of convulsions observed in two high-dose animals were not listed in the summary table. Similarly, instances of "lethargy" were noted in the summary table, but not in any individual animal line listing. You need to address the apparent discrepancies between the summary of clinical signs and the individual animal line listings.

    b. The study report did not include a signed Pathologist's Report. In order to document the gross pathology and histopathology findings in the chronic study, you need to provide a copy of this report.

3. You conducted a 14-day oral study of tetrabenazine to assess toxicokinetics and effects on serum prolactin in rats (Covance Study # 7425-114). The toxicokinetics data have been provided, but the serum prolactin data have not. You need to submit a final report of the serum prolactin data. These data are important for the interpretation of the results of the chronic toxicity study in rats.

4. The published findings of Satou et al. (Satou T et al. *Exp Toxicol Pathol* 53(4):303-308, 2001) raise a concern that tetrabenazine may have neurotoxic effects. Therefore, it is particularly important to understand how extensively the brain was examined in the 26-week and 9-month oral toxicity studies in rats and dogs, respectively. The reports of these studies do not provide sufficient detail regarding the methodology used in the microscopic examination of brain. You need to document that the microscopic examination of brain in the chronic studies was conducted using techniques sensitive enough to have detected, if present, neuropathological findings similar to those reported by Satou et al (2001).

5. The equivocal finding in females in the in vivo micronucleus assay in rat needs to be further investigated, particularly considering the lack of carcinogenicity data on tetrabenazine. The in vivo micronucleus assay needs to be repeated exploring a range of doses. Although the equivocal finding was only in females, it is difficult to understand why females would be more sensitive than males based on the available plasma exposure data; therefore, we ask that you include both males and females in the repeat assay.

6. You need to commit to initiating carcinogenicity studies. Your protocol for a 26-week p53 transgenic mouse assay has been reviewed by the Division and the Executive CAC; minutes of the Executive CAC meeting were sent to you on October 27, 2005. You have recently submitted a protocol for a 2-year carcinogenicity study in rats that is currently under review. You need to commit to a timeline for conduct of the studies and submission of final reports of these studies. Final study reports would not be required prior to approval.

CLINIAL PHARMACOLOGY & BIOPHARMACEUTICS

Before approval, we ask you to address the following:

1. Clarify the rotation speed at which the dissolution method was generated (previously requested on 1/2/06). If you have data to support the proposed rotation speed and agreement is reached between us regarding dissolution specifications, the method and agreed upon specifications can be accepted as interim method and specifications. The recommended dissolution method and specifications are as follows:

   Apparatus: USP Apparatus 2 (Paddles)
   Medium: 0.1 M HCl
   Volume: 900 ml
   Rotation Speed: 50 rpm
   Specification: ≥[redacted](Q) in 30 minutes

2. Since the 25 mg tablet is scored, you should demonstrate dissolution similarity (with f2 testing and using the interim dissolution method above) between 2 half-tablets and 1 whole 25 mg tablet.

3. The P16 component, identified as the largest circulating component in the mass balance study, should be characterized. In addition, the extent to which the mono- and bis-dealkyl tetrabenazine metabolites (and other individual metabolites) are circulating should be clarified.

4. You should submit adequately performed *in vitro* metabolism studies to address the potential for inhibition or induction of P450s by TBZ and its metabolites. You should also characterize the *in vitro* metabolism of TBZ and its metabolites as well as the role of PgP in TBZ disposition. Finally, you should adequately address the role for TBZ as a PgP inhibitor *in vitro*. There is currently insufficient information to allow for adequate labeling regarding the potential for drug interactions. Please see our comments below about performing the in vitro drug metabolism studies (communicated to you in an email of 12/21/05).

   1. You have not taken a step-wise approach to understanding the metabolism of TBZ or its metabolites. The preferred first approach would be to directly identify metabolites after incubation with hepatocytes or liver slices. Subsequent studies can also eliminate non CYP oxidative pathways.
   2. The studies to evaluate CYP pathways of TBZ and HTBZ metabolism are methodologically deficient. It is recommended that recombinant enzymes not be used alone, but in combination with other methods (such as use of inhibitors) for identifying drug metabolizing P450 isozymes. In addition, the probes used as controls in the submitted studies are not classical, preferred probes, and you have not provided justification, so it is difficult to understand the acceptability of the reactions.
   3. Studies characterizing the metabolism of TBZ *in vitro* should include measurement of the formation of metabolites (including the oxidative metabolites of TBZ and the oxidative metabolites of HTBZ) to identify the pathways by which they are formed.
   4. You should follow-up the results of the submitted studies with *in vitro* inhibition studies that use well accepted methodology and preferred substrates to confirm lack of involvement of TBZ and its metabolites in inhibition of P450s.
   5. The *in vitro* study of TBZ inhibition of PgP provided from the literature was not conducted with methods that are in agreement with current Agency thinking. The *in vivo* TBZ-digoxin interaction study was performed with a low dose of TBZ, and does not allow for conclusions

about higher doses that will be used clinically. You should perform an adequate *in vitro* inhibition study using preferred methodology to determine the need for further *in vivo* study.

6. The results of adequate *in vitro* drug metabolism studies will guide the need for further *in vivo* drug interaction studies.
7. Since CYP2D6 appears to be involved in the metabolism of TBZ and HTBZ, we recommend genotyping for CYP2D6 in future TBZ clinical trials.
8. The thorough QT study did not assess exposure to TBZ or metabolites outside of the ranges that might be normally observed after administration. The results of the in vitro drug metabolism studies may help guide decisions regarding the need and approach for further metabolically-based evaluation of QT.

Phase 4 Commitments
NON-CLINICAL

We ask that you address the following issues as Phase 4 commitments:

1. Submission of final study reports for the 26-week p53 transgenic mouse assay and the 2-year carcinogenicity study in rats.

2. Conduct of a fertility and early embryonic development (to implantation) study. You should commit to a timeline for conduct of the study and submission of the final study report.

3. The following apparent discrepancies in the report of the pre- and post-natal development study need to be addressed:

   a. the lack of corpora lutea and preimplantation loss data in F1 females. These data need to be submitted if collected.

   b. the number of stillbirths versus early postnatal deaths. You need to specify which pups were determined to be stillborn due only to the lack of milk in the stomach versus those determined to be stillborn by the lack of lung floatation (with or without lack of milk in the stomach); the lack of milk in the stomach alone does not necessarily indicate a stillborn pup. In addition, you need to explain why the summary table (page 39) indicates a dose-related increase in stillbirths, whereas the individual line listings (page 204-207) fail to indicate a stillbirth in any litter.

   c. apparent discrepancies in the data for individual dams, low-dose female B73509, mid-dose female B73526, and high-dose female B73557. You need to provide all data (including pregnancy, litter, and final disposition) for these dams.

Although not needed prior to approval, we ask that you address these issues in a timely manner.

CLINICAL PHARMACOLOGY

We ask that you address the following issues as Phase 4 commitments:

1. Perform an in vivo study of the effect of CYP2D6 inhibition on TBZ disposition using a strong CYP2D6 inhibitor since CYP2D6 inhibition may increase the exposure to the inactive β- HTBZ relative to the active moiety α-HTBZ (based on evaluation of plasma concentrations in Phase III studies).

2. Evaluate the clinical relevance of CYP2D6 inhibition after administration of TBZ in vivo using a sensitive CYP2D6 substrate (such as desipramine) since in vitro studies suggest involvement of CYP2D6.

3. Other *in vivo* drug interaction studies should be guided by the results of the *in vitro* drug metabolism studies, in agreement with the Agency.

4. The discriminatory ability of the interim dissolution method should be determined in order to determine the final dissolution specifications.

In addition, it will be necessary for you to submit draft labeling revised as attached.

Please submit the final printed labeling (FPL) electronically according to the guidance for industry titled Providing Regulatory Submissions in Electronic Format - NDA (January 1999). Alternatively, you may submit 20 paper copies of the FPL as soon as it is available but no more than 30 days after it is printed. Please individually mount ten of the copies on heavy-weight paper or similar material.

If additional information relating to the safety or effectiveness of this drug becomes available, revision of the labeling may be required.

When you respond to the above deficiencies, include a safety update as described at 21 CFR 314.50(d)(5)(vi)(b). The safety update should include data from all non-clinical and clinical studies of the drug under consideration regardless of indication, dosage form, or dose level.

1. Describe in detail any significant changes or findings in the safety profile.

2. When assembling the sections describing discontinuations due to adverse events, serious adverse events, and common adverse events, incorporate new safety data as follows:

    - Present new safety data from the studies for the proposed indication using the same format as the original NDA submission.
    - Present tabulations of the new safety data combined with the original NDA data. Include tables that compare frequencies of adverse events in the original NDA with the retabulated frequencies described in the bullet above.
    - For indications other than the proposed indication, provide separate tables for the frequencies of adverse events occurring in clinical trials.

3. Present a retabulation of the reasons for premature study discontinuation by incorporating the drop-outs from the newly completed studies. Describe any new trends or patterns identified.

4. Provide case report forms and narrative summaries for each patient who died during a clinical study or who did not complete a study because of an adverse event. In addition, provide narrative summaries for serious adverse events.

5. Describe any information that suggests a substantial change in the incidence of common, but less serious, adverse events between the new data and the original NDA data.

Provide a summary of worldwide experience on the safety of this drug. Include an updated estimate of use for drug marketed in other countries.

Provide English translations of current approved foreign labeling not previously submitted.

In addition, submit three copies of the introductory promotional materials that you propose to use for this product. Submit all proposed materials in draft or mock-up form, not final print. Send one copy to this division and two copies of both the promotional materials and the package insert directly to:

Division of Drug Marketing, Advertising, and Communications, HFD-42
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857

Within 10 days after the date of this letter, you are required to amend this application, notify us of your intent to file an amendment, or follow one of your other options under 21 CFR 314.110. If you do not follow one of these options, we will consider your lack of response a request to withdraw the application under 21 CFR 314.65. Any amendment should respond to all the deficiencies listed. We will not process a partial reply as a major amendment nor will the review clock be reactivated until all deficiencies have been addressed.

Under 21 CFR 314.102(d), you may request an informal meeting or telephone conference with this division to discuss what steps need to be taken before the application may be approved.

The drug product may not be legally marketed until you have been notified in writing that the application is approved.

If you have any questions, call CDR Teresa Wheelous, Sr. Regulatory Project Manager, at (301) 796-1161.

Sincerely,

{See appended electronic signature page}

Robert Temple, M.D.
Office Director
Office of Drug Evaluation I
Center for Drug Evaluation and Research

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

Robert Temple
3/24/2006 05:58:02 PM

Drug Information Journal, Vol. 36, pp. 517–534, 2002
Printed in the USA.

0092-8615/2002

REPORT OF A WORKSHOP ON CONFIRMATORY EVIDENCE TO SUPPORT A SINGLE CLINICAL TRIAL AS A BASIS FOR NEW DRUG APPROVAL*

CARL C. PECK, MD
Director, Center for Drug Development Science, Georgetown University Medical Center, Washington, District of Columbia
JILL WECHSLER, MA
Editor and Writer, Chevy Chase, Maryland

A novel approach for more efficient, economical, and faster clinical evaluation of new drugs is to couple effectiveness data from phase 2 and other studies with a phase 3 single clinical trial (SCT) for regulatory approval. Sanctioned in the Food and Drug Administration Modernization Act (FDAMA), this approach challenges the traditional requirement for two controlled phase 3 trials for approving a new therapy. Academic, industry, legal, and regulatory experts participated in a workshop to define adequate confirmatory evidence of effectiveness to support regulatory approval based on data from an SCT. Participants examined qualities of confirmatory evidence and SCTs and implications of this model for clinical safety information. Acknowledging the evolutionary nature of scientific evidence of effectiveness, participants identified risks and benefits of this approach, concerns of FDA and pharmaceutical companies, and policy changes that may further encourage widespread use of the confirmatory evidence-SCT model. For example, these policy changes include explicit FDA publication of the basis for effectiveness determinations of new drugs, use of end-of-phase 1 industry-regulator meetings for prospective planning of the confirmatory evidence-SCT program, and more use of established pharmacological knowledge to qualify confirmatory evidence biomarkers and surrogate endpoints.

Key Words: Confirmatory evidence; Single clinical trial; Evidence of effectiveness; FDAMA Section 115a; Drug approval standards

A WORKSHOP ENTITLED "Confirmatory Evidence (CE) to Support a Single Clinical Trial (SCT) as a Basis for Drug Approval—An Exploratory Workshop" was held January 15–16, 2002 at Georgetown University in Washington, District of Columbia.

Reprint address: Carl C. Peck, MD, director, Center for Drug Development Science, Georgetown University Medical Center, Med-Dent NE 405, 3900 Reservoir Road NW, Washington, DC 20007–2195.

*This article is based upon the workshop entitled "Confirmatory Evidence to Support a Single Clinical Trial (SCT) as a Basis for Drug Approval," January 15–16, 2002, Washington, DC. This workshop was supported by an unrestricted educational grant from the Yamanouchi USA Foundation and unrestricted supporting grants from AstraZeneca, Inhale, Lilly, Immunex, Novartis, and Wyeth-Ayerst.

Sponsored by the Center for Drug Development Science (1), the workshop aimed to achieve the following objectives:

1. Explain the legislative and legal implications of using confirmatory evidence and an SCT,
2. Define the nature and sources of "evidence of effectiveness,"
3. Identify satisfactory requirements for confirmatory evidence to support an SCT in the determination of drug effectiveness,
4. Define requirements for an SCT in conjunction with adequate confirmatory evidence, and
5. Define requirements to establish an adequate safety database, assuming that effectiveness is independently affirmed.

Organized by the Georgetown University Center for Drug Development Science and academic collaborators, the conference benefited from strong participation by leaders of FDA's centers for drug and biologics evaluation and research (CDER, CBER) and pharmaceutical industry research executives. Speakers examined the legal context supporting the confirmatory evidence-SCT model and FDA's history in setting standards for evidence. There was considerable discussion about statistical and modeling developments to support the derivation of effectiveness evidence from non-phase-3 trials. An important focus of dialogue was the need to better define what constitutes credible scientific confirmatory evidence to support phase 3 empirical clinical trial data and how that fits into the drug development and approval planning processes.

The workshop commenced with plenary presentations by experts on drug development and regulation from academia, the pharmaceutical industry, regulatory agencies, and the legal profession. In four breakout sessions, participants discussed objectives 2 to 5 and prepared reports that were presented and openly discussed in the final plenary session. The workshop agenda, presentations of plenary speakers, and breakout group reports are posted on the Center for Drug Development Science Web site (1). This summary of the workshop was derived by the authors from the plenary presentations, breakout session reports, and open discussions. Draft text and editorial reviews of draft versions of breakout session reports and the overall summary were provided by individuals identified in the list of presenters and contributors (see Acknowledgments).

BACKGROUND AND PLENARY SESSION VIEWS

The unmet medical needs of patients and the rising cost of drug development continue to stimulate efforts to streamline the pharmaceutical research and review process. The challenge for pharmaceutical/biotech firms, regulators, and academic researchers is to design more efficient and economical clinical research programs that preserve safety and ensure effectiveness of medical products coming to market. Scientific advances in the understanding of disease biology and mechanisms of therapeutic interventions, coupled with advanced data analysis, modeling, and simulation methods, offer new possibilities for devising more targeted and more informative clinical research programs that avoid ineffectual and redundant efforts. These developments are encouraging the pharmaceutical research community to explore new approaches for accelerating the clinical development process. A lead proposal is to integrate evidence of effectiveness of a new therapy from phase 2 and other studies, with a single, well-designed phase 3 clinical trial.

The idea of intentional utilization of scientifically-derived effectiveness information from all relevant sources in a new drug development program, although not a new concept, was formally proposed in the course of Congressional hearings in the mid-1990s concerning the role of the FDA in reducing drug development time and costs. This led to the concept of coupling such information ("confirmatory evidence") with an empirical showing of effectiveness in an SCT.

Members of Congress included a provision in the FDAMA to clarify FDA's authority to use the confirmatory evidence-SCT

model (2). However, some Congressional supporters of the measure have expressed their concern that their intent has not yet been fully implemented. Pharmaceutical and FDA representatives acknowledge that five years after passage, the legislation has not had a significant impact on the design of drug industry research and development (R&D) programs.

One reason appears to be a general risk-averse posture of major pharmaceutical companies, which admit a willingness to spend more money on large traditional clinical study programs to assure speedy market approval by FDA. Big pharma industry leaders display a general reluctance to use the confirmatory evidence-SCT model except in limited special cases. Despite FDA's view that sufficient guidance has been offered to the industry, some observers have detected a lack of consistency within FDA regarding the definition of confirmatory evidence as it relates to single clinical trial approvals, leading to mixed messages to the industry.

While FDA officials support efforts to streamline clinical study programs, they are cautious of approving a New Drug Application (NDA) based on other than phase 3 clinical trial data. There appears to be general uncertainty among both regulators and manufacturers over what constitutes sufficient "confirmatory evidence" to support a drug development program based on one phase 3 clinical study.

Most workshop attendees agreed that greater emphasis on exploratory studies of drug candidates and target disease states at early stages in clinical development—studies designed to increase understanding and quantification of mechanisms—could lead to more rational and successful phase 3 studies and greatly reduce wasted expenditures on drug development. However, many wondered whether the availability of the confirmatory evidence-SCT paradigm alone would effectively encourage that change in emphasis. Several speakers suggested that leaders at FDA and other regulatory authorities need to instill a more encouraging, flexible, and open-minded approach to managing the review process among agency review staffs if sponsors are to benefit from investing in this paradigm shift. Regulators, however, are committed to maintaining high standards for sponsors to document product safety and effectiveness and often are reluctant to accept nontraditional evidence. Clearer published guidances and information from FDA could help researchers understand better what constitutes acceptable confirmatory evidence of clinical effectiveness, other than merely a replicated clinical study.

One proposal offered was to employ economic time and value analysis of different drug development paradigms to understand more fully the benefits of adopting new approaches to the R&D process. A general conclusion was that both sponsors and regulators should look more to expanded phase 2 study data and analyses to improve the design of phase 3 studies.

LEGISLATION AND POLICIES SHAPE EFFECTIVENESS STANDARDS

The standard that substantial evidence of effectiveness must be demonstrated prior to obtaining FDA approval of a new drug dates back to enactment of the 1962 Drug Amendments (Public Law # 87–781). Under the 1938 Federal Food, Drug, and Cosmetic Act (FDC Act), a manufacturer was required to show only that a drug product was safe to market it in the United States. The 1962 Drug Amendments called for all new drugs to show "substantial evidence" of effectiveness from "adequate and well-controlled investigations." This generally has been interpreted by FDA to mean that most new drugs require data from at least two adequate and well-controlled studies.

Legal experts have debated how much flexibility Congress intended in this language and whether it really means that new drug approval requires two separate clinical trials. A CDER official explained at the Georgetown confirmatory evidence workshop that FDA's long-standing interpretation is that Congress generally intended to require at least two ade-

quate and well-controlled studies to establish effectiveness. Despite this long-standing interpretation, FDA has approved NDAs based on the results of a single phase 3 study. Such actions generally have been limited to approval of applications for orphan drugs; for treatments for "serious and life-threatening diseases or conditions"; or where a single study produced statistically very strong responses.

FDA has explained its views on evidence of effectiveness in several agency documents. In 1988, FDA published a guidance document, entitled *Guideline for the Format and Content of the Clinical and Statistical Sections of an Application* that discusses study types and correct presentation of trial data (3). However, the guidance provides little information on the quantity and quality of evidence of effectiveness that can be derived from non-phase 3 trials to support approval. In discussing one example of a drug approval based upon the results of a single trial (FDA's approval of timolol for reduction of postinfarction mortality), the document states that: "There [are] instances in which a single particularly persuasive study [may be] accepted in support of claims because the study was considered unrepeatable on ethical grounds."

In August 1995, FDA issued a *Federal Register* notice (4) which offered similar clarification to the "substantial evidence" standard. The 1995 notice states that while a second study may well be needed to replicate results demonstrated in a first study, in some instances "it is possible to replicate results within one large, well-designed, multi-center study." FDA has emphasized that this approach can be successful only when results are strong.

In March 1997, FDA issued a draft version of its *Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* (commonly referred to as the "Evidence Document"). This draft document was responsive to efforts then underway to include language in the 1997 FDAMA legislation to clarify that FDA could determine that data from one clinical trial provide "substantial evidence" of effectiveness, combined with support from "relevant science" and other "confirmatory evidence."

FDA finalized this guidance document in May 1998 (5). The authors of the FDAMA provision believed that amending the FDC Act's effectiveness standard would encourage sponsors to develop more targeted and streamlined drug development and testing programs. However, it appears that FDA officials regarded the measure more as codifying current agency practice rather than as a change in the approval standard.

Congressional authors of the FDAMA "effectiveness" provision have expressed disappointment that the measure, in fact, has done little since its enactment in 1997 to reduce the types and amount of data required to affirm drug effectiveness, according to Representative Richard Burr (R-NC), who sent a letter to the confirmatory evidence workshop (posted on the Center for Drug Development Science's Web site) expressing his support for the intent of FDAMA Section 115(a). The provision, he said, sought to encourage FDA to assist companies in minimizing extraneous data and information collection and filing, and to encourage more complete use of all relevant effectiveness data. Congress intended that such efforts would improve public health by reducing the number of patients in clinical trials, increasing the number of new drugs under investigation, reducing drug development time, reducing the cost of drug development and, ultimately, lowering the cost of new drugs (in a letter to workshop participants, January 15, 2002).

FDAMA Section 115(a) was written to enable this streamlined study paradigm by affirming that approval based upon confirmatory evidence plus a single trial is not confined to certain diseases, or to cases where the data are so compelling that it is unethical to repeat a placebo-controlled study, explained Frank Sasinowski, a former FDA regulatory counsel. Congressional committee reports on FDAMA defined confirmatory evidence to be "scientifically sound data from any investigation in the NDA that pro-

vides substantiation as to the safety and effectiveness of the new drug." This evidence may consist of "earlier clinical trials, pharmacokinetic data, or other appropriate scientific studies" (6). However, this language has proved to be insufficient to fully clarify what constitutes confirmatory evidence. Sasinowski noted that the legislation leaves it up to the FDA [via authority delegated from the Secretary of the Department of Health and Human Services] to decide when and where to apply the confirmatory evidence-SCT standard and that the agency is not compelled to do so.

Although the May 1998 final evidence document describes some reasons for FDA's traditional two-study research requirement, Sasinowski noted that FDA has never fully defined the quantum of evidence needed to establish effectiveness. The 1998 evidence document fails to define confirmatory evidence and includes only scenarios limited to modifications of already approved drugs. No consideration of approval of new molecular entities based on a confirmatory evidence-SCT model is presented. Consequently, there remains significant misunderstanding as to what FDA's policy is with regard to approval based on this approach.

DOCUMENTING EFFECTIVENESS AND SAFETY

This shortcoming has generated an ongoing debate over the scope and type of evidence required to document the effectiveness of a new medical product. The procedures and science involved in designing clinical studies to produce convincing evidence of effectiveness have been evolving since the 1960s. Today, statisticians generally employ statistical tests to show that prescription of a specific drug to a group of study subjects causes a clinical benefit relative to a control group which does not receive the drug. The validity of these statistical methods depends essentially on the effectiveness of the randomization procedures in the basic study design.

FDA requires that clinical studies provide credible evidence that a beneficial clinical effect can be expected in future patients. Traditionally, regulatory authorities have believed that if a certain clinical benefit can be replicated in two randomized controlled clinical trials empirically supporting effectiveness, this demonstrates sufficient evidence of future replicability. One statistically strong clinical trial is enough to document empirical certainty (ie, that an observed beneficial effect in a drug-treated group was due to the drug and not to chance), explained Professor Lewis Sheiner (University of California at San Francisco), a leading expert on the confirmatory evidence-SCT model.

However, FDA has usually required more than one clinical empirical trial to provide sufficient "causal certainty" to support approval of an NDA. Such certainty requires that the benefit demonstrated empirically will extrapolate in time and space. The only rational basis for believing in such extrapolation is that a drug benefit is due to one or more intrinsic pharmacological property(ies) of the drug and drug-response "system," that is, the patient. Replication of a randomized clinical trial is one way to establish pharmacological causality, that is, to rule out that some extraneous factor in the test or control group of a first trial could have been responsible, along with the drug effect, for the beneficial clinical response seen in the drug treated group in that trial.

Peck, Sheiner, and Rubin have proposed another strategy that can provide superior causal evidence of effectiveness: coupling early studies of pharmacological action with one empirical phase 3 trial (7). They advance the idea that scientifically-sound pharmacological phase 2 studies, coupled with observations compatible with pharmacological action in a single phase 3 trial demonstrating clinical effectiveness, can better provide this information.

To illustrate the concept, they propose taking into account randomized, blinded phase 2 "learning trials" that document dose- and/or concentration-response relationships. These can provide evidence of graded pharmacological effects on causal chain biomarkers, surrogate endpoints, or clinical outcomes.

Such causal evidence of effectiveness may be particularly persuasive when it is demonstrated that biomarkers change in the anticipated direction in response to graded dosages, especially when prospectively incorporated in both phase 2 and phase 3 studies. If such a paradigm (coupling studies of pharmacological action with one empirical phase 3 trial) were to be accepted by FDA as evidence of effectiveness, it might encourage sponsors to employ more scientific methods and models in phase 2 clinical trials, with the expectation of less numerous phase 3 trials.

This "learning while confirming" model raises a number of statistical and design issues, which Sheiner identified as important concerns. This includes adequacy of individual biomarkers that link with clinical benefits. However, he noted that the generally accepted two-clinical-trial approach runs the risk of repeating earlier errors and often results in excessive research efforts. He believes that the confirmatory evidence-SCT model will strengthen the drug approval process, lead to better designed and executed phase 2/3 trials, and, through this more efficient process, free up resources for better safety evaluation. Moreover, this approach will provide greater mechanistic understanding of how a class of drugs works, which will facilitate development of new molecules in a drug class and help establish surrogacy for rapidly responding and inexpensive biomarkers.

FDA CONSIDERS ONE PHASE 3 STUDY USUALLY INSUFFICIENT

Robert Temple (CDER) said that a second empirical clinical trial should not be an exact repetition of the first study, but that it should provide independent substantiation of an initial study result. Such substantiation is needed to rule out unidentified biases in study design, chance results, peculiarities to a certain study site, and outright fraud. Temple acknowledged that such concerns were greater in the past when clinical trial designs often were mediocre and marked by nonspecific endpoints, inadequate blinding, unclear rules on analysis, and sketchy protocols.

In addition to reducing the potential for error, FDA prefers that sponsors conduct more than one phase 3 study to gain information that is generalizable to additional patient populations, as well as to obtain additional safety data. FDA officials consider it important to learn about how a drug works with other drugs, how it varies with disease severity, how an effect is maintained with continuing therapy, and how a drug affects different endpoints. Although FDA typically receives NDAs with more studies than it requires, often this is because companies themselves choose to conduct multiple studies for pharmacoeconomic and other marketing purposes.

According to the FDA speakers, whether data to provide independent substantiation of a clinical study constitute a second phase 3 clinical trial or confirmatory evidence may be a semantic issue. For instance, a well-controlled phase 2 study with a clinical endpoint is usually considered as one of two required studies, since FDA does not mandate that two studies necessarily must be conducted during phase 3. FDA has accepted confirmatory evidence derived from other controlled studies of different doses, combination treatments, and related diseases or other phases of the same disease. The credibility of the confirmatory evidence varies, according to how similar the subsequent study is to the initial dose, disease, or treatment. FDA may require fewer effectiveness data for the tenth drug in a class, but also may have questions about how greatly agency reviewers should rely on clinical results from pharmacologically-related drugs.

Overall, FDA is skeptical that confirmatory evidence from studies of pharmacological action is stronger evidence than clinical trial empirical replication. Few drugs are thought by FDA to have pharmacological effects that can be documented with sufficient rigor to link quantitatively to clinical benefit. For example, there have been numerous plausible mechanisms of action proposed for sep-

sis drugs that have failed to be affirmed in phase 3 effectiveness trials.

FDA regards existing law as sufficiently flexible to allow the agency to approve a new drug based on a strong single phase 3 study or a study supported by other clinical evidence. Drug classes that have established long-term benefit, such as estrogens, lipid-lowering statins, ace inhibitors, and antihypertensives may provide cases for relying on pharmacologic effects for additional members of the class.

FDA claims it has displayed flexibility in setting evidence requirements, supported by several cases where the agency approved new drugs based on single studies. These are usually large, independently conducted multicenter mortality trials, most involving cardiovascular therapies. The studies displayed strong empirical results with internal consistency across multiple outcomes, were reviewed by independent drug monitoring committees, and had such good results that the sponsors faced ethical barriers to running another placebo-based trial. FDA considers that its evidence guidance of 1998 (3) describes how evidence other than that derived from two phase 3 trials could be "appropriate" to support a particular claim or product.

THE CONFIRMATORY EVIDENCE-SCT MODEL AND SAFETY ISSUES

In addition to obtaining effectiveness data, clinical development programs aim to provide sufficient understanding of safety issues that might affect drug use once on the market. Peter Honig (CDER) explained that clinical trials generally study too few patients for too short a time to provide a completely adequate safety database. Moreover, most premarket studies involve homogenous populations that provide little information on what safety issues might emerge when the drug is used in medical practice. Even when safety issues fail to arise in clinical studies, problems may appear during marketing, particularly with drugs intended for long-term treatment of chronic conditions. When more patients are studied in clinical trials, it is more likely that a rare adverse event may appear. An increased understanding of the risk factors related to a new therapy can assist FDA and the sponsor in designing clinical studies that adequately explore patient factors (age, sex, race, genetic vulnerabilities, target illness, comorbidities) and drug factors (dose, plasma level, duration, concomitant medications, route of elimination) that are most likely to raise safety issues.

The Center for Drug Development Science's viewpoint is that by conceptually separating the effectiveness determination from that of safety using the confirmatory evidence-SCT approach, conserved resources may be redirected toward efforts to establish an improved safety database. In the traditional drug development paradigm (dozens of clinical trials in phases 1 and 2, accompanied by several large empirical phase 3 trials), safety and effectiveness are coupled. However, the emphasis on effectiveness considerations (power, precision) often consigns the study population to nonrepresentative cohorts. In the confirmatory evidence-SCT model, once effectiveness is established, research can focus on the deliberate evaluation of safety through a large (>10000 subjects) very simple trial under the typical conditions of medical use. This type of study can use a short case report form with only a few pages that captures only significant adverse events. Honig explained that FDA has found that large simple trials conducted in Phase 4 may help improve the safety database. However, he noted that even very broad postmarketing studies do not override the public safety need for FDA to be able to evaluate an adequate clinical database to assure that a drug is safe and effective at the time of initial marketing.

THE EUROPEAN VIEW

While FDA and manufacturers have been debating these issues in the United States, regulatory authorities in the European Union (17 national and one supranational agency) are examining similar concerns as part of efforts

to streamline and accelerate drug development and approval. An expert working group of the Committee for Proprietary Medicinal Products (CPMP) has developed a "Points to Consider" document to advise when one empirical phase 3 trial may provide sufficient evidence of effectiveness (8). Armin Koch (German Federal Institute for Drugs and Medical Devices) explained that the CPMP working group had adopted a progressive position in stating that the minimum requirement for phase 3 data is generally one controlled study with statistically compelling and clinically relevant results. The group's paper concludes that "there is no formal requirement to include two or more empirical phase 3 studies in the phase 3 program."

At the same time, Koch acknowledged that it may be prudent for sponsors to plan for more than one phase 3 trial, particularly where there is a lack of pharmacological rationale; when studying a new pharmacological principle; or where phase 1 and 2 data are limited or unconvincing. On the other hand, he commented that a sponsor may decide to rely on the SCT model when developing a therapy with a clinically valuable treatment effect, with high quality data, and with consistent results among centers, population subgroups, and for different endpoints.

The CPMP expert working group notes that in most cases, a development program with several studies may be the only feasible way to provide the variety of data needed to confirm the usefulness of a product in the intended population(s). A submission with only one phase 3 study has to be "particularly compelling" with respect to internal and external validity, clinical relevance, statistical significance, data quality, and internal consistency. Although the assumed purpose of phase 3 is to confirm findings of earlier studies, in reality, many later trials are based on vague assumptions. European regulators appear to be open for discussion of this topic.

INDUSTRY PRACTICE

A survey of about 50 pharmaceutical and biotech companies two years ago by the Tufts Center for the Study of Drug Development found that two-thirds of the 36 respondents had used the confirmatory evidence-SCT approach or planned to do so. However, many of the approvals resulting from these development programs involved orphan drugs or effectiveness supplements, with a good number for supplemental pediatric indications (9).

In presenting an industry viewpoint, Ronald Krall (AstraZeneca) observed that large pharmaceutical company drug development programs are designed to deliver commercially successful products, not just approved products. The result is that most programs require multiple controlled clinical trials in order to demonstrate the advantages of a new drug over existing therapy. In this situation the confirmatory evidence-SCT paradigm for drug approval is largely moot. The confirmatory evidence-SCT approach is more attractive: when the market is underserved and it is unnecessary to do multiple controlled trials to develop a competitive profile; when a product is expected to have an obvious clinical advantage (eg, a new cancer agent with evidence of superior tumor response); for supplemental and pediatric indications; and when speed to market is paramount. As the regulators noted, the confirmatory evidence-SCT approach is desirable and has been industry practice for outcome trials that require many thousands of subjects, especially when mortality is the trial endpoint.

Krall cautioned that there are risks with the confirmatory evidence-SCT approach because it is dependent on the success of the single clinical trial, and even well designed trials sometimes have unanticipated variances. This may occur when important subgroups do not respond like the entire population. In his view, these occurances make the approach inherently more risky. He also cautioned about a "one-size-fits-all" definition of what constitutes confirmatory evidence.

To reduce the cost of drug development and to get to market faster, Krall observed that simplification of clinical trials—collecting fewer data, doing less source data verification, reducing data errors—is more likely to be productive, as is the application of

pharmacology-based modeling approaches to early definition of dose and dose regimen. Cost pressures on the industry make it imperative for companies to demonstrate effectiveness and clinical comparisons with as few trials as possible.

BIOTECH FIRMS FIND THE CONFIRMATORY EVIDENCE-SCT PARADIGM APPEALING

Smaller pharmaceutical and biotech companies, as well as medical device manufacturers, may be more eager to try new drug development models, according to several workshop participants. Large pharmaceutical companies have ample resources to conduct long and expensive studies, while smaller manufacturers with limited resources generally are looking for ways to make clinical research more affordable. One biotech company executive explained that some biotech researchers tend to "front-load" phase 1 and phase 2 studies with informative learning trials in order to optimize the chances for the success of one large phase 3 trial.

Craig Smith (Guilford Pharmaceuticals) observed that the confirmatory evidence-SCT proposal provides a healthy basis for challenging conventional wisdom and for considering new views for drug development. Industry finds, on balance, that two or more phase 3 studies provide a reliable basis for approving new drugs, he commented, adding that it is important to question how much evidence is enough.

One significant concern expressed was that if sponsors seek FDA advice on an innovative study plan, agency officials will give them a long list of issues to address, without sharing any of the risk in doing so. Industry executives recognize that FDA cannot say, "If you do this, we'll approve it." But sponsors would like some upfront indication from the regulators that following an agreed-upon research path will yield some regulatory benefit. Sponsors also reported that some FDA reviewers are unwilling to consider new research approaches. Temple advised industry to bring such difficulties to him or other CDER officials, but companies developing new drugs often fear that such action may alienate future reviewers.

BREAKOUT GROUP FINDINGS

Individual working groups at the Georgetown workshop discussed these issues and sought to develop recommendations for defining more clearly:

1. The nature, scope, and standards of evidence of effectiveness,
2. The qualities and standards for confirmatory evidence,
3. SCT requirements when supported by confirmatory evidence, and
4. Qualities of an adequate safety database.

Challenged with several issues to consider, breakout group facilitators guided their group's discussion. Following are the reports of each breakout group's deliberations and recommendations.

Breakout Group 1: Nature, Sources, and Standards for "Evidence of Effectiveness." *Facilitators: Janet Woodcock and Frank Sasinowski*

What is Causal Evidence of Effectiveness? Causal evidence, as a subset of confirmatory evidence, constitutes evidence of pharmacologic activity along the pathophysiologic chain that correlates to dose response and is associated with the clinical outcome. The consensus of the group was that such causal evidence might be sufficient to affirm effectiveness, depending upon the strength of its association with the clinical outcome and dose response. In addition, the group clarified that causal evidence must be derived from adequate and well-controlled clinical trials, while other forms of confirmatory evidence may not need to be.

Confirmatory Evidence Includes Prior Scientific Knowledge. Woodcock stated that confirmatory evidence is broader than causal evidence and may include "prior scientific

knowledge." To illustrate, she provided a hypothetical illustration in which a company was seeking to obtain FDA approval of an estrogen that was a new chemical entity. For the indication of "prevention of osteoporosis" the firm would need to demonstrate improvement in bone mineral density in a single clinical trial, as well as substantial evidence supporting the lowest effective dose (in order to minimize toxicity). The evidence on dosing could come from the bone mineral density trial, or from a phase 2 dose-ranging trial. Additional studies would not be required because of extensive scientific understanding and acceptance of estrogen's pharmacological effects on bone mineral density as a surrogate endpoint for osteoporotic fracture rate. In contrast, if an indication of "treatment of osteoporosis" is sought (ie, for women with existing osteoporotic fractures), sponsors would be required to demonstrate a beneficial effect on osteoporotic fractures in a single clinical trial, plus provide confirmatory evidence of an effect on bone mineral density, including dose response.

Cautions. The group also identified risks that may be associated with relying on causal evidence to confirm effectiveness. For example, pharmacologic activity, though initially demonstrated, may not ultimately correlate with clinical outcome. Moreover, studies to identify causal evidence may be costly and complex, and may raise unanticipated questions and issues. In sum, conducting two identical empirical studies (that is, replication) may carry less commercial risk in some cases, but is far less informative to FDA and may prevent companies from drafting fully informative product labeling.

Applicability to all Drugs. The group addressed whether causal evidence of pharmacologic activity, in combination with a single empirical trial, is sufficient to affirm the effectiveness of any drug product. Current FDA practice generally limits single study approvals to cases in which the single study has a very persuasive statistical finding, or where a study cannot be replicated for ethical reasons because the disease or condition studied is serious or life-threatening. The group concluded that the confirmatory evidence-SCT approval standard in FDAMA § 115(a) is broad enough that it should apply, in theory, to all drug products. In practice, however, application is likely to be incremental and evolutionary, by class of drug as well as by disease. For example, in studying the first drug in a class, it may be difficult to establish that pharmacologic activity correlates with clinical outcome. Early development programs, however, may provide information on appropriate biomarkers for subsequent drugs in the same class or to treat the same disease.

Clarification of FDA Policy. Finally, the group discussed an inconsistency in the May 1998 FDA effectiveness document as to whether the examples it contains apply to new drug products (including new molecular entities) or only to a new use of already approved drug products. Woodcock noted that in practice FDA applies the concepts of the effectiveness document to new molecular entities as well as new uses of approved drug products.

Breakout Group 2: Satisfactory Requirements for Confirmatory Evidence to Support a Single Clinical Trial for the Determination of Drug Effectiveness. ***Facilitators: Lewis Sheiner (University of California at San Francisco), Karen Weiss (CBER), and Larry Lesko (CDER)***

What is Confirmatory Evidence? Confirmatory evidence is defined as evidence other than a second phase 3 randomized clinical trial (RCT) that supports the generalization of results of the SCT to future patients. The breakout group noted that replication (ie, two RCTs) is not an absolute confirmation that current results will be similar in future patients. Thus, the use of confirmatory evidence for this purpose need only match, not exceed, the predictive value of a second RCT.

TABLE 1
Hierarchy of Causal Evidence

Dimension	Extrapolate From	Extrapolate To
Pharmacology	Drugs in class	Target drug
Pathophysiology	Related condition	Target condition
Causal chain biomarker/time	Biomarker/short term	Clinical endpoint/long term
Biology	Animal	Human

Goal and Incentives. A principle goal of the confirmatory evidence-SCT paradigm is to render drug development more informative and efficient through increased use of modern scientific methods, including principles and techniques of clinical pharmacology. One question is whether the possibility of gaining market access more quickly and more economically by following the confirmatory evidence-SCT paradigm provides a sufficiently strong incentive for change. Some argue that one additional effectiveness trial does not materially increase the total cost of clinical development. Thus the confirmatory evidence-SCT approach alone would not appear to economically motivate a "system" change that makes greater use of mechanistic science.

However, phase 3 RCTs, preceded by several phase 2 trials that currently do not "count" for effectiveness determination, frequently involve large numbers of patients and can be expensive and lengthy. This is true especially when clinical endpoints are delayed, the active treatment produces only small effects, and/or the indication is rare. In such cases, the availability of a confirmatory evidence-SCT approach for establishing effectiveness may be attractive to a sponsor aiming for efficient drug development. The confirmatory evidence could come from phase 2 trials of the following types:

1. Exposure (dose, pharmacokinetics)-response, using a rapidly-responding continuous biomarker, for example, blood pressure,
2. Results from an RCT that studies a closely related disease or drug, and/or
3. An appropriate response in an animal model of the disease in question.

Several specific examples of potential confirmatory evidence are provided in FDA's 1998 effectiveness guideline (5).

Extrapolation. The basic concept underlying the use of confirmatory evidence to establish effectiveness is extrapolation of desired effects from a biomarker to clinical effectiveness. There are several dimensions in which this extrapolation can take place, as shown in Table 1.

Although it may vary case-by-case, the credibility of the extrapolation depends upon the dimension in the following order: pharmacology > pathophysiology > causal chain biomarker > biology. Credibility depends upon the strength of the evidence. For pharmacology and pathophysiology, the key factors are the strength of clinical data in studies of other drugs in a class or related diseases that share a similar action or disease mechanism. For a causal-chain biomarker, the main criteria are the state of scientific knowledge of the disease mechanisms, consistency of association of the clinically approvable endpoint and biomarker, the proximity of the biomarker to the clinical endpoint on the causal path, multiple biomarkers changing in "correct" temporal sequence, and similarity of biomarker exposure and clinical exposure-response when both are studied.

Implementation of the confirmatory evidence-SCT paradigm proposed by Peck, Rubin, and Sheiner requires that the SCT show a strong association between the clinical endpoint(s) and the confirmatory evidence biomarker(s). At least one confirmatory evidence clinical study of the biomarker (with or without clinical endpoints) should show

similar exposure-response relationships in the SCT. This links evidence effectiveness in both the confirmatory evidence and SCT studies.

Encouragement, Consensus, and Evaluations. If the confirmatory evidence-SCT paradigm is to gain acceptance, it will be through an evolutionary process. For example, there is currently no consensus on some technical issues that would enable the paradigm, such as what degree of data-driven model building is acceptable for confirmatory evidence, and how one quantifies the evidence for establishing pharmacological action. As next steps, these and related issues need to be investigated. New methods should be proposed and evaluated using actual examples from the drug development process.

All parties are urged to explore what can be done to further encourage the above described evolution if the inducement of market application approval based upon the confirmatory evidence-SCT model is not sufficient to stimulate more scientific drug development. Some current ideas are to employ "enrichment designs" (dose-response assessed in responders, excluding nonadherers); within-individual dose-response; Bayesian trials incorporating adaptive dose-ranging, adaptive allocation, and/or "seamless" transition from phase 2 to phase 3; or other "learning while confirming" ideas. It is unclear how best to coordinate the efforts of academics, industry, and regulatory agencies in devising, testing, and utilizing such approaches. Perhaps this could be a topic for subsequent working group conferences.

Breakout Group 3. Satisfactory Requirements for a Single Clinical Trial in Conjunction with Adequate Confirmatory Evidence. ***Facilitators: Donald Rubin (Harvard) and Bob O'Neill (CDER), assisted by Karl Peace (University of Georgia) and Joachim Vollmar (Pharmaceutical Research Associates Int.).***

Applicability of Confirmatory Evidence. The group found that the confirmatory evidence concept still is rather vague and appears to be specific to the particular case situation. For this reason, the potential for confirmatory evidence to support the SCT approach should focus on those areas that may benefit most from the availability of significant prior information that meets the spirit of confirmatory evidence. The group viewed the confirmatory evidence-SCT concept to be immediately applicable to a subset of new drug development candidates, mainly new formulations, dosage regimen changes, line extensions, or new indications of already approved drugs for which there is substantial prior empirical data and information. The application of the confirmatory evidence concept to new molecular entities may require more knowledge of disease and pharmacology than has traditionally been available.

Benefits of Confirmatory Evidence. Some industry participants acknowledged that they often regard phase 2 studies as an obstacle to advancing to phase 3 quickly. However, recent industry experience suggests that the costs and risks of inadequate understanding of confirmatory evidence prior to undertaking large phase 3 trials are great. These later studies often fail or have disappointing results due to a lack of understanding of pharmacology and disease. The group recognized the negative consequences of inadequate planning and research in advance of proceeding with an SCT. The value added by earlier confirmatory evidence planning, data collection, and analysis is reduction of risks of late phase 3 trial failures.

In attempting to address the design of an SCT, concerns were raised about the kinds of confirmatory evidence that would be available in advance of planning such a study. At the same time, to take full advantage of confirmatory evidence-SCT thinking, it is worth considering how early in the drug development process such planning should occur. The end-of-phase 2 meeting is usually too late for formal discussion between the sponsor and FDA concerning use of the confirmatory evidence-SCT paradigm, leading to the desirability of an earlier sponsor-FDA

discussion, for example, in an end-of-phase 1 meeting.

Sponsors may need a better reward system for utilizing the confirmatory evidence-SCT approach to offset apparent potential risks involved in planning for an SCT to provide the primary evidence of effectiveness. One incentive may reside in the opportunity to reduce the scope of additional development programs for product extensions or expanded indications.

SCT Study Design. To implement the confirmatory evidence-SCT approach, trial designers need to address the types of data analyses, models, complexity of models and assumptions, and the challenge of achieving multiple trial objectives embedded in a single study. When confirmatory evidence is planned as part of the entire NDA database, it is probably necessary to consider new study designs for an SCT. This may involve trials with larger sample sizes to address several questions simultaneously and variability of observations, and more flexible and adaptive designs that incorporate sequential decision making utilizing the confirmatory evidence information.

Complex Models. The confirmatory evidence-SCT paradigm may involve the use of disease and drug action models, prospective identification of covariates, and prospective sequential model-building strategies. If models are relied upon for analysis, empirical validation of the models is important. Greater use of more complex models for analysis, based upon knowledge gained from the confirmatory evidence investigations, raises concern about whether this contributes to or detracts from trial success. For example, additional trial objectives may be added. There may be an impact due to missing data and informative censoring associated with patient withdrawals from a trial, complicating interpretation using model-dependent analyses.

To satisfy the need for more intensive data analyses than are routinely employed in clinical trials when fewer model assumptions are made, sequential staged decision making may be required. This may involve discussions with regulators at critical points in the development process. However, regulators may be concerned about the risks of using unvalidated approaches. Such sponsor-regulator dialogue also may be risky for industry if reliance on an uncertain future outcome delays market entry date. One approach may be to determine the number and type of clinical studies needed as well as their sequence according to agreed upon final labeling requirements. Deciding how many studies are needed requires consensus on the core evidentiary database.

Important lessons can be learned from the evaluation of meta-analyses, especially of studies that appear to be nonreplicable. One problem area has been the evaluation of heterogeneous treatment effects in several studies in which the design—not the drug—induces the observed effect. Often traditional meta-analytic-based statistical methods are not powerful enough to evaluate consistency of effects and to dissect the reasons for differences.

P-value is not the Central Issue. The real quantitative and statistical issues associated with implementation of an SCT in the confirmatory evidence-SCT paradigm are not about type 1 error alone or about the merits of the p value alone. Essentially, they concern broader issues about the demonstration and interpretation of the evidence. This concept includes evaluation of treatment effects in a variety of target populations and the relationships to treatment effect sizes, the precision of these estimates, their variation among subgroups, and variation according to the multiple circumstances of conditions of use.

Value of the Confirmatory Evidence-SCT Paradigm. Questions may arise as to whether the confirmatory evidence-SCT approach may reduce or increase current effectiveness standards, thus requiring a case-by-case assessment of value added. However, this group emphasized that it is important to encourage the overall philosophy of confirmatory evidence.

The current rush-to-market drug development paradigm encourages premature entry to phase 3 trials without the earlier planning needed to avoid failure. A common reason for disappointments in late phase trials is an overly optimistic estimate of treatment effect, associated with premature selection of suboptimal dose(s). Confirmatory evidence may enable more realistic forecasting of treatment effects that can be taken into account in phase 3 trial planning.

Breakout Group 4: Adequate Requirements of a Safety Database, Assuming that Effectiveness is Independently Affirmed. ***Facilitators: Roger Porter (Wyeth-Ayerst) and Peter Honig (CDER)***

The group examined how the confirmatory evidence-SCT model relates to the need to establish an adequate safety database for a new drug. It examined trial design, trial analysis, and reporting requirements in seeking to determine what constitutes an ideal safety database.

The workshop examined the goals, components, and critical issues in the design of clinical trial safety programs, plus critical issues for future discussion such as the relative importance of understanding the mechanism of action of toxicity.

The group concluded that the ideal clinical program safety database should enable:

1. Generalization of the drug's safety profile in the treated population(s),
2. Prediction of the drug's safety profile in subpopulations (age, gender, concomitant drugs, nutritionals, over-the-counter products, etc.),
3. Prediction of the drug's safety based upon the severity of disease,
4. Prediction of the drug's cumulative toxicity with regard to duration of treatment (acute and chronic) and dose level,
5. Derivation of dose-response relationships with regard to safety, and
6. Derivation of a benefit-to-safety ratio for the subpopulations.

The members of this workshop emphasized that each clinical trial safety program needs to be customized with regard to:

1. The type and quantity of clinical safety data derived from the clinical trial effectiveness program, specific clinical safety trials, and a clinical trial safety program focused on the effects of drug duration and drug dose,
2. The size of the patient population,
3. The incidence of the disease being treated, and
4. The severity of the disease being treated.

The availability of robust safety databases from well-designed (controlled, blinded, and randomized) clinical studies will permit drug development scientists and regulatory scientists to be in a position to make informed safety assessments during both the drug development and the drug review phases, in order to ensure that safe drugs are available to patients who need them.

The overarching theme expressed by the members of the workshop was that the best safety database in a clinical program will be derived from controlled clinical trials. Although supplementation by specific safety clinical trials often is needed, these supplemental clinical trials by themselves will not be adequate to provide the safety knowledge and information needed for approval of a new drug registration dossier. Controlled trials may be needed to rule out significant toxicities known to be associated with other related compounds or drugs in the same class.

Thus, the ideal database for registration is one that will permit determination of a benefit-to-safety ratio. It was recommended that analysis may best be derived when both the clinical benefit (or its surrogate) and safety are measured in the same clinical trial. The benefit to safety assessments will be particularly revealing when the data are derived from dose response trials.

Altogether, the ideal clinical program safety database will be derived from:

1. The clinical safety database from the effectiveness trials (phase 2 and 3, and any ongoing effectiveness trials at the time of registration review), and
2. Specific supplemental acute and chronic safety clinical trials that are enriched with the appropriate subpopulations to determine the relationship of age, gender, concomitant medications, severity of disease, higher than recommended dose levels, and longer drug exposure to determine safety with chronic treatment.

Once the patient population, dose, and dose regimen have been established from a well-planned and executed phase 2 clinical program, a clinical trial safety program can be conducted in parallel with the phase 3 clinical effectiveness program. The clinical safety trials can be:

1. Continuations, extensions, or part of the clinical effectiveness trials,
2. New medium-size trials with a longer duration of treatment and/or higher dose than is planned for the effectiveness clinical program, or
3. "Very simple trials" that are focused on capturing only significant adverse reaction data.

In order for the safety database to be meaningful, the workshop members recommended that the clinical safety program trials (including simple trials):

1. Have a definitive hypothesis,
2. Be controlled, randomized and blinded trials, and
3. Capture effectiveness measures so a realistic benefit to safety assessment can be made.

The workshop participants agreed that with regard to a clinical safety program, "one size does not fit all." There is a clear need to be adaptable and flexible in the safety and benefit/risk evaluation of new drugs. This conclusion arose from discussion of whether the level of risk varies with the population to be studied, with the severity of the disease, or with the vulnerability of the patient population and the incidence of the disease. These issues confound efforts to generalize regarding the design of clinical safety programs. For example, if there are known and significant drug interactions, a "naturalistic" study evaluating the effectiveness of labeling or other information dissemination programs might be tested.

Several case studies were developed during the workshop to serve as "straw men" for the quantification of clinical trial safety programs. They were designed with regard to level of risk as a function of either disease severity or subject population size:

Case Study 1: Large Patient Population: Class IV Congestive Heart Failure. Specific large safety clinical trial programs are usually not required since the effectiveness clinical trials are generally large and with an endpoint of mortality. One large effectiveness trial may include enough patients for a full safety evaluation.

Case Study 2: Medium-size Patient Population: Epilepsy. Phase 2/3 trials of an epilepsy treatment typically involve 300 to 600 patients per trial, with a total number of 2000 to 3000 subjects in an NDA clinical program. Additional specific supplemental acute and chronic safety clinical trials as outlined above would be recommended for a drug that displayed a modest increase in effectiveness over already marketed entities.

However, the requirements might differ in a situation where the new drug is a significant breakthrough in epilepsy effectiveness, as in the case where never-before-seen effectiveness is demonstrated in a single 600-subject, 8-month RCT, along with confirmatory evidence from two other smaller phase 2 trials. That may raise questions about whether data from a single, simple safety study would be

adequate to substitute for missing safety data, and if this type of clinical safety trial should be required at the time of NDA review as an ongoing study, or as a phase 4 commitment. Presumably, regulatory agencies would weigh the merits of earlier patient access to such breakthrough therapies over requirements for immediate additional safety data.

Case Study 3: Very Small Patient Population: Gaucher's Disease. If we assume that there is a good understanding of the mechanism of the disease and the therapy, one clinical effectiveness trial may be convincing. Even if the one trial is not completely convincing, it may be difficult to repeat such a clinical trial for ethical and practical reasons. In this case, a single clinical effectiveness trial might be all the safety data that one will be able to generate, unless there is follow up with long-term clinical trials or a patient registry is established.

WORKSHOP CONCLUSIONS AND RECOMMENDATIONS

These conclusions and recommendations were selected from many proposals.

General Conclusions and Recommendations

1. Employing confirmatory evidence to support an SCT as a legal and practical basis for drug approval is a novel approach that differs significantly from the traditional approval requirement for two controlled phase 3 trials or acceptance of a particularly strong SCT without confirmatory evidence. This promising new paradigm requires further development to become widely practiced,
2. Any shift to a confirmatory evidence-SCT paradigm will be evolutionary. Advances needed are technical (eg, greater causal knowledge of disease pathophysiology, biomarkers and pharmacology; paradigms for integration of confirmatory evidence with an SCT; novel trial designs and data analytic techniques; etc.), and cultural (willingness of industry to depart from traditional empirical multitrial phase 3 programs, and receptivity and encouragement by regulatory agencies),
3. Presently, the confirmatory evidence-SCT paradigm may be most useful for product extensions and additional indications of already approved drugs, and is not yet sufficiently developed for routine application to development of new molecular entities with novel mechanisms of action. However, the confirmatory evidence-SCT paradigm may be ideal for use in development of orphan, subpart E, or Fast Track products,
4. The confirmatory evidence-SCT paradigm is not widely embraced due to the fact that well-financed pharmaceutical companies employ multiple phase 3 trials additionally for marketing and safety database purposes. Some sponsors would like to see the confirmatory evidence-SCT approach count more, but find its use hindered by lack of sufficient agency guidance and fear of agency conservatism. Another concern is that competitors may use a demonstrated confirmatory evidence-SCT pathway. Sponsors should reject the view that phase 2 is a barrier to advancement to phase 3. Rushing through phase 2 and failing to develop confirmatory evidence and a good understanding of a drug candidate often lead to overly large phase 3 trials that sometimes fail,
5. Biotech and smaller companies may regard the confirmatory evidence-SCT model as consistent with mechanistic/causal approaches used in discovery programs and as an affordable way to reduce the risks of failure,
6. Some academic scientists view confirmatory evidence-SCT as a pathway for improving and strengthening the science of drug development and regulation. For example, the inclusion of causal effectiveness data in both confirmatory evidence and SCT trials (Peck, Sheiner, and Rubin proposal), reflects a theory that integrates both causal and empirical evidence of effectiveness as a basis for confirmatory

evidence-SCT linkage of causal-chain biomarkers with pharmacology and empirically demonstrated clinical effectiveness, and
7. Effectiveness and safety may be decoupled for purposes of optimizing resource utilization and informativeness, providing opportunities to strengthen the safety database using a large very simple safety trial or specific clinical pharmacology investigations.

CE and Causal Pharmacological Evidence

1. Confirmatory evidence may be found in randomized, blinded, dose-response studies with a causal chain biomarker or clinical endpoint; clinical trial results from a closely related disease; and responses in an animal model of disease,
2. Causal evidence of pharmacological activity may support effectiveness in combination with a single clinical trial, depending on the strength of its association with the clinical outcome and dose response, and
3. Overreliance on causal evidence carries risk if pharmacological activity fails to correlate with clinical outcome. Studies to identify causal evidence may be complex and expensive and raise unanticipated questions and issues. In some cases, consequently, it may be more efficient to conduct two empirical phase 3 studies.

Recommendations

1. Publication by FDA of its rationale and database utilized as the accepted evidentiary basis of effectiveness in each new drug's approval would provide clear guidance to industry for designing new drug development programs, especially for employment of the confirmatory evidence-SCT approach,
2. End-of-phase 1 meetings are encouraged. Both regulators and sponsors agree that end-of-phase 2 is usually too late to adequately consider the confirmatory evidence-SCT approach and that an end-of-phase 1 meeting between the sponsor and FDA would be a preferable time for planning the confirmatory evidence-SCT approach,
3. More complete and timely analysis of contemporary phase 2 trials is encouraged, as this may provide already available causal evidence of effectiveness for use in the confirmatory evidence-SCT paradigm,
4. Using the estrogen-osteoporosis prevention example, regulatory receptivity to similar circumstances is encouraged, where established disease and pharmacological knowledge bases enable qualification of a surrogate endpoint for use in the confirmatory evidence-SCT paradigm, and
5. FDA guidances usually reflect past experience and well-established techniques, positioning newer methodologies as a more risky approach for sponsors. A suggestion from academia is that FDA include a section on "Practices to be Encouraged" in new guidances; this could lead to practical advances for use with the confirmatory evidence-SCT approach.

EPILOGUE

It is apparent that the confirmatory evidence-SCT model has not yet "caught fire." This may be because regulatory agencies are skeptical of it, sponsors are reluctant to displease regulatory authorities, and some researchers are uncertain that this approach is preferable to multiple clinical trials in terms of enhancing safety and ensuring effectiveness.

In the discussion, Sheiner (University of California at San Francisco) stated that he is very mechanistically oriented, while Bob Temple (CDER) is very empirical in the approach to decisions on a drug's action. Sheiner quipped that "I'm the mad modeler, and Temple doesn't believe anything unless you show him." These are caricatures of our positions, he acknowledged, noting that the real challenge is to better use increased knowledge about drug action to improve the drug development process. Temple responded that mechanistic explanations certainly should be factored into clinical trials, but

that the confirmatory evidence-SCT model increases the risk that the sponsor could make a mistake.

Janet Woodcock (CDER) added that she stands between the two sides, supporting FDA efforts to rely on prior knowledge to inform the link between pharmacologic effect and outcome. She pointed out that "system" problems among industry, academia, and regulators can prevent adoption of newer practices, despite the feeling of most individual members of these groups that innovative approaches would ultimately be beneficial.

Even if most drug applications do not currently rely on one clinical trial, increased focus on the scientific and technical issues underlying the clinical effectiveness-SCT model may encourage efforts to invest more in phase 2 studies. This strategy can reduce the number and excessive size of later clinical trials and help sponsors decide earlier to cancel unpromising research efforts.

Acknowledgments—The following participants are gratefully acknowledged for their contributions to conference management, planning committee participation, plenary session presentations, breakout session facilitations, and (marked by an asterisk) drafting or editing of texts of breakout session reports or the overall report: Richard Burr (Representative, North Carolina); David Feigal (Center for Devices and Radiological Health); Charles Grudzinskas (Center for Drug Development Science [CDDS])*; Dana Hilt (Guilford Pharmaceuticals)*; Samah Habayeb (CDDS); Jennifer Hansen (Staff, House of Representatives); Peter Honig (Center for Drug Evaluation and Research [CDER])*; Kurt Karst (Hyman, Phelps & McNamara)*; Armin Koch (German Federal Institute for Drugs and Medical Products); Ronald Krall (Astra-Zeneca)*; Cynthia Kroger (CDDS); Sandra Kweder (CDER)*; Larry Lesko (CDER)*; Carol McCullough (Enterprise Consulting Conference Management)*; Florian Michel (CDDS)*; Christopher Milne (Tufts Center for Study of Drug Development); Anne Marie Murphy (Hyman, Phelps & McNamara)*; Julie Nelson (CDDS)*; Robert O'Neill (CDER)*; Karl Peace (Georgia Southern University); Roger Porter (Wyeth-Ayerst)*; Donald Rubin (Harvard University); Frank Sasinowski (Hyman, Phelps & McNamara)*; Lewis Sheiner (University of California at San Francisco)*; Craig Smith (Guilford Pharmaceuticals)*; Robert Temple (CDER); Joachim Vollmar (Pharmaceutical Research Associates); Karen Weiss (CBER)*; and Janet Woodcock (CDER)*.

REFERENCES

1. Center for Drug Development Sciences. http://CDDS.georgetown.edu.
2. Food and Drug Administration Modernization Act (FDAMA) (Pub. L. 105-115); 1977.
3. U.S. Food and Drug Administration. *Guideline for the Format and Content of the Clinical and Statistical Sections of an Application*. Rockville, MD: U.S. Food and Drug Administration, Center for Drug Evaluation and Research; 1988.
4. Statement Regarding the Demonstration of Effectiveness of Human Drug Products and Devices. *Federal Register*. 60;147, August 1, 1995:39,180.
5. U.S. Food and Drug Administration. *Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products*. Rockville, MD: U.S. Food and Drug Administration Center for Drug Evaluation and Research; 1998.
6. Senate Report 105-43; Food and Drug Administration Accountability Act of 1997, Title IV Improving Certainty and Clarity of Rules, pg. 30.
7. Peck CC, Rubin DB, Sheiner LB. Causal evidence of effectiveness to support a single clinical trial approval. *Clin Pharmacol Ther*. (in press).
8. Committee for Proprietary Medicinal Products, European Medicines Evaluation Agency. *Points to consider on application with 1. Meta-analyses; 2. One pivotal study (draft)*. London: Committee for Proprietary Medicinal Products, European Medicines Evaluation Agency; May 31, 2001.
9. Milne CP. The single controlled trial: industry survey indicates implementation is still a work in progress. *Drug Inf J*. 2002.

VOLUME 46
NUMBER 2
MARCH 2012

Drug Information Journal

OFFICIAL PUBLICATION OF DIA

dij.sagepub.com
ISSN: 0092-8615

Editor's Note: As this special feature discusses, a rare disease is defined as one that affects fewer than 200,000 patients in the United States. The tremendous importance of this topic lies in the fact that there are thousands of such diseases, tens of millions of patients suffer from them, the majority are of genetic origin and are life threatening, and far too few have available treatments. I would like to thank Mr Sasinowski for this contribution to the Journal.

Quantum of Effectiveness Evidence in FDA's Approval of Orphan Drugs

Cataloguing FDA's Flexibility in Regulating Therapies for Persons with Rare Disorders

by Frank J. Sasinowski, M.S., M.P.H., J.D.[1]
Chairman of the Board
National Organization for Rare Disorders

One of the key underlying issues facing the development of all drugs, and particularly orphan drugs, is what kind of evidence the Food and Drug Administration (FDA) requires for approval. The Federal Food, Drug, and Cosmetic [FD&C] Act provides that for FDA to grant approval for a new drug, there must be "substantial evidence" of effectiveness derived from "adequate and well-controlled investigations." This language, which dates from 1962, provides leeway for FDA medical reviewers to make judgments as to what constitutes "substantial evidence" of a drug's effectiveness, that is, of its benefit to patients.

The sole law that applies specifically to orphan drugs, the Orphan Drug Act of 1983, provided financial incentives for drug companies to develop orphan drugs, which is legally defined as products that treat diseases that affect 200,000 or fewer patients in the U.S. But the Orphan Drug Act, whose enactment was championed by the National Organization for Rare Disorders (NORD), did not amend or revise the statutory standards in the law for establishing that a new medicine is safe and effective for its proposed use. From a strict regulatory standpoint, the standard for orphan drugs is identical to the standard required for all other drugs, namely that "substantial evidence" demonstrates the effectiveness of the drug for its intended uses.

In the past decades FDA has moved in two broad formal ways to establish policies that provide greater flexibility for medical reviewers in assessing applications for new drugs. Neither of these efforts was designed specifically for orphan therapies. First, in response to the AIDS crisis and need for new cancer therapies, FDA established regulatory systems that formally recognized the need for flexibility in FDA's review of therapies for serious diseases for which there is an unmet medical need. Such systems found expression in FDA's promulgation in 1988 of the IND Subpart E regulation (21 C.F.R. Part 312) and in 1992 of the NDA Subpart H regulation (21 C.F.R. Part 314) (sometimes referred to as the "accelerated approval" regulation). Second, in its pursuit of good regulatory science, FDA announced a seminal guidance in May 1998 on "Providing Clinical Evidence of Effectiveness" in which FDA described nine different ways for a new therapy to get approved on the basis of a single adequate and well-controlled trial. With this guidance, FDA created new regulatory tools for addressing the needs of patients while meeting the legal obligations to ensure that all new therapies are both safe and effective for their intended uses.

FDA has for many decades acknowledged that there is a need for flexibility in applying its standard for approval. For example, one of FDA's regulations states that: "FDA will approve an application after it determines that the drug meets the statutory standards for safety and effectiveness... While the statutory standards apply to all drugs, the many kinds of drugs that are subject to the statutory standards and the wide range of uses for those drugs demand flexibility in applying the standards. Thus FDA is required to exercise its scientific judgment to determine the kind and quantity of data and information an applicant is required to provide for a particular drug to meet the statutory standards." 21 C.F.R. § 314.105(c).

FDA publicly has expressed sensitivity to applying this flexibility to new therapies for rare disorders. For example, in his testimony to the United States Senate on June 23, 2010, Dr. Jesse Goodman, FDA Chief Scientist and Deputy Commissioner for Science and Public Health, testifying on "FDA's Efforts on Rare and Neglected Diseases," said: "FDA is fully committed to applying the requisite flexibility in the development and review of products for rare diseases, while fulfilling its important responsibility to assure that the products are safe and effective for these highly vulnerable populations. There are numerous examples of drugs approved for treating rare diseases where FDA's flexibility and sensitivity to the obstacles of drug development for rare diseases has brought forth a successful treatment. Many of the 357 approved orphan drugs have been successfully tested on extremely limited numbers of patients, serving as a testament to FDA's commitment to these patients. This is possible when the best science is flexibly applied and when therapies are truly effective."

Dr. Goodman cited as successful examples the following:

"* Carbaglu (carglumic acid) for the treatment of NAGS deficiency, the rarest of the Urea Cycle Disorders (UCDs): This disease affects fewer than 10 patients in the U.S. at any given time and fewer than 50 patients worldwide. This drug was approved in March 2010 based on a case series derived from fewer than 20 patients and comparison to a historical control group.

* VPRIV (velaglucerase) for the treatment of Gaucher disease, a rare genetic disorder: This disease affects approximately 2,000 people in the U.S. and approximately 5,000 worldwide. This drug was approved in February 2010 based on a development program that included about 100 patients and a pivotal study of 25 patients.

[1] Frank J. Sasinowski is a director with the law firm of Hyman, Phelps & McNamara, P.C. and Chairman of the Board of Directors of the National Organization for Rare Disorders (NORD).

* Myozyme (alglucosidase alfa) for the treatment of infantile variant, a rapidly fatal form of Gaucher disease: The variant of this disease affects about 1,000 patients in the U.S. and about 3,000 patients worldwide. This drug was approved in April 2006 based on a clinical development program of fewer than 80 patients and a pivotal study that included 18 patients.

* Ceprotin (human plasma derived protein C concentrate) for the treatment of severe congenital Protein C deficiency: There are fewer than 20 known patients with this disorder in the United States. This biological drug product was approved in March 2007 based on a study of 18 patients using comparison to historical control data." [2]

PURPOSE OF THIS STUDY

NORD designed this study to examine closely how much flexibility FDA provides in reviewing orphan drugs – that is, to determine whether FDA requires that orphan drug applications provide the conventional or traditional level of proof of effectiveness that is ordinarily expected for most drugs for more prevalent diseases. This issue is especially critical because the patient population available for testing of orphan drugs is by definition more limited than for drugs for more prevalent diseases. The National Institutes of Health estimates that there are as many as 7,000 rare diseases, with some affecting only a handful of patients. The numbers of persons with such disorders can vary, as for example, cystic fibrosis which affects 35,000 Americans, or infant botulism, which affects, at most, only a few hundred infants per year. This study examines whether FDA exercises flexibility when reviewing applications for these diseases and, if so, illustrates the nature and scope of that flexibility.

This paper specifically examines the quantum of effectiveness evidence that provided the basis for FDA's approval of the 135 non-cancer orphan drug new chemical entities that were approved between the orphan drug law's creation in 1983 and June 30, 2010. The intent was to catalogue each of the 135 orphan drugs according to whether its approval had demonstrated any exercise of scientific judgment or flexibility by FDA in reaching its conclusion that the statutory requirement for demonstrating that drug's effectiveness had been met. The study aims to determine, based on an examination of the publicly-available information used to support approval, whether the amount of data presented would have satisfied the conventional requirements for proving the effectiveness of the drug.

The examination of 135 orphan drugs found that 90 approvals were based on some exercise of flexibility by FDA. That is, the study supports the FDA assertion that it exercises flexibility when reviewing applications for orphan drugs. This study also catalogues the types of situations in which the FDA has elected to exercise that flexibility.

METHODS

To identify the non-cancer orphan drugs approved as new chemical entities, NORD relied upon FDA's publicly-available documents for drugs approved by FDA from January 1983 to June 30, 2010.

For each approved drug, NORD sought to access the FDA approval letter, the labeling at the time of that approval (in order to exclude subsequent supplemental information that later added new clinical data), the decision memoranda of the FDA officials who approved the products, and the reviews of the medical and statistical officers. While such documents were retrievable in most cases, only subsets of these documents were recoverable for some drugs, especially for several of the earliest approved orphan therapies.

The evidence explaining the basis for each drug's approval was analyzed and classified, in the judgment of NORD, as whether or not it would have met the usual and customary conventional showings of effectiveness that would ordinarily be expected for any disorder, including a common or prevalent disorder. In addition to this classification, the category of 90 non-cancer orphan drugs whose approval was based on some exercise of FDA flexibility was further analyzed and subdivided into either those which were based on a formal, expressed FDA system for flexibility ("administrative flexibility") or were not based on any such formal FDA expression of flexibility ("case-by-case flexibility").

In summary, this paper classifies the 135 orphan drug approvals into one of three categories based on the analysis of the quantum of effectiveness evidence:

1. "conventional" or traditional quantum of evidence;
2. evidence consistent with some formal FDA system for exercising discretion or "administrative flexibility"; or
3. evidence that is consistent with a "case-by-case flexibility".

The first two of these classifications are described below and the third category is one by exclusion. All available source documents were gathered and analyzed for each FDA approval in order to classify each approved orphan therapy approval as "Conventional," "Administrative Flexibility" or "Case-by-Case Flexibility" (see Figure 1).

1. Conventional or Traditional Showing of Effectiveness

This category is for those drugs whose quantum of effectiveness evidence would satisfy the usual, conventional, traditional showing of effectiveness, which most often is colloquially and commonly referred to as "the two adequate and well-controlled studies" standard.

The 1962 Amendments to the FD&C Act added the requirement that for FDA to approve for commercial marketing any drug, it had to conclude that there exists "substantial evidence...consisting of adequate and well-controlled investigations, including clinical investigations" such that "experts

[2] FDA Deputy Commissioner Dr. Jesse Goodman, Testimony before U.S. Senate Appropriations Committee Agriculture Subcommittee, at p.2 (June 23, 2010).

qualified by scientific training and experience to evaluate the effectiveness of the drug involved" could "fairly and responsibly" conclude that the drug will have the effects that the drug purports or claims to have in the sponsor's proposed labeling for that therapy. FD&C Act § 505(d). FDA has interpreted "adequate and well-controlled studies" to mean generally a minimum of two such studies. FDA has promulgated regulations defining the types of trial designs that are "adequate and well-controlled studies." 21 C.F.R. § 314.126. Traditionally, FDA has accepted two adequate and well controlled trials when each meets its primary endpoint by its prespecified primary analysis with a p value of less than 0.05.

2. Administrative Flexibility in Formal Expressed Systems

There are three major expressions of formal ways in which FDA exercises scientific discretion in assessing the effectiveness evidence of all drugs, not just for orphan therapies: 1. FDA Guidance for Industry: "Providing Clinical Evidence of Effectiveness" (May 1998) ("Evidence Guidance"); 2. FDAMA 115 "one adequate and well-controlled clinical investigation and confirmatory evidence"; and 3. Subpart H, 21 C.F.R. Part 314, or "accelerated approval" regulations ("Subpart H").

A. Evidence Guidance and FDAMA 115

In its May 1998 Evidence Guidance, FDA describes nine[3] circumstances in which a single trial may meet the statutorily-required effectiveness evidence. Generally FDA had set a standard of requiring at least two adequate and well-controlled studies, following the language of the 1962 Amendments which used the plural "investigations" to describe the basic requirement for effectiveness. There had been times prior to 1998 when FDA had approved drugs based on a single study, especially when the AIDS crisis was just starting, but for most diseases the agency held drug approval to the "at least two" studies standard. The 1998 Evidence Guidance described circumstances in which a single study might be sufficient, such as where it may be unethical to conduct a second study and where the single study has a "statistically very persuasive finding" with other indicia of reliability, such as a multi-center trial with no single center dominating the results.

At the same time that FDA was developing its May 1998 guidance, Congress was enacting an amendment to the 1962 effectiveness standard that created a new alternative statutory standard for establishing a drug's effectiveness. This new alternative statutory standard is: "one adequate and well-controlled study and confirmatory evidence." This provision of the law is referred to as FDAMA 115 (after the section in the law called the FDA Modernization Act or FDAMA that established this alternate statutory standard for substantial evidence of effectiveness). The May 1998 Evidence Guidance and FDAMA 115 can be seen as qualitatively similar, in that both spoke to new ways of establishing substantial evidence of effectiveness, and both were issued almost simultaneously.

B. Subpart H and Fast Track

The same 1997 law that created FDAMA 115 also created the statutory authority for "Fast Track" drugs, which is a modest elaboration by Congress of an FDA regulation known by its section of the drug regulations, Subpart H of 21 C.F.R. Part 314, or the so-called "accelerated approval" regulations (for biologics, the parallel regulation is at 21 C.F.R. Part 601, Subpart E). Both Fast Track and Subpart H are programs whereby a therapy for a serious or life-threatening disease for which there is no FDA-approved "available therapy" may be approved based either on an unvalidated surrogate that is reasonably likely to predict ultimate clinical outcome, or on an outcome other than irreversible morbidity or mortality. However, in such cases, there is also an additional post-approval requirement to conduct a study to establish the ultimate clinical outcome benefit, and if that study fails to do so, FDA may withdraw its approval as an expedited basis.

Subpart H represents a formal FDA system established to introduce an element of flexibility in executing FDA's responsibilities for ensuring that investigational therapies have adequately demonstrated their treatment benefit prior to marketing authorization. FDA created this system in response to the need of patients contracting HIV infections in the 1980's and the attendant public health crisis. This paper notes which orphan drugs were approved under Subpart H as well as which ones were designated as Fast Track therapies by way of a footnote in Figure 1.

RESULTS

Figure 1 records the classification for each of the 135 non-cancer orphan therapies approved as new chemical entities from the enactment of the Orphan Drug Act in 1983 through June 30, 2010, with 45 classified as "Conventional", 32 as "Administrative Flexibility", and 58 as "Case-by-Case Flexibility". In Appendix 1, there is a narrative text that briefly describes the basis for each "case-by-case flexibility" classification, except for two therapies in that category: #75 Lanreotide Acetate and #116 Sodium Phenylbutyrate. In addition, there are textual comments about particular aspects of interest regarding eight other therapies: #6 Ambrisentan, #7 Amifostin, #8 Anagrelide, #35 Coagulation Factor IX, #58 Fosphenytoin, #59 Gallium, #60 Gangciclovir, and #79 Levomethadyl Acetate. Textural comments are included for these eight even though they are not classified as "case-by-case flexibility" in order to provide breadth of perspective and depth of understanding to the analytical processes employed. All of the therapies are listed alphabetically by chemical names.

3 See Appendix 2 for more detailed discussion of how these nine types of single study approval examples apply to orphan drugs and of how FDAMA 115 relates to these.

DISCUSSION

When asked how much evidence of safety and effectiveness an orphan drug must provide, FDA officials have generally explained to the Agency's public Advisory Committees, patient organizations, pharmaceutical companies and Wall Street that the Orphan Drug Act did not change the statutory requirements for establishing the safety and effectiveness of a proposed new medicine. For example, in a March 2010 FDA briefing document for the FDA Advisory Committee on an orphan drug, pirfenidone, being considered for patients with a rare, fatal pulmonary condition called idiopathic pulmonary fibrosis (IPF), FDA said:

> In accord with our regulations, the Agency requires substantial evidence of effectiveness. Substantial evidence consists of adequate and well-controlled investigations on the basis of which it could be concluded that the drug will have the effect it is purported or labeled to have. The Agency usually requires more than one trial to provide independent substantiation of efficacy. Although IPF is an orphan disease, the requirements to establish effectiveness are not different, with the exception that the overall database may be smaller. We ask that you consider whether the results of PIPF-004 and PIPF-006 provide substantial evidence of efficacy to support the proposed indication to reduce decline in lung function in patients with IPF.[4]

This statement represents the most common FDA response when FDA is asked about the quantity and quality of effectiveness evidence required of an orphan drug. However, on some occasions, FDA has noted that it has the ability to be flexible within those statutory limits. For instance, the FDA briefing document for an Advisory Committee meeting on January 13, 2010 concerning the orphan drug Carbaglu (N-carbamylglutamate) for hyperammonemia had the following statement:

> FDA has been flexible within the limits imposed by the congressional scheme, broadly interpreting the statutory requirements to the extent possible where the data on a particular drug were convincing. The Code of Federal Regulations (21 C.F.R. § 314.126) allows for studies without concurrent controls to be used to provide substantial evidence of effectiveness in diseases with high and predictable mortality, or in studies in which the effect of the drug is self-evident. Thus, the evidence obtained from retrospectively reviewed case studies could be considered as substantial evidence of effectiveness under those particular circumstances. The fact that the case series presented in this application is retrospective, un-blinded, and uncontrolled, precludes any meaningful formal statistical analyses of the data. Under these conditions, any statistical inference from confidence intervals and/or p-values is uninterpretable and, consequently, should not be utilized to inform clinical decision-making. To help frame the Committee's deliberations on whether the evidence standard in this application has been met, an FDA guidance document, '[Evidence Guidance]' is provided as background on the regulatory requirements for evidence of effectiveness.[5]

Thus, while the norm has been for FDA to respond simply that the statutory standard for effectiveness was not amended by the 1983 Orphan Drug Act, there have been ample occasions on which FDA has observed that it also has the legal authority and scientific right to be flexible in applying those statutory standards to orphan drug therapies.

There have been documents in which FDA has made abundantly clear its commitment to flexibility in applying the standard of safety and effectiveness, most notably during the AIDS crisis. In the mid-1980's, FDA promulgated Subpart E of the IND regulations for "drugs intended to treat life-threatening and severely-debilitating illnesses." FDA stated:

> [The purpose of Subpart E is] to establish procedures designed to expedite the development, evaluation, and marketing of new therapies intended to treat persons with life-threatening and severely-debilitating illnesses, especially where no satisfactory alternative therapy exists. As stated [in section] 314.105(c) of this chapter, while the statutory standards of safety and effectiveness apply to all drugs, the many kinds of drugs that are subject to them, and the wide range of uses for those drugs, demand flexibility in applying the standards. The FDA has determined that it is appropriate to exercise *the broadest flexibility* in applying the statutory standards, while preserving appropriate guarantees for safety and effectiveness. *These procedures reflect the recognition that physicians and patients are generally willing to accept greater risks or side effects from products that treat life-threatening and severely-debilitating illnesses, than they would accept from products that treat less serious illnesses. These procedures also reflect the recognition that the benefits of the drug need to be evaluated in light of the severity of the disease being treated.*[6]

The regulation that FDA references in its Subpart E regulation is section 21 C.F.R. § 314.105(c) which predates the Subpart E regulation and illustrates again FDA's historic position on applying the same statutory standards in a flexible way depending upon the circumstances. According to 21 C.F.R. § 314.105(c):

> FDA will approve an application after it determines that the drug meets the statutory standards for safety and effectiveness, manufacturing and controls, and labeling. While the statutory standards apply to all drugs, the many kinds of drugs that are subject to them and the wide range of uses for those drugs demand flexibility in applying the standards. Thus FDA is required to exercise its scientific judgment to determine the kind and quantity of data and information an applicant is required to provide for a particular drug to meet them. FDA makes its views on drugs products and classes of drugs available though guidelines, recommendations and statements of policy.

An example of a formal regulatory policy or guidance that expresses this concept of "flexibility" in FDA's application of the statutory standards of safety and efficacy is seen in the International Conference on Harmonisation (ICH) of the Technical Requirements for Registration of Pharmaceuticals for Human Use E1A guidance. This FDA-adopted international guidance stipulates the minimum quantum of safety exposures necessary for FDA to even accept a marketing application for review when the medicine is intended for a chronic

4 FDA Pulmonary-Allergy Drugs Advisory Committee Division Memorandum, Feb. 12, 2010, at pp. 15-16.

5 FDA Briefing Document, at pp. 9-10, attached to FDA Dr. Donna Griebel's December 16, 2009 memo to the Advisory Committee. See also June 23, 2010 statement of Deputy Commissioner Goodman to the U.S. Senate hearing cited in opening paragraphs of this paper.

6 21 C.F.R. Part 312, Subpart E, emphasis added.

condition.[7] Most rare disorders are chronic in nature and not acute, and so this guidance applies to most rare disorder therapies. The guidance states that the minimum number of safety exposures to meet the statutory standard for safety are 1,500 persons exposed to the investigational therapy with 300 to 600 of those exposed for at least 6 months and with at least 100 exposed for one year. However, the guidance states that these minimum safety thresholds do not apply to therapies for rare disorders. Importantly, the guidance then does NOT state what is required in the alternative, whereas it could have stated an algorithm such as at least 1% of the U.S. population with the rare disease must be exposed with half of them for at least one year. Rather, the guidance relies upon the exercise of FDA's scientific judgment to determine what is appropriate to meet the statutory standard for safety in each particular rare disorder therapy.

In other areas FDA can exercise similar flexibility. For instance, where the potential number of subjects is limited, the degree to which FDA demands dose selection to be optimized in pre-approval studies may be reduced, as can FDA's requirements for validation of a patient-reported outcome instrument in a rare disorder population or proof of the sensitivity, specificity and clinical meaningfulness of a primary endpoint.

NORD has requested that FDA issue a formal policy statement on FDA's regulation of therapies for persons with rare diseases (see footnote 9). Given that each investigational therapy for a rare disorder will present unique features, NORD understands that the granularity of the requested statement of policy may necessarily be limited. However, even cataloging the nature and scope of the orphan drug precedents that illustrate FDA's flexibility may enable key stakeholders to better understand FDA's position. That is, even while FDA states correctly that the statutory standards are the same for prevalent and rare conditions, FDA should develop and issue a formal companion statement of the equally important and consistent FDA historic position that FDA will exercise its scientific judgment to interpret and apply those statutory standards in a flexible manner, tailored to the circumstances of each investigational therapy for each rare disease and disorder.

It is this cataloguing of orphan drug precedents that is the chief purpose of this analysis and paper. This review of FDA actions on rare disorder therapy marketing applications concludes that two of every three orphan drugs approved manifests FDA's historic flexibility in applying to therapies for rare disorders the statutory standard for establishing effectiveness. By this classification, 32 of the 135 orphan drugs analyzed reflect administrative flexibility, that is, FDA application of statutes and FDA regulations and guidance documents to those particular orphan therapies, and another 58 orphan therapies were approved on a case-by-case application of flexibility.

There is an element of subjectivity and judgment in making these classifications. NORD does not have access to non-public information, which both FDA and the sponsors have. It is therefore possible that FDA and drug manufacturers will disagree about into which one of these three categories any therapy may be classified.[8]

However, NORD believes that the overall thrust of the findings of this analysis is immovable – that FDA's approval actions on a considerable portion of therapies for those patients afflicted with rare disorders demonstrate a consistently applied flexibility in assessing the effectiveness of such therapies.

Ironically and unfortunately, there has not been any statement from FDA as to how that flexibility finds expression. At the first FDA public hearing on orphan drugs which was held on June 29 & 30, 2010, NORD called on FDA to issue a "clear, granular expression of FDA's historic commitment to exercise flexibility in its review of therapies for rare disorders."[9]

CONCLUSION

Research resources in the universe of rare disorders are precious, with the most precious being the persons with the rare disorders who heroically volunteer to participate in a trial, usually under conditions where there is less known than in trials of therapies for prevalent diseases about the safety and potential effectiveness of the investigational therapy from animal models, animal toxicology and early human trials. So, when these trials are conducted, sometimes with designs with which all parties may not be in full concurrence, including FDA, great deference should be afforded to the design of these trials and flexibility applied in the interpretation of their results. If such a principle were to be addressed and accepted by FDA, much good would come of it.

In the more than 28 years since its enactment, the Orphan Drug Act has proven a resounding success. This is best seen in the 357 new medicines for more than 200 different rare disorders approved by FDA over the first quarter of a century of the law's existence.

NORD believes that this study's confirmation of FDA's flexibility in reviewing applications for orphan drugs reinforces the need for its public acknowledgement that orphan drugs are indeed meritorious of special consideration. Such a statement by the FDA would provide the impetus for greater attention to orphan drug therapies within the academic community as well as within the drug development and investment communities.

With health care reform measures inevitably changing how medicine is practiced and how patients are treated and reimbursed, the need for such attention to the rare disease community is especially critical. Patients with rare diseases can easily be left behind during this transitional period. FDA has demon-

7 Note that this paper consists of an analysis only of the quantum of evidence of effectiveness information determined to be adequate by FDA to support an approval, and the FDA-adopted ICH guidance that is the subject of this paragraph refers to a formal expression by FDA of its flexibility with respect to the quantum of safety information required for orphan drug therapies.

8 A further cautionary note is that every drug approval, whether for a rare condition or a common one, stands on a unique set of empiric evidence judged against a backdrop of specific scientific and clinical considerations in light of the relative degree of the medical needs of that particular set of patients. Therefore, caution must be exercised in any attempt to extrapolate from any one or more of these case studies to current or future therapies in development or under FDA review.

9 Statement of NORD, presented by Chairman Frank Sasinowski (June 29, 2010).

strated in its review of orphan products that it recognizes the importance of therapies for persons with rare disorders. It is time for that policy to be clearly enunciated as a formal FDA policy, and for FDA medical reviewers to incorporate and recognize this flexibility in a systematic way into their evaluations of each new therapy in development and under FDA review for Americans with any rare disease. Much that is very good for all persons with rare disorders could come of this.

NORD exhorts FDA to continue to embrace even more fully the historic flexibility it has long noted and exercised in FDA's regulation of medicines for those Americans with rare disorders. [10]

10 Author's note: The author commends FDA on its stellar, worldwide leadership on critical matters affecting persons with rare diseases for the past 28 years, and exhorts FDA to continue to embrace even more fully the historic flexibility FDA has long noted and exercised in FDA's regulation of medicines for those Americans with rare diseases. In the over 28 years since its enactment, the Orphan Drug Act has proven a resounding success. This is best seen in the over 350 new medicines for more than 200 different rare disorders approved by FDA. However, there are still about 6,800 disorders for which there is not one FDA-approved therapy. Perhaps most discouraging is that many affected with rare disorders do not even see any research being conducted on their conditions. It seems as though the proverbial low-hanging fruit has been harvested in the first quarter of a century of the law's existence, while the vast majority of persons with rare diseases see only that there is no medicine within their reach, and sometimes even within the reach of reasonable hope. In sum, much has been accomplished by FDA, by NIH, by medical and scientific researchers, by the pharmaceutical industry, by the financial community and by patient advocates in these first 28 years, but much more beckons each of us to respond to the needs of those with rare diseases. The author's heartfelt hope is that this analysis helps to advance the development of those medicines to aid all in need of them.

FIGURE 1

	Chemical and Brand Names	Approval (mm/yy)	Type of Efficacy Evidence: Conventional	Administrative Flexibility	Case-by-case Flexibility
1	Agalsidase beta - Fabrazyme[1]	04/2003		X	
2	Albendazole - Albenza	06/1996			X
3	Alglucerase - Ceredase	04/1991			X
4	Alitretinoin - Panretin	02/1999			X
5	Alpha1-Proteinase Inhibitor (Human) - Prolastin	12/1987			X
6	Ambrisentan - Letairis[2]	06/2007	X		
7	Amifostine - Ethyol[1]	12/1995		X	
8	Anagrelide HCl - Agrylin	03/1997	X		
9	Antihemophilic Factor Recombinant - Kogenate	02/1993		X	
10	Antithrombin III (Human) - Thrombate III	12/1991		X	
11	Antithrombin III (Human) - Atnativ	12/1989		X	
12	Antivenin, Crotalidae Polyvalent Immune Fab (Ovine) - CroFab	10/2000	X		
13	Apomorphine HCl - Apokyn[3]	04/2004	X		
14	Aprotinin - Trasylol	12/1993	X		
15	Artemether/Lumefantrine - Coartem[3]	04/2009			X
16	Atovaquone - Mepron	11/1992	X		
17	Azacitidine - Vidaza[3]	05/2004		X	
18	Basiliximab - Simulect	05/1998	X		
19	Sodium Benzoate and Sodium Phenylacetate - Ucephan	12/1987			X
20	Beractant - Survanta	07/1991	X		
21	Betaine HCl - Cystadane	10/1996			X
22	Bosentan - Tracleer[2]	11/2001	X		
23	Botulinum Toxin Type A - Botox, Botox Cosmetic	12/1991	X		
24	Botulinum Toxin Type B - Myobloc	12/2000	X		
25	Botulism Immune Globulin Intravenous (Human) - BabyBIG	10/2003		X	
26	C1 Esterase Inhibitor (Human) - Cinryze	10/2008		X	
27	C1 Esterase Inhibitor (Human) - Berinert P	10/2009		X	
28	Calcitonin-Human for Injection - Cibacalcin	10/1986	X		
29	Canakinumab - Ilaris	06/2009		X	
30	Capsaicin - Qutenza	11/2009	X		
31	Carglumic acid - Carbaglu	03/2010		X	
32	Chenodiol - Chenix	07/1983			X
33	Cinacalcet HCl - Sensipar	03/2004			X
34	Clofazimine - Lamprene	12/1986	X		
35	Coagulation Factor IX (Recombinant) - Benefix	02/1997		X	
36	Coagulation Factor IX - Mononine	08/1992		X	
37	Coagulation Factor IX (Human) - Alphanine	12/1990		X	
38	Coagulation Factor VIIa (Recombinant) - NovoSeven	03/1999			X
39	Colfosceril Palmitate, Cetyl Alcohol, Tyloxapol - Exosurf Neonatal for Intrathecal Suspension	08/1990	X		
40	Collagenase clostridium histolyticum) - Xiaflex	02/2010	X		
41	Corticorelin Ovine Triflutate - Acthrel	05/1996			X
42	Cysteamine Bitartrate - Cystagon	08/1994			X
43	Cytomegalovirus Immune Globulin (Human) - Cytogam	12/1998	X		
44	Daclizumab - Zenapax	12/1997	X		
45	Dalfampridine - Ampyra	01/2010	X		
46	Deferasirox - Exjade[1,3]	11/2005			X
47	Dexrazoxane HCl - Zinecard[1]	05/1995			X
48	Diethylenetriamine pentaacetic acid (Dtpa) - none	08/2004			X
49	Digoxin Immune Fab (Ovine) - Digibind	04/1986	X		
50	Dornase Alfa - Pulmozyme	12/1993		X	
51	Ecallantide - Kalbitor	11/2009	X		
52	Eculizumab - Soliris	03/2010	X		
53	Eflornithine HCl - Ornidyl	11/1990	X		

1-Subpart H for Efficacy; 2-Subpart H for Safety; 3-Fast Track

	Chemical and Brand Names	Approval (mm/yy)	Type of Efficacy Evidence: Conventional	Administrative Flexibility	Case-by-case Flexibility
54	Eltrombopag Olamine - Promacta	11/2008	X		
55	Ethanolamine Oleate - Ethamolin	12/1988			X
56	Felbamate - Felbatol	07/1993	X		
57	Fomepizole - Antizol	12/1997			X
58	Fosphenytoin Sodium - Cerebyx	08/1996		X	
59	Gallium Nitrate - Ganite	01/1991		X	
60	Ganciclovir Sodium - Cytovene	06/1989		X	
61	Glatiramer Acetate - Copaxone	12/1996		X	
62	Glutamine - Nutrestore	06/2004		X	
63	Halofantrine HCl - Halfan	07/1992	X		
64	Hemin - Panhematin	07/1983			X
65	Histrelin Acetate - Supprelin	12/1991	X		
66	Human Fibrinogen Concentrate, Pasteurized - RiaSTAP	01/2009		X	
67	Icodextrin Peritoneal Dialysis Solution - Extraneal	12/2002		X	
68	Idursulfase - Elaprase	07/2006		X	
69	Iloprost - Ventavis	12/2004		X	
70	Imiglucerase - Cerezyme	05/1994			X
71	Interferon Beta-1a - Avonex	05/1996			X
72	Interferon Beta-1b - Betaseron[1]	07/1993		X	
73	Interferon Gamma 1B - Actimmune	12/1990		X	
74	Ferric Hexacyanoferrate(II) - Radiogardase	10/2003			X
75	Lanreotide Acetate - Somatuline Depot	08/2007			X
76	Laronidase - Aldurazyme	04/2003			X
77	Lenalidomide - Revlimid[2,3]	06/2010			X
78	Lepirudin - Refludan	03/1998			X
79	Levomethadyl Acetate HCl - Orlaam	07/1993		X	
80	Lodoxamide Tromethamine - Alomide	09/1993	X		
81	Mecasermin Rinfabate Recombinant - Iplex	12/2005			X
82	Mecasermin Recombinant - Increlex	08/2005			X
83	Mesna - Mesnex	12/1988		X	
84	Midodrine HCl - Proamatine (Amatine)[1]	09/1996			X
85	Miglustat - Zavesca[3]	07/2003			X
86	Modafinil - Provigil	12/1998	X		
87	Monoctanoin - Moctanin	10/1985			X
88	N-acetylgalactosamine 4-sulphatase, Recombinant Human [Galsulfase] - Naglazyme	05/2005			X
89	Nitazoxanide - Alinia	11/2002	X		
90	Nitisinone - Orfadin[3]	01/2002		X	
91	Nitric Oxide - Inomax	12/1999	X		
92	Oprelvekin - Neumega	11/1997			X
93	Pegademase Bovine - Adagen	03/1990			X
94	Pegvisomant - Somavert[3]	03/2003		X	
95	Pentamidine Isethionate - Pentam	10/1984	X		
96	Pentastarch - Pentaspan	05/1987	X		
97	Pentosan Polysulfate Sodium - Elmiron	09/1996	X		
98	Plerixafor - Mozobil	12/2008	X		
99	Protein C Concentrate - Ceprotin	03/2007			X
100	Rasburicase - Elitek	07/2002			X
101	Recombinant Human Acid Alpha-Glucosidase [Alglucosidase ALFA] - Myozyme	04/2006			X
102	Recombinant Human Antithrombin - ATryn	02/2009			X
103	Respiratory Syncytial Virus Immune Globulin (Human) - Respigam	01/1996			X
104	Rho (D) Immune Globulin Intravenous (Human) - Winrho SD	03/1995			X

1-Subpart H for Efficacy; 2-Subpart H for Safety; 3-Fast Track

	Chemical and Brand Names	Approval (mm/yy)	Type of Efficacy Evidence: Conventional	Administrative Flexibility	Case-by-case Flexibility
105	Rifabutin - Mycobutin	12/1992	X		
106	Rifapentine - Priftin[1]	06/1998			X
107	Rilonacept - Arcalyst	02/2008			X
108	Riluzole - Rilutek	12/1995			X
109	Romiplostim - Nplate	08/2008	X		
110	Rufinamide - Banzel	11/2008			X
111	Sacrosidase - Sucraid	04/1998			X
112	Sapropterin Dihydrochloride - Kuvan	12/2007			X
113	Sargramostim - Leukine	03/1991	X		
114	Selegiline HCl - Eldepryl	06/1989	X		
115	Sodium oxybate - Xyrem[2]	07/2002	X		
116	Sodium Phenylbutyrate - Buphenyl	04/1996			X
117	Somatrem - Protropin	10/1985	X		
118	Sotalol HCl - Betapace	10/1992	X		
119	Sterile Talc Powder - Sclerosol	12/1997			X
120	Succimer - Chemet	01/1991	X		
121	Teriparatide Acetate - Parathar	12/1987		X	
122	Tetrabenazine - Xenazine	08/2008			X
123	Thalidomide - Thalomid[2]	07/1998			X
124	Tiopronin - Thiola	08/1988			X
125	Tranexamic Acid - Cyklokapron	12/1986			X
126	Treprostinil sodium - Remodulin[1]	05/2002			X
127	Trientine HCl - Syprine	11/1985			X
128	Trimetrexate Glucuronate - Neutrexin	12/1993			X
129	Vaccinia Immune Globulin (Human) Intravenous - N/A	02/2005			X
130	Velaglucerase alfa - Vpriv	02/2010	X		
131	Vigabatrin - Sabril	08/2009			X
132	von Willebrand Factor/Coagulation Factor VIII Complex (Human) - Wilate	12/2009			X
133	Zalcitabine - Hivid[3]	06/1992			X
134	Zidovudine - Retrovir	03/1987		X	
135	Zoledronic Acid - Zometa	08/2001	X		
	Sub Totals:		**45**	**32**	**58**
	Flexibility:		**Not Needed: 45**	**Yes: 90**	

1-Subpart H for Efficacy; 2-Subpart H for Safety; 3-Fast Track

APPENDIX 1

This Appendix provides commentary on the basis for approval only for products categorized as "case-by-case flexibility." The Appendix is keyed to the product numbering system in Figure 1 (thus, the Appendix starts with the second drug listed because there is no commentary on the first drug listed).

2. Albendazole - Albenza

The 1996 approval for this antihelminthic drug for treating infectious diseases caused by pork tapeworms and by dog tapeworms was based by FDA on:

1. a single study that was either a well-controlled study (Medical Review) or not well-controlled (Statistical Review, 28);
2. supporting literature (which was not all positive, one October 1995 paper in the Annals of Internal Medicine concluded from this study of 138 subjects over 2 years that "previous reports of favorable response to treatment of necrilocysticercosis (pork tapeworm) with...albendazole are by no means definitive and may be a reflection of the natural history of the condition");
3. compassionate use information; and
4. existing approvals in Australia, The Netherlands, Germany, United Kingdom, South Africa, India, Japan, and Spain.

The one study characterized by the medical reviewer as "well-controlled" was the Gelman study in Peru which compared 55 subjects in those with pork tapeworm disease with approximately half of the subjects randomized either to 7 days or 14 days of Albendazole. After 90 days there was no difference between the two groups ($p = 1.00$, Medical Review, 42[11]), but at one year the group on 7 days of therapy had a statistically significant greater reduction in cysts (primary endpoint) than did the group on 14 days of therapy ($p = 0.037$, Medical Review, 42 .) (The statistical review had concluded that "there were no formal statistical analysis of the clinical data in the NDA, [and] only descriptive statistics were used." Statistical Review, 28). The medical review cited multiple deficiencies in the Gelman study when it was audited (Medical Review, 98). Overall, the medical review catalogued the litany of clinical factors which hindered this regulatory review: (1) there is little or no such disease in the U.S.; (2) The natural history of the disease is not completely understood; (3) There is a lack of gold standard for diagnosis; (4) There is a lack of reliable clinical endpoints; and (5) need for long term follow-up.[12]

The statistical review concluded, with respect to the therapy for dog tapeworm, that "due to the very limited data available...the statistical conclusion toward the efficacy...of [albendazole] can not be reached" (Statistical Review, 30) and with respect to the therapy for pork tapeworm, "the results do not sufficiently provide comprehensive evidence to confirm [albendazole] as an effective...medicine...due to the weakness of the nature of these studies. Upon considering the particularity of [albendazole] for orphan drug status, the reviewer does not preclude to endorse this application and regulatory actions will be adopted after soliciting for standpoints of clinicians" (Statistical Review, 30-31). In the medical officer's concluding statement of factors that were considered in arriving at the approval recommendation it was noted that albendazole "qualifies for orphan drug designation" (Statistical Review, 31).

3. Alglucerase - Ceredase

In the April 1991 approval of this lysosomal enzyme, which is deficient in those with Gaucher's disease, the medical group leader noted, as the first of three issues to be considered in approving this drug, that "no well-controlled studies were conducted" (Medical Group Leader Memo, Dec. 26, 1990, 1). She went on to explain that, to her, there were 2 studies that demonstrated efficacy. One was a study in which liver biopsies were conducted before and 44 hours after a single infusion in 22 subjects. The other (seemingly more convincing) study was a 6 month study that compared 2 groups of 12 subjects on drug and 12 not on drug (or placebo). The subjects were not randomized and there were major differences in key baseline prognostic variables. Therefore, the most compelling data from the study were the change from baseline to end of study in the 12 subjects on drug in the key clinical parameters of anemia and spleen and liver organomegaly. The Medical Group Leader concluded that "this was convincing evidence of efficacy and because of the rarity of patients and the difficulty of following placebo-treated or untreated patients with severe disease for long periods of time, randomized studies were not required." (Medical Group Leader Memo, 1). This approval illustrated FDA's ability to exercise scientific judgment as well as to extend itself in aiding a sponsor with compiling the NDA as the FDA medical reviewer noted in his reviews.[13] [14]

11 This between group difference, even if statistically significant, hardly lends considerable weight to the biological plausibility that the drug "works" in that the group dosed for 7 days fared better than the group that was dosed for 14 days.

12 This list of factors is noteworthy because, although articulated by this FDA reviewer over 15 years ago, these same factors apply to many, many orphan diseases, yesterday, today and likely tomorrow's well.

13 NORD considered whether to classify this application as meeting the May 1998 clinical evidence of effectiveness standard for a single study because in that Guidance, FDA explains that "a single clearly positive trial can be sufficient to support approval of a replacement therapy...when it is combined with clear evidence that the condition being treated is caused by a deficiency of that factor. Demonstration of physical replacement of the deficient factor...provides strong substantiation of the clinical effect." However, in this case, there is no "single clearly positive trial." Evidence Guidance, 11.

14 This study could also be seen as a historically controlled study. FDA regulations recognize a historically controlled study as one of 5 enumerated types of "adequate and well-controlled studies", 21 C.F.R. § 314.126(b)(2)(v). Use of a patient as his/her own control is a variant of the historically controlled study model. However, FDA in its regulations, notes that historically controlled studies should be reserved for "special circumstances" because pertinent variables can not be controlled and such special circumstances include where the effect of the drug is "self-evident"; however, here the effect of the drug was measured on organ volume and anemia which are less "self-evident" as caused by the investigational drug than examples given in FDA's regulation such as general anesthetics.

4. Alitretinoin - Panretin

In this February 1998 approval for the treatment of AIDS-related Karposi's Sarcoma (KS), FDA found that one of two Phase 3 trials was clearly positive but the second Phase 3 trial was stopped early and as the November 17, 1998 memo of the statistical reviewer concluded, "doubts remain as to the appropriateness of the interim analysis and robustness of the response rates in the trial that was stopped early." As noted earlier in that same review, "if the [FDA] medical reviewer's assessment would have been used as evidence for stopping the trial early, the trial would not have been stopped and one would conclude that there was no statistically significant difference between the arms" (Statistical Review, 14). In addition, the secondary endpoints measuring various "time to event" outcomes did not show numerically different results in the two treatment arms, except that in one of these secondary endpoints (median time to progression) the placebo arm results were much better (that is, took much longer for subjects to progress on placebo than on drug). The FDA reviewer noted that "this is somewhat unexpected considering the superiority of response rate in the sponsor's assessment" (Statistical Review, 13).[15]

5. Alpha1 – Proteinase Inhibitor (Human) - Prolastin

In this December 1987 approval of a replacement protein for those who are genetically deficient in alpha-1-antitrypsin, FDA in the approved labeling cites one uncontrolled study of 19 subjects, all with the same phenotypic variant of this deficiency, the most severally affected variant of which there are many variants. In that study, there was within a few weeks a change from baseline reported in two measures, alpha-1-antitrypsin levels and antineutrophil elastase capacity, as ascertained by bronchoalveolar lavage. However, FDA also notes that the disease manifests itself as emphysema in the third or fourth decade of life but that the "pathogenesis of development of emphysema in alpha1-antitrypsin deficiency is not well understood at this time." (Label, 1).

This approval clearly demonstrates the exercise of scientific judgment by FDA. The May 1998 FDA Guidance speaks to a single study sometimes being sufficient to support approval of a replacement therapy "when the pathophysiology of a disease and the mechanism of action of a therapy are very well understood." (Evidence Guidance, 11). Therefore, while NORD classifies this approval as "case-by-case" flexibility, if one were to conclude that the conditions of the May 1998 FDA Guidance had been met, then the classification instead would be "administrative" flexibility, which is evidence of FDA flexibility nonetheless.

6. Ambrisentan - Letairis

This June 2007 approval of a drug for pulmonary hypertension was approved under 21 C.F.R. Part 314, Subpart H. However, it was not approved under Subpart H for reasons related to its evidence of efficacy, such as its registration studies having been conducted using an unvalidated surrogate as their primary endpoint. Instead, this was approved under Subpart H restrictions on distribution for safety concerns. NORD surmises that this will be the last drug ever so approved because several months later, the Food and Drug Administration Amendments Act granted FDA authority to impose a REMS as part of marketing approval and as part of the REMS to include in some cases, Elements to Assume Safe Use (ETASU). Therefore, NORD believes that drugs formerly approved under Subpart H with safety restrictions such as Actiq, Thalomid, and Bosentan will in the future be approved with a REMS that includes ETASU. NORD includes this discussion here to illustrate a point made earlier in this paper that can otherwise be confusing and that is, Figure 1 to this paper includes footnotes that denote each drug approved under Subpart H for efficacy reasons, under Subpart H for safety concerns and under Fast Track. NORD thought it critical to include this discussion of Ambrisentan so that the diligent reader who checks on all the Subpart H orphan drug approvals and discovers some NORD would not have otherwise included because they are Subpart H orphan drugs but only because of safety concerns can now understand the reason for the apparent discrepancy.

7. Amifostine - Ethyol

This drug was approved in December 1995 and illustrates all the principles that would later be articulated by FDA in its "single study" with a "statistically very persuasive finding" and where another study is likely unethical. (Evidence Guidance, 12-16). Therefore, NORD classifies this as a case of "administrative" flexibility even though this approval predated the issuance of FDA's May 1998 Guidance.

8. Anagrelide HCl - Agrylin

In this March 1997 approval for treating essential thrombocytopenia, the approved indication was "to reduce the elevated platelet count and the risk of thrombosis and to ameliorate associated symptoms." The FDA-approved labeling refers to two "historically controlled, unblinded" studies in a total of about 300 subjects. The statistical review states that these two trials were both Phase 2 open-label trials that were "patient controlled," and baseline-controlled or "patient as own control" (Statistical Review, 2), which, as discussed earlier, are a form of historical control. The statistical analysis review supports this by only describing changes in each subject from that subject's baseline platelet count (without any reference to any natural history control group). While the statistical review mentions that associated symptoms were a secondary endpoint in the larger of these two Phase 2 trials, the review never mentions any results of that analysis of symptoms in its memo and moreover, there is no mention at all in either study that risk of thrombosis was assessed as measured by thrombosis events or any endpoint or instrument. (The medical reviewer's memo for efficacy is not publicly available.) While this drug's approval may illustrate some exercise of scientific judgment, NORD classifies this approval as having met the 2

15 To NORD, there are not two adequate and well-controlled studies clearly positive in this case, so this approval shows an exercise of scientific judgment. Also, while this indication is for cancer, this approval was closely followed and seen by the AIDS patient community as an FDA action related to AIDS, more than for cancer, and so this approval has been included in this analysis.

adequate and well-controlled study efficacy standard or "conventional" approval.

This approval and text of its analysis is included to alert the readers that there are many cases that were classified as "conventional" which also have elements of flexibility. NORD anticipates that readers reviewing any one of these approvals classified as "case-by-case" flexibility may come to a different conclusion if asked to adjudicate that case. Therefore, for comprehensiveness, any reader would also have to re-adjudicate each of the 34 cases classified as "administrative flexibility" and 46 cases classified as "conventional" approvals in order to score all the cases, and the text of these analyses is not presented but for one or two exceptions such as this one for illustrative purposes. NORD appreciates that there is subjectivity in making these judgment calls, but the overall "gestalt" is clear.

15. Artemether/Lumefantrine - Coartem

In this April 2009 approval of this fixed-dose combination product for treating malaria, FDA medical and statistical reviewers both noted that there were only two trials that tested the combination against the single components and both studies were conducted at the same single center in China with a single racial group. Therefore, the statistical review of August 21, 2008 questions whether study results can be extrapolated beyond this region and this ethnic group. (Statistical Review, 11). The FDA medical review of November 25, 2008 states that the 2007 FDA draft Malaria Guidance recommends that the primary endpoint be 28-day cure as defined by FDA. (Medical Review, 34). However, the statistical review explains that "evaluation of FDA-defined cure rate is not possible [in these 2 studies] due to lack of information on clinical signs and symptoms as well as malaria-related laboratory abnormalities from the sponsor." (Statistical Review, 8).

The key finding here is that on the primary endpoint of 28 day cure rate, even without being able to employ the FDA defined cure rate, the combination failed to beat lumefantrine: in one study the p value for this comparison was 0.49 and in the other study, there were two comparisons of the combination to the lumefantrine component because the study had both a lumefantrine capsule arm ($p = 0.675$) and a lumefantrine tablet arm ($p = 0.16$). However, there were other non-primary endpoints that showed the value of the combination over both monotherapy components. Therefore, this approval required an exercise of scientific judgment.

19. Sodium Benzoate/Sodium Phenylacetate - Ucephan

In this December 1987 approval to treat urea cycle disorder, FDA demonstrated flexibility in that the March 20, 1986 medical review states that about 80% of subjects on this therapy in a study of 56 subjects in 45 sites survived compared to about a 15% survival rate historically for persons on dietary modification alone. The medical review concludes by noting that, "The usual requirements for a statistical evidence of efficacy though not fulfilled, the volume of data accrued over almost 6 years in a multicenter study appear reasonably adequate." (Medical Review, 43-44). Furthermore, the review's final paragraph before its approval recommendation notes that this drug has been designated as an orphan drug.[16]

21. Betaine HCl - Cystadane

In this November 1996 approval to treat an inborn error of metabolism, homocystinuria, FDA exhibited enlightened exercise of scientific judgment in that all data were drawn from published literature and there was only one randomized, double-blind placebo-controlled trial and it failed. This study looked at the effect of vertebral bone density. This six patient trial was one year in length. According to the medical review of June 19, 1996, "results showed that bone density measurements determined after 6 and 12 months Bentaine prescription did not differ from those after 6 and 12 months of placebo." (Medical Review, 9). However, in the other 17 published trials including 78 patients, the sponsor concluded that homocysteine levels were reduced by bentaine, and 48 of these 78 patients also reported some clinical response in addition to the biochemical response. The medical reviewer's first observation under "Discussion and Conclusions" was the following:

> This NDA is generally in poor condition, and the sponsor has made relatively poor use even of the published articles.... There are several reasons for such limited exertion. First, the company is a relatively newer entity and has had little previous experience with drug development; this is in fact the first NDA it has ever submitted to FDA. Other and larger companies show little or no interest in submission of an NDA for this drug in this disorder after [the FDA Office of Orphan Products Development] inquired after a sponsor for the product. Additionally, the disorder for which this new treatment is to be indicated was described only within the past four decades, and it is rare. Homocystinuria...has been estimated that only 800-1,000 cases in total have been found and reported in the United States. It is obvious that this company was not willing and/or able to spend much on original work in homocystinuria; it has depended entirely upon knowledge already in the medical literature.
>
> (Medical Review, 14).[17]

32. Chenodiol – Chenix

In this July 1983 approval for treating certain gallstones in patients at increased surgical risk, the preponderance of the clinical experience came from a placebo-controlled Natural Cooperative Gallstone Study (NCGS) of 916 subjects who were not at high surgical risk, and that studied two lower doses of this drug than the doses approved. The dose range approved

16 This case could alternatively be considered for classification as "administrative" rather than "case-by-case" flexibility but flexibility nevertheless.

17 To NORD, this seems to have been a candidate for Subpart H approval (because reduction of homocystine levels would seem to be an unvalidated surrogate that would be reasonably likely to predict ultimate clinical outcome) but this drug was not classified by FDA as a Subpart H drug. However, the difficulty of conducting a confirmatory Phase 4 trial may have been a factor (although following subjects on chronic administration of drug and comparing their outcomes to natural history/historical controls could have been explored and maybe it was.)

came from several uncontrolled studies. Since there are no publicly available medical or statistical reviews (and the drug has been discontinued from marketing), the label at the time of approval is the only available source of information on the efficacy evidence in the approved orphan population of subjects at high surgical risk and the label has only the following single sentence about one study considered in that population: "In a prospective trial using 15 mg/kg/day, 31% enrolled surgical risk patients treated more than 6 months (n=86) achieved complete confirmed dissolutions." (Label, 1).

Given that most of the discussion in the labeling is of the 916 subject NCGS that was in a different type of subject and at different doses than those approved, and given that the other clinical data are several uncontrolled studies, none of which appear to be restricted to high surgical risk patients, FDA appears to have extrapolated from these clinical data set to the dose approved in the high surgical risk orphan patient population. Therefore, this seems to exhibit an exercise of some modicum of scientific judgment, although this classification necessarily has to be tentative given the lack of medical and statistical reviews in this case.[18]

33. Cinacalcet HCl - Sensipar

In this March 2004 approval for treating hypercalcemia in patients with parathyroid carcinoma, the data on patients with the orphan condition came from a Phase 2, open-label study of ten subjects. (However, there was ample clinical data from three randomized, double-blind placebo controlled trials in chronic kidney disease (CKD) patients with secondary hyperparathyroidism with about 1200 total subjects enrolled, and this clinical evidence from a prevalent disease [likely showing the drug's ability to reduce serum calcium] may have played a significant role in FDA's consideration of the orphan condition.) The primary endpoint of the Phase 2 study was a reduction in the two-to-sixteen week titration phase in serum calcium of 1.0mg/dl or more and seven of the ten subjects met this, but the medical review of February 14, 2004 went on to note that: "None of the patients, however, normalized their serum calcium levels." (Medical Review, 18). This review of the efficacy evidence for the orphan condition concludes: "To state the obvious, the data upon which Amgen is requesting approval for the treatment of parathyroid carcinomas are very limited. Yet, parathyroid carcinoma is a rare disease and patients have few treatment options for the hypercalcemia associated with the condition. Cinacalcet offers the potential to satisfy an unmet medical need in this population of seriously ill patients." (Medical Review, 18-19, emphasis added).[19]

35. Coagulation Factor IX (Recombinant) - Benefix

In this February 1997 approval for treating hemophilia B, there were two studies that evaluated clinical results in the Summary Basis of Approval. In one study of 37 subjects who had been previously treated with moderate to severe hemophilia, 82% of all bleeding episodes in the peri-operative period required only one infusion for resolution. In the second study, the drug was administered for 13 procedures in 12 subjects. Ninety-seven percent of clinical responses were rated subjectively by the physician or patients as excellent or good, and transfusion of blood products was needed in only three of the 13 surgical procedures with hemostasis maintained throughout the surgical period without any clinical evidence of thrombotic complications.

Since there was no discussion of a historical control group nor of the prior experiences of any of the subjects in either trial (therefore, no analysis could be made using each patient as his/her own control), there was, in NORD's opinion, some exercise of scientific judgment in this approval. NORD would have classified this as "case-by-case flexibility" had NORD not consulted with a senior FDA hematologist on this and most of the other orphan blood disorder biologic approvals (see also # 9, 10, 11, 36, 37, 38, 99, 102, 104 & 132). This set of blood disorder biologic approvals is, to NORD, the most opaque in terms of understanding whether the quantum of efficacy evidence would have been sufficient for approval even if these disorders were prevalent and not rare and if these approvals required any exercise of FDA administrative or case-by-case flexibility. NORD includes this one only to illustrate the value of the insights provided by the FDA official. In this case, the FDA official explained that while NORD's catalogue of the evidence is accurate, the conclusion is wrong because, to the FDA official, even if one million Americans had the condition, this quantum of evidence would have been adequate for approval. The official explained that this therapy simply replaces a protein that is missing and that replacing the missing protein by giving one unit of this product will predictably raise blood levels of that protein by a certain amount and it is well established in hematology what the blood levels of that protein should be for surgery, for satisfactory hemostasis and for other situations. (See also #36 Mononine and #37 Alphanine, which are plasma-derived and for which this same paradigm applies.) However, there are two issues to note: one, Benefix was the first recombinant Factor IX, and therefore, there were safety issues that needed to be addressed such as immunogenicity concerns, and two, this approval was in 1997 and FDA likely would hold a new Factor IX product today to a more demanding efficacy requirement that includes a demonstration of the drug's effect in surgery and on the treatment of bleeding generally. But, as of 1997, the quantum of efficacy evidence provided with this application was not only sufficient for approval for this orphan condition but would have been adequate even if the condition had been prevalent. Therefore, the classification here is "administrative flexibil-

18 It is to be noted that the lack of medical and statistical reviews is nearly unique to this drug in this analysis of non-cancer orphan drug approvals. Also, it is worth noting that this drug was designated as an orphan in September 1984 but had been approved in July 1983.

19 While NORD did not see any statement in the FDA review documents that in this orphan condition, the serum calcium levels never fall spontaneously, if this is nevertheless the case in this condition, then this drug may be classified as having been approved under the May 1998 Guidance conditions and could then be classified as "administrative" flexibility, but again, still flexibility.

ity" and not "case-by-case flexibility."[20]

38. Coagulation Factor VIIa (Recombinant) - NovoSeven

In this March 1999 approval for treating bleeding episodes in hemophilia A or B patients with inhibitors to Factor VIII or Factor IX, the situation was wholly different than, for comparison, for Benefix (Factor IX, #35 above) because with Benefix, the protein was simply providing that which was missing in that individual, whereas the scientific basis for NovoSeven was mainly to provide Factor VII in order to bypass the cascade of Factors VIII or IX since these patients had inhibitors to these other two factors. (Note—there are persons who are deficient in Factor VII but that is not the approved indication here.) In other words, this was not the case of simply supplying exogenously that which was missing endogenously, but this was more akin to a more typical pharmacotherapeutic intervention that relies upon pharmacological intervention to achieve its therapeutic benefit.

Therefore, in this case, the standard expectation of clinical evidence of effectiveness would be expected. However, in this case, there were compassionate use, open-label studies of NovoSeven but, as described in the FDA summary basis of approval (SBA) of March 22, 1999 for these, the "clinical data from [these] were insufficient to evaluate the safety and efficacy of the product by statistical methods." (SBA, 7). There was also one double-blind, randomized trial comparing two doses of NovoSeven for which FDA states that: "No comparisons between NovoSeven and other coagulation products have been made; therefore, no conclusions regarding the comparative safety or efficacy of NovoSeven can be made." (SBA, 7). In consulting with an FDA hematologist, there was, of course, an adequate and sufficient scientific basis for this product approval based on the therapy, the condition, the compassionate use data, the dose ranging study, pharmacodynamic and pharmacokinetic studies and animal data; however, this quantum of evidence would not have been sufficient had this therapy been proposed for a prevalent use. Most importantly, FDA here was exercising "case-by-case flexibility."

41. Corticorelin Ovine Triflutate - Acthrel

In this May 1996 approval for differentiating between pituitary and ectopic production of ACTH in patients with ACTH-dependent Cushing's syndrome, the medical officer's review of April 5, 1981 notes that the NDA rests upon two "pivotal" bioequivalence trials comparing the sponsor's ovine corticorelin releasing hormone (oCRH) and the NIH preparation in order to rely upon all the NIH published data to support the efficacy (and safety) of this product. However, the medical reviewer notes that the corticorelin releasing hormone (CRH) in the published studies were different formulations and sometimes human and not oCRH was used. Moreover, the hormonal response to the oCRH was not defined in either the two "pivotal" bioequivalence trials (which were in 20 "normal" subjects and 10 "normal" subjects) or in the published literature. As for the published literature the medical reviewer notes that all were submitted under the heading of "well-controlled studies without case report forms" (Medical Review, 4) and: (1) that the oCRH formulations differed from study to study and in some, human CRH was used (and with respect to these, the medical reviewer states that "these reports...using [human CRH] do not support any claims on oCRH"); (2) "that efficacy is defined differently from report to report"; (3) that the overall quality of the publications differ widely; and (4) "that the Agency does not have access to the original data to support or discuss the sponsor's claims" (Medical Review, 5).

42. Cysteamine Bitartrate - Cystagon

In this August 1994 approval for treating nephropathic cystinosis, the medical reviewer (in the medical review of October 27, 1993) relied upon 3 open-label multicenter studies: the National Collaborative Cystinosis Study (NCCS), a so-called "Long Term Study" and a UK retrospective study. (Medical Review, 44-45). However, the medical reviewer states that the UK retrospective study only provides "supportive evidence of efficacy" because "a minimal deterioration in renal function [the study's primary endpoint] occurred in the treated group in the UK study" (Medical Review, 45), and therefore, if any inference can be made about the efficacy of the compound in this disease from this study, it would be against, and not for, the drug's efficacy.

As for the "Long Term Study," the "primary endpoints of death and renal death (need for dialysis or transplant) were compared to historical controls represented by 205 unselected, unassociated cystinotic patients analyzed in a retrospective European study. The comparison between the two groups was not prespecified in the study protocol. Inferential statistical testing of the differences was not done because numerical values for each data point were not available for the untreated controls." (Medical Review, 45, emphasis added). Consistent with this, the statistical review characterized this "Long Term Study" as not well-controlled. Therefore, it is difficult to regard this as a positive trial.

As for the NCCS, the comparison group was noted by the medical reviewer as "a group of patients treated with placebo in a previous double-blind study of ascorbic acid for the treatment of cystinosis." (Medical Review, 4). The statistical review of December 13, 1993 stated that, "there were statistically significant differences between the cysteamine and placebo groups in terms of age at diagnosis, age at entry, height and renal function for evaluating patients. Due to historically controlled study and insufficient sample size for placebo (n=17) it is very difficult to have meaningful inference between treatment comparisons." (Statistical Review, 16). It is difficult as

20 This is the only product approved for which NORD provides this kind of explanatory text, because it illuminates the rationale and reasoning behind the classifications for several other blood disorder biologic orphans as being approved with an exercise of "administrative flexibility" (see also #9, 10, 11, 36 & 37). Including this explanation of Benefix also provides an opportunity to illustrate how efficacy standards evolve over time and evidence sufficient at some earlier point in time may no longer be prognostic for FDA action on a later similar product.

well, therefore, to consider the NCCS as a positive adequate and well-controlled trial. However, the role of historical controls in this setting likely provides the basis for the finding of efficacy here.

46. Deferasirox - Exjade

In this November 2005 approval for treating chronic iron overload in patients with transfusion-dependent anemia, the Subpart H/Fast Track approval was based essentially on one single, non-inferiority trial on an unvalidated surrogate primary endpoint. At the pre-NDA meeting, the Division had generally agreed to the statistical methods but had "indicated that the efficacy of DFO [deferoxamine mesylate, the active control] would have to be established and that the margin of -15% would have to be justified in the NDA." (Medical Review, 39 [October 26, 2005]).

DFO had been approved prior to 1982 and is the only FDA approved drug for this use. (Medical Review, 1). The reviewer stated the following with respect to the primary endpoint results of this study: "Exjade [deferasirox] was to be declared non-inferior to DFO if the lower limit of the 2 sided 95% CI for the difference in the percentage of treatment success between Exjade and DFO in the [primary population] was above -15%. For the entire primary population this goal was not achieved. [The success in Exjade was 52.9% and in DFO was 66.4% and the lower limit of 95% CI for difference in percentage of treatment success was -21.6%.] This led the sponsor to segment the primary population into multiple subcategories to determine whether or not non-inferiority could be achieved for any subgroup." (Medical Review, 47). The review then shows the results for eight post-hoc subgroup analysis; it is unclear if this is an exhaustive list of all subgroups analyzed post-hoc. The reviewer then comments that: "These results are problematic. Analysis should be prespecified, not retrospective. The identification of a subgroup in which efficacy is demonstrated can be used for hypothesis generation, but not to provide support for efficacy to gain approval of the drug. Subgroup analysis should lead to a prospective study to establish efficacy in that subgroup. However, the sponsor's argument has merit even though the sponsor's predicament is of its own devise." (Medical Review, 49).

47. Dexrazoxane HCl - Zinecard

This drug was approved in May 1995 for preventing cardiomyopathy associated with doxorubicin. While it appears from three randomized, placebo-controlled trials that the drug is able to prevent and/or reduce the incidence and severity of doxorubicin-induced cardiomyopathy, FDA also notes that in the largest of these 3 trials, which was in breast cancer patients, the patients on the doxorubicin arm with dexrazoxane, "had a lower response rate (48% vs. 63%, p = 0.007) and a shorter time to progression than those who received [doxorubicin without dexrazoxane], although survival of patients who did or did not receive [dexrazoxane] was similar." (Label, 13 [comments by the medical reviewer], April 28, 1995).

48. Diethylenetriamine pentaacetic acid - DTPA

In this August 2004 approval for treating patients with known or suspected contamination with plutonium, americium or curium to increase the rates of elimination, FDA had announced in a September 15, 2003 Federal Register notice (that was prior to submission of this NDA) that FDA had already concluded that the drug would be effective based on FDA's review of the Federal government's "database on 646 patients who received one or more doses...during the past 40 years.... In these patients, administration of [this drug] increased the rate of radiation elimination in the urine on average of 39-fold." (60 Fed. Reg. 53,984, 53,986). FDA had established in 2002 a regulation for assessing the safety and efficacy of drugs to deal with the radiation that may be emitted from a "dirty bomb" or other bioterrorism agents. (21 C.F.R. Part 314, Subpart I (so-called "animal efficacy rule"). However, FDA did not require the sponsor to conduct such animal studies either pre- or post-approval.[21]

55. Ethanolamine Oleate - Ethamolin

In this December 1988 approval for treating patients to prevent rebleeding in esophageal varices that have recently bled, it appears from the FDA approved labeling at the time of approval that the demonstration of efficacy was based upon the clinical pharmacology of the drug that causes "fibrosis and occlusion of the vein" when injected intravenously. "The time course of these findings [from human autopsy studies] suggests that sclerosis of esophageal varices will be a delayed rather than an immediate effect of the drug."[22] (Label, 1).

57. Fomepizole - Antizol

In this December 1997 approval for the treatment of methanol or ethylene glycol poisoning, the medical reviewer concludes that:

> [T]here seem only two courses possible for this application at this time: (1) [Not Approvable] the entire application so that the company may then perform some decent studies...as a sole therapy employed; (2) approve the NDA for use...only as an adjunct to use of hemodialysis and require the studies under (1) immediately above as phase IV trials. It is true that even if this preparation were completely unavailable at this time...there would be no great hardship or loss. Ethanol, even though it may be more difficult to use, is still an adequate therapy." (Medical Review, Nov. 13, 1987,13). Earlier in his review the medical officer stated the following conclusion on efficacy: "Efficacy when fomepizole is given as a single...agent has not been demonstrated in any sort of controlled study (even historical control).

(Medical Review, 10). The NDA had 2 studies submitted and some historical control data dating back to 1946. The statis-

21 If one regards the experiences of the 646 persons over 40 years who received at least one dose of this drug as having had their results compared to historical controls, then this classification may move from "case-by-case flexibility" to "administrative flexibility".

22 Even if the clinical pharmacology of the drug is a very good surrogate, two positive adequate and well-controlled studies with that surrogate would be needed to satisfy the conventional showing of evidence. If any reclassification were to be considered here on the basis of the use of the clinical pharmacology of the drug, the alternate classification would be under the May 1998 Guidance and therefore, "administrative flexibility".

tical reviewer observes that, "from 1965 to the present, the administration of ethanol as an antidote and the use of renal dialysis have been the treatments of choice." (Statistical Review, Oct. 16, 1997, 4). As for the two studies, the statistical reviewer states that "interpretation of the efficacy results are confounded by the use of ethanol...and the use of hemodialysis in both studies." (Statistical Review, 3). As for the historical data, the statistical reviewer concludes that, "this reviewer does not believe that the historical data is helpful in establishing the efficacy of the drug." (Statistical Review, 5). The statistical reviewer's overall conclusion is that "the sponsor's efficacy database consists solely of data from open-label uncontrolled studies; therefore, there are no statistical issues [because there are no data to analyze statistically]. The descriptive data does not clearly delineate the effects of [fomeprizole] alone since the majority of the patients were treated... in combination with ethanol and/or hemodialysis." (Statistical Review, 4, emphasis in original). Consistent with the statistical reviewer's report, the medical reviewer recommended either not approval of the NDA or approval as an adjunctive indication and the NDA was approved as first line therapy.

58. Fosphenytoin Sodium - Cerebyx

In this August 1996 approval for the acute treatment of patients with status epileptic of the grand mal type, the drug approval is the prodrug use of phenytoin and the medical reviewer notes that it is, "rapidly and completely converted to phenytoin in vivo." (Medical Review, Feb. 1, 1996, 18). The medical reviewer further comments:

> This NDA is unique in many ways. First, there are no controlled trials to support the efficacy of [this drug]. The 'controlled trials' submitted were really not designed to show a difference between treatment groups on a protocol specified efficacy outcome. The majority of patients studied were not having seizures, but were only at risk for seizures.... Secondly, the bioequivalence data... really only applies to the isolated instance of IV loading. To my knowledge, no bioequivalence data for IV maintenance dosing, IM loading, or IM maintenance dosing has been submitted.
>
> (Medical Review, 19).

59. Gallium Nitrate - Ganite

In this January 1991 approval for treating clearly symptomatic cancer-related hypercalcemia that has not responded to adequate hydration, the Division Director expressed serious concerns about the nature of the efficacy evidence, specifically, "the participation of only one principal investigation... in the pivotal clinical trials, the performance of the clinical studies, essentially in only one clinical center (Sloan Kettering) [and] Sloan Kettering holds the use patent on the drug." (Division Director's memo, Sept. 28, 1990, 1).

While the single randomized trial comparing gallium to calcitonin reported "achieving normocalcemia in much higher percentage of gallium treated patients than calcitonin treated patients, the overall survival of patients in both treatment groups was poor (median survival time was 29 days for gallium and 35 days for the [calcitonin] group." (NDA Review, Dec. 11, 1989, 4). With respect to the treatment effect of more patients on gallium achieving normocalcemia, "the treatment effect would not be significant if the expected percentage (60%) of calcitonin patients [had] achieved normocalcemia."[23] (Statistical Review, 8, September 20, 1989).

60. Ganciclovir Sodium - Cytovene

The June 1989 approval for treating cytomegalovirus (CMV) retinitis in immunocompromised patients with AIDS presents a very complicated regulatory history. FDA urged, and the sponsor had agreed, to conduct a prospective, randomized, no-treatment controlled study; however, the NDA ended up being approved on a post-hoc, retrospective review of a case series of subjects treated by one physician at the Johns Hopkins University, and it is that "study" and only that study whose results are shown in the FDA approved labeling. [24]

64. Hemin - Panhematin

In this July 1983 approval for ameliorating recurrent attacks of acute intermittent porphyria (AIP) and similar symptoms in other patients with AIP, porphyria variegata and hereditary coproporphyria, the SBA notes that, "sorbitol serves as a useful stabilizer" in this drug product (SBA, 1), but the SBA later lists five published open studies that were conducted with a formulation without sorbitol and only one, "progress report" of an open-label study with a formulation containing sorbitol. (SBA, 6-8). These six reports together with a couple single dose case reports totaled 125 subjects, of whom over 85% experienced symptom relief on this drug. (SBA, 6). Of the five studies of the formulation without sorbitol, study 1 administered the drug to seven subjects for three to 13 days, study 2 treated 28 subjects for one to six days, study 3 treated 11 patients for three to 13 days, study 4 treated 57 patients for "an unspecified time period" and study 5 treated eight subjects for three to five days, and despite the short duration of these treatments a total of 13 of these 111 subjects died. (SBA, 6-8).

In the single "progress report" that administered the drug in a formulation with sorbitol, these seven patients received drug for two to five days and none were reported to have died. (SBA, 8).

Overall, there was no concurrent control in any study and no reference to any historical control. Moreover, if these studies relied upon each patient's prior clinical experience as his/her own control, there were no reports of the previous patient experiences without the drug. Given the design of these very short duration, open-label, uncontrolled studies for which no mention was made whether line listings, case report forms or even protocols were ever made available to FDA, this may be a case in which FDA relied upon historical controls that were not well documented.

70. Imiglucerase - Cerezyme

In this May 1994 approval for treating Type 1 Gaucher's Dis-

23 Achieving normocalcemia may be an appropriate surrogate and putting aside the concerns expressed about the single investigator at a single site that has a financial interest in the outcome of the study and the overall poor survival results, then this drug's classification may change to "administrative flexibility".

24 If one regards that the Hopkins case series was reviewed as though it was compared to a historical control, then this approval may be "administrative flexibility".

ease, the single "pivotal" study compared the 1981 human placenta-derived form of this drug (Ceredase) to the proposed recombinant version, imiglucerase (Cerezyme). The statistical review of September 2, 1983 reports that the primary endpoints were an increase in hemoglobin concentration of at least 1 g/dl, an increase in platelet count and a decrease in liver and spleen volumes over the 6 month study duration in the 15 imiglucerase and 15 Ceredase subjects. (Statistical Review, 2).

According to the statistical review, "The sponsor failed to detect a statistically significant difference with regard to the proportion of patients ([Cerezyme] 11/15, Ceredase 12/15) who achieved an increase of at least 1 g/dl in hemoglobin concentration from baseline to conclusion of 6 months of double-blind treatment. A 95% confidence interval for the between-treatment group difference ([Cerezyme]-Ceredase)...is (-61.5%, 48.1%) which of course is extremely wide." (Statistical Review, 3). Similar nonsignificant results with wide confidence intervals are seen in the other primary endpoints as well. The statistician was concerned over these wide confidence intervals and to illustrate this noted that, "the 95% confidence interval...indicates that it is statistically conceivable that the Ceredase increase [in hemoglobin concentration] may be as much as 0.52 g/dl greater than the [Cerezyme hemoglobin] increase." (Statistical Review, 3).

Applying non-inferiority margins to the first efficacy parameter (preserving at least 50% of the benefit seen in the approved active control) would lead to the following:

> (1) mean hemoglobin concentration was increased by 1.53 g/dl over baseline in the Ceredase arm and therefore, to be non-inferior the Cerezyme would need to have a lower 95% confidence interval that was greater than +0.765 g/dl, when the lower CI for Cerezyme was -0.52; and
>
> (2) mean platelet count in the Cerezyme arm was increased by 16.13 x 10-3/mm-3 which means that the lower 95% CI in the Cerezyme arm needed to be greater than +8.065 x 10-3/mm-3, but it was -8.11.[25]

71. Interferon Beta-1a - Avonex

In this May 1996 approval for treating patients with relapsing forms of multiple sclerosis, evidence of efficacy at the time of initial approval came from one randomized, double-blind placebo-controlled study in 301 subjects. The primary endpoint, time to progression, was statistically significant at a p value of 0.02 and the secondary clinical endpoints were generally significant: change in Expanded Disability Status Scale ($p = 0.006$), number of exacerbations ($p = 0.03$), percentage exacerbation free ($p = 0.10$, not significant) and annual exacerbation rate ($p = 0.04$). The secondary MRI endpoints were number of lesions at end of year 1 ($p = 0.02$), at end of year 2 ($p = 0.05$), T2 lesion volume at end of year 1 ($p = 0.02$) and at end of year 2 ($p = 0.36$). (See Label, 8-9).

While FDA had previously approved another interferon beta compound for MS in 1993, FDA had determined that, for orphan drug purposes, these two were different drugs. Accordingly, Avonex was approved on the basis of this single study (and without reliance on the efficacy results for the previously approved interferon beta drug for MS) in which the primary endpoint results are not "very persuasive" (that is, not less than a p of 0.01).[26]

74. Ferric Hexacyanoferrate (II) - Prussian Blue, Radiogardase

In this October 2003 approval to treat patients with known or suspected internal contamination with radioactive cesium or thallium, there were no prospectively randomized controlled clinical trials and the "best human data on the efficacy of Prussian Blue will come from retrospective analysis of data on accidentally contaminated patients...treated with Prussian Blue." Such studies cannot, of course, be powered to achieve statistical significance [and] no formal statistical analysis has been performed on this data." (Medical Review, Sept. 15, 2003, 19).

Nevertheless, the medical reviewer concluded that, "although these publications all describe retrospective studies and the number of patients is small compared to a typical Phase 3 clinical trial, the evidence [of effectiveness] is compelling." (Medical Review, 42). The reviewer explains that, "in this retrospective study each patient served as his/her own control. For each patient the half-life during treatment was compared to the half-life after treatment had stopped, which was assumed to be equal to the half-life if no treatment had been given." (Medical Review, 42). Also, the reviewer pointed to animal efficacy data including that, "Prussian Blue has been shown to consistently decrease the half-life of 137 CS [radioactive Cesium] in dogs, rats and farm animals." (Medical Review, 43).

76. Laronidase - Aldurazyme

In this April 2003 approval for treating patients with mucopolysaccharidosis-I (MPS-I), efficacy was established on the basis of a single randomized, placebo-controlled trial of 45 MPS-I patients. The medical review of April 25, 2003 explained, that, "the study would be considered statistically significant if both primary endpoints of forced vital capacity and 6 minute walk meet or exceed the critical p-value of 0.05 in the difference between the treatment groups." (Medical Review, 20, emphasis in original). While the forced vital capacity between treatment group difference was statistically significant ($p = 0.02$), the 6 minute walk between group difference was not statistically significant ($p = 0.07$). (See Label, 1). Moreover, there were four prespecified secondary endpoints and only one was statistically significant: apnea ($p = 0.14$), liver volume ($p = 0.001$), Disability Index ($p = 0.99$), and shoulder

25 If the comparison between the investigational and the active control is not a non-inferiority comparison, but rather the investigational arm's results are being compared to historical control, then this classification may change to "administrative flexibility".

26 This approval can be alternatively read as consistent with the Evidence Guidance if that single study example can be read broadly so as to regard the multiple positive secondary endpoint results and MRI lesion results, in addition to a novel primary endpoint, as fulfilling the May 1998 Guidance for a single study, and in that case, this approval would be considered "administrative flexibility".

flex (p = 0.99). However, the first tertiary endpoint of urinary glycosaminoglycans (GAG) was statistically significant (p = 0.001). The medical reviewer concludes that "two markers of in vivo enzyme activity were associated with significant reductions during the 26 weeks of [the pivotal trial]: liver size reductions and urinary GAG concentration. The response of these markers to laronidase has been consistently shown also in the pre-clinical experiments and in the Phase 1 clinical trial, as well as in the placebo-treated subjects switched to laronidase treatment during the open-label extension." (Medical Review, 113). Therefore, the reviewer concluded, "given the lack of alternative treatments in a rare disease with severe or fatal consequences, this reviewer recommends approval of laronidase, supported by the evidence of efficacy in the co-primary endpoints and favorable trends in subsets of MPS-I in secondary endpoints." (Medical Review, 113).

77. Lenalidomide - Revlimid

In this December 2005 approval for treating patients with transfusion-dependent anemia due to myelodysplastic syndromes (MDS), the approval was based primarily on the results of one single arm, non-randomized, not controlled study. The demonstration of clinical benefit was RBC transfusion independence, defined as having had any rolling 56 day period without need for any RBC transfusion during a treatment duration of up to 672 days. The reviewer commented that, "in MDS, which is a heterogeneous disease, single arm studies using patients as their own controls are generally not acceptable. The sponsor definition of transfusion independence with a rolling duration as defined here is problematic in an unblinded study. In an end-of-Phase 1 meeting...FDA recommended a randomized controlled trial using an endpoint with a longer duration of response." (Medical Review, April 7, 2005, 65). FDA noted a randomized controlled trial with a longer duration of responses was ongoing at the time of approval. (Medical Review, 135). The first question put to the Oncologic Drugs Advisory Committee (ODAC) for a vote at its September 14, 2005 meeting on this drug was:

> Randomized controlled trials allow for direct comparisons of treatment effects and safety between treatment arms. A single arm study has been submitted using an 8-week run-in period to serve as baseline for each patient's transfusion requirements. A comparison is subsequently made to a follow up 8-week period on [lenalidomide] to compare transfusion requirements. Does the study design allow adequate characterization of [lenalidomide's] treatment effect in the population described in the proposed indication?

The ODAC voted yes = 11, no = 5. (Medical Review, 130). However, the ODAC may not have been aware that the comparison was not between periods of equal duration, that is, the comparison was not between the percentage of subjects who were transfusion independent in the run-in eight-week period to the first on-drug eight-week period, but instead the comparison was between the eight week run-in period and any rolling 56 day period of on-drug transfusion independence over a total of up to 672 days (that is, day 1 to day 56, day 2 to day 57, day 3 to day 58). The comparison was between each subject's transfusion independence over a single 56 day run-in period compared to up to as many as 671 rolling 56-day periods. In addition, to be included in the trial a subject had to have received "at least 2 or more units of RCBs within 8 weeks of study treatment." So, to be enrolled, a subject had to have had a run-in period with a transfusion of 2 or more units. (See, e.g., study inclusion criteria at Medical Review, 25. See also Medical Review, 43). Therefore, by definition, the "comparator" run-in eight-week period had to have had no subject who was transfusion independent, and there is no mention in the comprehensive 152 page Medical Review to that comparison between each subject's transfusion requirements during the run-in period and during the treatment phase except that FDA notes that 4.7% of the study subjects had only one transfusion in the eight-week run-in period (Medical Review, 44, but these were excluded from FDA's primary analysis of estimating the percentage that were transfusion independent in the treatment phase as protocol violators) and, "the statistical reviewer noted that there was a correlation in the number of pre-treatment RBC transfusions and the transfusion response. It is more likely for those patients with less than or equal to 5 pre-treatment transfusions to develop a transfusion independence response." (Medical Review, 65).

78. Lepirudin - Refludan

In this March 1998 approval for "anticoagulation in patients with heparin-induced thrombocytopenia and associated thromboembolic disease in order to prevent further thromboembolic complications," the efficacy evidence came from two non-randomized, open-label multi-center (all sites in Germany) trials using a historical control comparator group. However, as noted in the FDA approved labeling, "the key criteria of efficacy ...[was] platelet recovery...[but] comparable rates for the historical control group cannot be given, because [...] platelet counts were not monitored as closely as in the Refluden group."

Reliance upon a historical control group is fraught with uncertainty generally for many reasons which have been well-articulated elsewhere. However, FDA has relied upon such comparators in the case of rare conditions where the ability to have sufficient subjects to randomize to both the investigational and a concurrent control arm is limited, if not non-existent. (See Label; see also FDA approvals of Myozyme #101 and Ceprotin #99 for infantile-onset Pompe disease).

79. Levomethadyl Acetate HCl - Orlaam

In this July 1993 approval for treating heroin addicts suitable for maintenance on opiate agonists, active control (methadone) Phase 3 trials established that response to treatment for levomethadyl acetate was similar to that for methadone. However, there was no formal non-inferiority testing and, although an Advisory Committee indicated it was willing to accept a placebo-control in this patient population, there were no Phase 3 methadone-controlled studies that also included a placebo arm to establish the assay sensitivity of that study design and conduct. Because of the lack of formal statistical

comparisons of the treatment effects of levomethadyl acetate to methadone and the lack of a concurrent placebo arm in any of the Phase 3 trials, the classification is "case-by-case flexibility."

81. Mecasermin Rinfabate Recombinant - Iplex

In this December 2005 approval for treating growth hormone insensitivity syndrome, FDA accepted the sponsor's arguments that "the need for a concurrent control group was obviated by obtaining a documented pre-treatment height velocity in each subject for comparison to on-treatment height velocity [and]...that it was furthermore unnecessary due to the well-known natural history of the condition, in which the poor height velocity is not expected to improve spontaneously." (Statistical Review, Aug. 28, 2005, 7). In a single prospective, open-label multicenter study 36 prepubertal subjects received either 1 mg/kg or 2 mg/kg daily and on the primary endpoint of height velocity, the 1 mg/kg pretreatment values were 3.4 cm/year compared to on-treatment values of 7.4 cm/year ($p < 0.0001$) and the 2 mg/kg cohort had pretreatment height velocities of 2.2 cm/year and on-treatment values of 8.8 cm/year ($p < 0.0001$). The statistician observed that "efficacy is supported by the fact that [the 2 mg/kg cohort] with a higher dose level had a larger growth velocity than [1 mg/kg cohort]," but the statistician also noted that because there was no randomization here, these differences have to be viewed with caution. (Statistical Review, 3).

82. Mecasermin Recombinant - Increlex

In this August 2005 approval for treating growth hormone insensitivity syndrome, FDA permitted the sponsor to pool post-hoc five small clinical trials (four open-label and one double-blind placebo-controlled) to permit a global efficacy analysis relying upon all 71 treated pediatric subjects from these five trials. The primary efficacy analysis was of the 58 subjects for which adequate pretreatment height velocity data were available so that paired t-tests could compare the pretreatment height velocities of the same subjects completing each year of treatment, and the pretreatment height velocity was 2.8 cm/year for these 58 subjects compared to 8.0 cm/year in the first year of treatment ($p < 0.0001$). Without FDA's exercise of scientific judgment in permitting this post-hoc pooling, the pairing of each of these five small trials separately for signs of efficacy would have been problematic.

84. Midodrine HCl - Proamatine

In the September 1996 approval for treating symptomatic orthostatic hypotension, three studies were submitted with the NDA, two with the original NDA and a third added later with respect to the two in the original NDA. The statistical reviewer stated the following conclusion:

> The first one...was supposedly a multicenter study, but only one site collected data, and only for 7 patients. This is too few data for the results to be useful. Because of the other difficulties with the study...this reviewer feels that the medical reviewer's (negative) conclusion for the study should be heeded. [Note: it is unusual for the medical reviewer to be an outside consultant as it was in this case: Dr. Joel Morganroth.] The other study... randomized 97 patients.... The analyses...by the sponsor and...by this reviewer did not show midodrine treatment effect.... There was no midodrine treatment effect compared to placebo as measured by the syncopal symptoms endpoint.

(Statistical Review, March 13, 1996, 9). With respect to the third study, the statistical reviewer said that:

> [T]his study demonstrates that midodrine treatment has a significant effect on systolic blood pressure, and appears to affect standing time and dizziness in this highly selective group of patients. This study is unable by design to show that this temporary effect can be sustained over long-term use. The study contributes very little toward establishing that midodrine is an effective treatment for orthostatic hypotension. The study was too short (seven hours), involved only one dose at the upper level of the dosing range and a three hour dosing interval, was compromised by potential unblinding, and was limited to an enriched population of patients known to respond to midodrine treatment.

(Statistical Review, Sept. 2, 1993, 6).

85. Miglustat - Zavesca

In this July 2003 approval for treating mild to moderate type 1 Gaucher's disease patients for whom enzyme replacement therapy (ERT) is not an option, the NDA was supported by two Phase 1/2 studies and one Phase 2 study with extension studies to each. In the two open-label uncontrolled monotherapy Phase 1/2 studies, there were four primary endpoints: reductions from baseline in liver and spleen volumes and increases from baseline in platelet counts and hemoglobin. According to the Label, "In study 1...the results showed significant...reductions...in liver volume of 12% and spleen volume of 19%, a non-significant increase from baseline in...hemoglobin... and a [non-significant]...increase in platelet counts.... In study 2...the results showed significant...reductions...in liver volume of 6% and spleen volume of 5%. There was a non-significant...decrease...in hemoglobin...and a non-significant increase...in platelet counts." (Label, 5). The statistical reviewer stated that, "Study 004 was an open label, randomized, comparative study with Cerezyme monotherapy as the control group." (Statistical Review, April 27, 2002,3). "The primary objective for the comparative study was to assess the tolerability of [miglustat].... The efficacy analysis of liver volume was exploratory since no clinically meaningful difference was hypothesized and no sample size was determined." (Statistical Review, 27). As for the overall results of these trials and then applications for switching patients from ERT to miglustat, the medical reviewer concluded that: " These results suggest that switching to [miglustat] monotherapy may have a detrimental effect in 'well-controlled' patients with smaller liver and spleen volumes, and higher hemoglobin and platelet counts at baseline who had been receiving ERT." (Medical Review, May 2, 2002, ii).

87. Monoctanoin - Moctanin

In the October 1985 approval of this compound made from medium chain fatty acids derived from coconut oil to dissolve cholesterol gallstones retained in the common bile duct, FDA had issued a Federal Register notice on December 10,

1982 inviting submission of an NDA. In addition to published clinical data, FDA relied on the existence of 4 animal studies already reviewed by FDA, as well as one additional animal (dog) study proposed by FDA whose design is described in the Federal Register notice, as well as in vitro data showing dissolution of gallstones in this compound. (See SBA, 3). Also, "in her memo of September 1982, Dr. Finkel reviewed the reports of clinical trials...published throughout 1981... {the medical reviewer} added reviews of 7 reports which have been published since that time. Results published in the literature support the claim that infusion of monooctanoin into the biliary tract is effective in dissolution of cholesterol stones...["in about 1/3 of the patients" from Medical Review, 6]. The treatment is attended with a high incidence of adverse effects." (Medical Review, Nov. 26, 1984, 2). In a multicenter study of 377 patients, 32% of the subjects were considered to have had a complete response (Medical Review, 20), however, there was not only no concurrent control but no comparison to historical controls or to using each patient as his/her own control and no formal established analysis of success versus any control arm. (Medical Review, 2-3).

88. Galsulfase - Naglazyme

In this May 2005 approval for treating patients with mucopolysaccharidosis IV (MPS IV), the evidence of efficacy was derived essentially from a single, randomized, double-blind, placebo-controlled trial of 39 subjects for 48 weeks. The primary endpoint of 12 minute walk test had a p value of 0.025 (which is not the usual standard for single study in FDA's May 1998 Guidance in that it would not appear to be "a statistically very persuasive finding.") The two secondary endpoints of improvement in rate of stair climbing and urinary GAG levels have p values of 0.053 and less than 0.001, respectively. Also, "among patients who had been randomized initially to placebo [for the double-blind 24 week phase of the trial], the increases after 24 weeks of Naglazyme treatment compared to the start of the open-label period were [comparable in magnitude to the improvements seen in cohort initially randomized to Naglazyme for the 24 week double-blind phase]." (Label, 1). In sum, the primary endpoint of this single pivotal study was less than a p value of 0.05 but greater than 0.01 (that is, not a "statistically very persuasive finding") and one of the two prespecified secondary endpoints was not statistically significant.[27]

92. Oprelvekin - Neumega

In this November 1997 approval for preventing severe thrombocytopenia and relieving the need for platelet transfusions following thrombocytopenia chemotherapy in patients at high risk of thrombocytopenia, two randomized, double-blind, placebo-controlled Phase 3 trials formed the basis of the efficacy evidence. In one study of those who had recovered from an episode of chemotherapy-induced thrombocytopenia, the primary endpoint of whether the patient needed one or more platelet transfusions in the next course of chemotherapy was met with a p value of 0.04. The second study evaluated whether platelet transfusions were needed in either of the next two chemotherapy cycles in patients who had not previously experienced chemotherapy-induced thrombocytopenia. In this study the primary endpoint trended in favor of drug but was not statistically significant. The FDA approved labeling cited one additional positive analysis which, "in an unblinded, retrospective analysis of the 2 placebo-controlled studies, 19 of 69 patients (28%) receiving [oprelvekin] and 34 of 67 patients (51%) receiving placebo reported at least one hemorrhagic event which involved bleeding." (Label, 1).

93. Pegademase Bovine - PEG-ADA, Adagen

In this March 1990 approval of this enzyme replacement therapy for ADA deficiency in patients with severe combined immunodeficiency (SCID), the clinical evidence of this drug's efficacy comes from its use in 6 patients with ADA-deficiency SCID. The medical reviewer summarized his review this way:

> [I]n view of the rarity of the disease, insufficient cases to study, the orphan status of the disease, the potential lethality of the disease and the non-toxicity of PEG-ADA, the weak data provided might be enough evidence of efficacy in this case.... The strongest support of efficacy is the dramatic biochemical and in vitro immunological modulation by PEG-ADA in these patients, the trend of decreased infections in these patients, and the non-toxicity of PEG-ADA.... School attendance, hospitalizations, bouts of pneumonia, and growth data were inconclusive.

(Medical Review, Addendum IV, Jan. 12, 1989, 2).

99. Protein C Concentrate - Ceprotin

In this March 2010 approval for "patients with severe congenital Protein C deficiency for the prevention and treatment of venous thrombosis and purpura fulminans," there was a single, 18 subject, open-label, non-randomized Phase 2/3 trial with a historical control, as well as a retrospective analysis of 11 other subjects who had been on drug. As described in the case above for Refluden (see #78), a historical control comparator was appropriate here, but it is unlikely that if this condition were prevalent and there was no lack of subjects to enroll in a study, that this showing of efficacy would have been sufficient.

100. Rasburicase - Elitek

In this July 2002 approval for treating malignancy-associated or chemotherapy-induced hyperuricemia, the primary clinical efficacy evidence came from a single open-label, randomized, active control (allopurinol) Phase 3 study and two Phase 2 studies.

The Phase 3 study randomized 27 patients to rasburicase and 25 to allopurinol. The primary endpoint was a measure of plasma uric acid levels, and rasburicase was robustly statistically superior to allopurinol, p value of < 0.001. (Statistical Review, Nov. 28, 2000, 6). On each of the three prespecified secondary endpoints, rasburicase was also statistically superior to allopurinol. (Statistical Review, 8). The two Phase 2

27 In his clinical team leader's memo, Dr. Hyde notes that at the January 15, 2003 Advisory Committee on this drug, "some [panel] members expressed a sentiment for liberalizing p-value criteria in diseases as rare and difficult, but important, to study as this." Clinical Team Leader's memo of May 27, 2005 at page 13.

studies were both open-label single arm trials with a total of 238 subjects in both studies combined. The response rate was 99% and 95% in these two studies with uric acid levels reduced by 88% in these studies. (Statistical Review, 2).

If the reviews had indicated the uric acid levels do not spontaneously return to normal, then the implied historical control would have converted the two Phase 2 trials into "adequate and well-controlled" trials and there would be no exercise of judgment in this approval since the efficacy evidence would be straightforward. Similarly, if the review had indicated that it would be unethical to replicate the Phase 3 trial, then the data from the single Phase 3 study would satisfy the single study policy articulated at Section C.3 of FDA's Evidence Guidance. However, neither of those conditions apply and therefore this approval demonstrates the exercise of some scientific judgment and warrants a "case-by-case flexibility" classification.

101. Alglucosidase ALFA - Myozyme

In this April 2006 approval for treating Pompe disease patients, the clinical efficacy evidence is derived from a single open-label historically-controlled trial in infantile-onset Pompe disease patients. The study enrolled 18 patients on Myozyme and compared their one year performance on Myozyme against a historic control group of 62 untreated patients with a primary endpoint of invasive ventilator-free survival and proportion of patients alive. The statistical review of April 27, 2006 summarized its conclusion this way:

> The historical control subgroup contains data from subjects with birthdates over 20 years. The applicant's analysis points to the potential for improved outcome over time due to more aggressive therapy and better availability of the therapies in more diverse geographic regions. The result from the [historical control] cohort, however, support the contention that the long-term survival of patients with infantile-onset Pompe disease, which are not treated with Myozyme, is poor. The comparison of data between the historical control subgroup and the Myozyme-treated subjects does suggest a treatment effect. This observation is not based on statistical conclusions, per se, but more on the visual inspection of the results in the Myozyme-treated subjects compared with results in the historical control subgroup. The qualification of the treatment difference is almost impossible. Not only are there the issues of improved outcomes, however slight they may be, over time among the untreated subjects, but there remains the issue of selection bias among the Myozyme-treated subjects.
>
> (Statistical Review, 32).

102. Recombinant Human Antithrombin - ATryn

In this February 2009 approval for the "prevention of perioperative and peri-partum thromboembolic events in hereditary antithrombin deficient patients," the efficacy data came from combining one Phase 2, single arm, open-label trial (n=13 evaluable) with one Phase 3, single arm, open-label trial (n=18) to achieve a pooled cohort of 31 subjects on ATryn. The comparison was to those treated with plasma antithrombin and their data for comparison were collected from a prospectively-designed concurrently conducted retrospective chart review of 35 subjects. If this condition were not so rare, it is likely that a more substantial quantum of efficacy information would have been needed than the non-inferiority comparison based on a pooled comparison of 31 subjects on investigational therapy to a retrospective comparator arm of 35 subjects.[28]

103. Respiratory Syncytial Virus Immune Globulin (Human) - Respigam

In this January 1996 approval for prophylaxis of respiratory syncytial virus (RSV) lower respiratory tract infections in infants and young children at high risk of RSV disease, the principal efficacy evidence was from a randomized, double-blind placebo-controlled study in children under 24 months of age and at high risk of RSV disease. In this trial of 510 subjects, the primary endpoint was "the reduction of the incidence of RSV hospitalization ($p = 0.047$)." (Medical Review, April 30, 1998, 5). Almost all the secondary endpoints also showed a statistically significant separation between placebo and drug arms. (Medical Review, 5). There were two other key trials reviewed: the Cardiac trial and the NIAID trial. According to the medical reviewer:

> The Cardiac trial was a...randomized, non-placebo controlled, single-blind study conducted in 429 children with congenital heart disease of less than 48 months of age at enrollment. A 31% reduction in the primary endpoint (RSV hospitalization) was noted in the treatment group compared to the control group ($p = 0.164$). Not statistically significant reductions were observed in the treatment group of RSV ICU stay, RSV-associated mechanical ventilation and supplemental oxygen use... Adverse events were more severe in the [drug] group (64 children had severe AE compared to 44 control group children).
>
> (Medical Review, 6).

"The NIAID trial was reviewed in detail at the December 2, 1993 meeting of the Blood Products Advisory Committee. At this meeting it was pointed out that the trial conduct was flawed (unblinded, local randomization at a major site)," and "The NIAID trial and the Cardiac trial did not demonstrate efficacy in infants with congenital heart disease." (Clinical Review, 6).

104. Rho (D) Immune Globulin Intravenous (IGIV) (Human) - WinRho

In this March 1995 approval for treatment of chronic and acute immune thrombocytopenic purpura, the set of four clinical trials described in the FDA SBA included three small (n of 24, 24 and 63), open-label, single arm trials together with one trial in which 38 subjects were randomized to WinRho and others were randomized to prednisone with either high or low dose IGIV. There were no statistically significant differences found among treatment groups in any of the efficacy vari-

28 Also, while not affecting the quantum of efficacy evidence directly, it may be of interest to note that the drug was "manufactured" (made by!) genetically-altered cloned goats with the drug expressed in and purified from goat's milk. This is the first (and to date, only) FDA approved use of a cloned genetically-altered animal for drug production.

ables including response rate, peak platelet counts and times to achieve predefined platelet counts. In consultation with an FDA hematologist, such a quantum of evidence would not have been sufficient if this condition were prevalent and the number of subjects capable of being enrolled in trials was not a consideration.

106. Rifapentine - Priftin

This January 1998 approval to treat pulmonary tuberculosis "is based on the 6 month follow-up treatment outcome observed in the controlled trial as a surrogate for the 2 year follow-up accepted as evidence of efficacy in the treatment of pulmonary tuberculosis." (Label, 9). The primary endpoint for the single trial on which this drug was approved was "clinical equivalence on success rate to be no more than 10% worse than [the active approved control] rifampin with two-sided 95% confidence," which was met. (Statistical Review, July 27, 1997, 4). However, the statistical reviewer noted that "the two most important conclusions from this study are the following: 1. The cure rates are comparable between the rifampin (83%) and rifapentine (88%) arms [the primary endpoint].... and 2. There is a statistically [significant] difference between the arms in the chance of a relapse...the risk is 5% for rifampin ...and 11% for rifaptentine.... Rifapentine appears to be an effective drug in producing conversion to TB negative sputum...[but] [i]t is less effective than rifampin in preventing later relapse." (Statistical Review, 22-23). Furthermore, the medical reviewer noted that the CDC made a closed door presentation to the Advisory Committee which caused concern within the Committee over this drug's use in HIV positive patients because of "a study presented by the CDC where rifapentine resistance developed in the HIV-positive patients, and the potential for rifapentine to significantly reduce the AUC of the protease inhibitor, Indinavir." (Medical Review, June 19, 1998, 61).[29]

107. Rilonacept - Arcalyst

In this February 2008 approval for Cryopyrin-Associated Periodic Syndrome (CAPS), there was a single double-blind placebo-controlled study, but because of the rarity of this condition, FDA permitted there to be two segmented parts of the study, Parts A (n=47) and B (n=45), with separate randomizations for each part. Both Part A and Part B of the trial met their primary endpoints (p of less than 0.001 for each). Also, while the drug was designated as a Fast Track drug, it received a full approval without the need for a confirmatory Phase 4 study.

108. Riluzole - Rilutek

In this December 1995 approval for treating amyotrophic lateral sclerosis (ALS or Lou Gehrig's disease), the approval rested on the studies, both of which failed to hit their primary endpoints of time to tracheostomy or death according to the prespecified analysis in these placebo-controlled, randomized trials. The primary endpoint results by the prespecified analysis in these two trials were p values of 0.076 and 0.12. (Label, 2). In both cases, FDA salvaged each trial by permitting a post-hoc analysis that in each case yielded a p value of exactly 0.05 in each trial, not less than 0.05. (Label, 2). In addition, there had been one interim analysis in study 301 with an alpha "cost" of 0.001 so that the hypothesis was being treated to determine not if it were less than a p value of 0.05 but less than 0.049. It is also noteworthy that both trials had numerous secondary endpoints of muscle strength and neurological indices and not only did these not show any statistically significant separation between placebo and drug arms, there was hardly any numerical difference between the groups on these indices. Finally, in both studies, "there was no statistical significance in mortality at the end of the study." (Label, 2). The FDA medical reviewer notes that the apparent improvement in survival occurs early in each study period and the Kaplan-Meier curves came nearly together at the end of the study period, so that the FDA medical reviewer further observes: "Of course, the unanswered questions are whether the Kaplan-Meier curves eventually meet and follow a common path therefore [or] potentially the curves could cross with cumulative survival being worse on drug after 2-3 years." (Medical Review, 22 [Aug. 18, 1985]).

110. Rufinamide - Banzel

In this November 2008 approval for treating Lennox-Gastaut Syndrome, there was a single placebo-controlled randomized study (n=138) which was robustly statistically positive on all three co-primary endpoints of seizure activity (p values of 0.0015, <0.0001, 0.0041). However, Institutional Review Boards may not have found it unethical for a second study to be conducted. Therefore, the "statistically very persuasive finding" in this one trial may not have satisfied the strict application of FDA "single study" policy in its Evidence Guidance (See Section II.C.3). However, this sponsor also conducted 2 large studies of this drug in a prevalent disorder, "partial seizures," and while FDA did not find the efficacy evidence in these 2 "partial seizure" trials adequate to warrant the drug's approval for that prevalent indication, FDA found that the efficacy evidence in these 2 studies provided "additional support"[30] for the orphan indication as noted in the final sentence of the conclusionary paragraph by the medical reviewer on the efficacy evidence for the Lennox-Gastaut use: "The agent is additionally supported by the evidence from the partial seizure trials which indicate anticonvulsant activity." (Medical Review, Oct. 1, 2008, 77).[31]

111. Sacrosidase - Sucraid

29 Query: Does this approval mean that one positive non-inferiority trial versus an approved active control is the equivalent level of evidence as two positive placebo-controlled superiority trials? If so, would one adequate and well-controlled positive superiority trial over an approved active control also be considered the equivalent of two positive superiority trials versus placebo?

30 Query: Does this constitute "confirmatory evidence" under FDAMA 115?

31 The Evidence Guidance explains that studies in a closely related disease can essentially supply the "second" study necessary for approval and, coincidentally, FDA in this same 1998 Guidance cites an eerily nearly identical earlier precedent when FDA observes that: "The recent approval of lamotrigine for treatment of Lennox-Gastaut Syndrome (a rare, largely pediatric, seizure disorder) was based on a single adequate and well-controlled study, due in part to related data showing efficacy of the drug in partial-onset seizures in adults." (Evidence Guidance, 10). However, it is critical to note the difference between lamotrigine and rufinimide is that FDA viewed that lamotrigine had established proven efficacy in partial-onset seizures in adults.

In this April 1998 approval for treating congenital sucrase-isomaltase deficiency, there were two identically designed key trials: one of which was negative and one positive. However, FDA approval was based on a single double-blind, randomized, placebo-controlled, dose-response positive trial in face of a conflicting negative trial because the positive trial not only met its primary, but almost all of its secondary, endpoints which showed not only clinical improvements (e.g., fever, watery stools, more solid stools), but mechanistically showed that better results were observed in those who had higher enzyme (that is, drug) levels. In addition, there was a dose-response and subjects responded well to sucrose challenge. (Medical Review, Aug. 14, 1997, 82-84). However, the statistical reviewer concluded by recommending yet a third trial be conducted prior to approval. (Statistical Review, 19-20 [Sept. 15, 1997]).

112. Sapropterin Dihydrochloride - Kuvan

In this December 2007 approval for reducing blood phenylalanine (Phe) levels in patients with BH4-responsive phenylketonuria (PKU), there were four efficacy studies. The primary so-called "efficacy study" was a randomized, double-blind, placebo-controlled study (n=88) with a primary endpoint of mean change in Phe at week six ($p < 0.001$). The FDA medical review of December 7, 2007 concluded that this finding was both clinically meaningful and statistically significant, as well as noted important secondary endpoint results of clinically meaningful decreases in blood Phe levels at weeks one, two and four, which supported the primary endpoint finding. (Medical Review, 12). Other findings in a "Diet Study" and an "Extension Study" provided additional confirmatory evidence of efficacy. (Medical Review, 14). For instance, while the medical reviewer did not find the statistically significant primary endpoint results in Part II of the Diet Study to be clinically meaningful, the reviewer noted that a "secondary efficacy finding {in Part II of the Diet Study which was mean change in blood Phe from baseline to week 3} supports the primary efficacy finding of the Efficacy Study." (Medical Review, 13).

119. Sterile Talc Powder – Sclerosol

This December 1997 approval for treating malignant pleural effusions was based solely on published literature. The statistical review of January 5, 1996 notes that: "Talc has been used for years to treat patients with malignant pleural effussions, but talc has never been approved by the FDA for this purpose. It was felt that if approval were granted, there would be more control over the mechanism by which patients are treated with talc. For example, one concern is the asbestos which some talc contains." (Statistical Review, 2). In determining that substantial evidence of efficacy was provided in this NDA, FDA overcame concerns with both the quantum and quality of evidence as seen in the following comments about the published studies:

> Each study was sponsored by an investigator and there was no control body coordinating these research activities. Consequently, the studies use different study designs, different doses of talc, different routes of administration, different control groups, different definitions of response, and different lengths of follow-up. No CRFs are available, so it is impossible to determine exactly how the patients were treated and exactly how they responded. The quality of the safety data and prognostic factors for efficacy variables is then compromised.

(Statistical Review, 2). The statistician viewed five of the published studies as being of more reliable design and/or quality. Of these five studies, the statistical reviewer noted that in the intent to treat analysis, only one of these five had a statistically significant higher response rate in the talc group than in the control group. The other four of five studies had a statistically significant response rate in the evaluable population, but this analysis has a "potential bias" in that in three of the five studies the talc group, "was associated with a higher incidence of premature death than the control group." (Statistical Review, 10-14).

122. Tetrabenazine - Xenazine

This August 2008 approval for treating chorea associated with Huntington's disease relied upon one 12 week randomized, double-blind, placebo-controlled trial (n=84) and one five-day randomized, double-blind, placebo-controlled, staggered withdrawal study (n=30). In the larger efficacy trial, the primary endpoint of change from baseline in the Chorea Score (a subset of the Motor Assessment Scale of the Unified Huntington's Disease Rating Scale) for the average of weeks 9 and 12 was statistically significant ($p = 0.0001$); however, the primary endpoint of the smaller, staggered withdrawal study had only a trend suggestive of efficacy, but was not statistically significant for its primary endpoint.

123. Thalidomide - Thalomid

This July 1998 approval to treat erythema nodosum leprosum (ENL or leprosy) relied upon "primary data demonstrating the efficacy of thalidomide…[that] are from the published medical literature and from a retrospective study of 102 patients treated by the U.S. Public Health Service (PHS)." (Label, 7). With respect to the PHS study, the statistical review of August 7, 1997 stated:

> These [102] patients were treated from 1973 to 1997, which is a long period of time. Hence, the data generated from these medical records is of varying quality and completeness. No analytical protocol was available… no comparative drug or therapy was used, subjects were not randomized to treatment groups, and there is no fixed dose or duration of dose, no rules of titration up and down. This data set is of inferior quality as compared to the data from an adequate and well-controlled clinical trial…therefore the statistical analysis of this review will not contain any p values.

(Statistical Review, 1-2). Subsequently, the statistical reviewer stated that, "this data set is not from an adequate and well-controlled study." (Statistical Review, 36).

124. Tiopronin - Thiola

This August 1988 approval to prevent cystine nephrolithiasis in patients with homozygous cystinuria has an unusual regulatory history. The medical reviewer states,

> In 1979, the sponsor of this NDA was approached by FDA to consider obtaining an IND for Thiola and organizing a multiclinic trial with this drug.... The sponsor was advised that two other investigators had declined to undertake this task. A specific guideline for the preparation of the IND was provided to the sponsor [by the FDA]. A requirement of the inclusion of the placebo control group for the multiclinical trial was also deleted by the FDA, when potential co-investigators refused to conduct [a] randomized trial for bioethical reasons.... On December 5, 1985, the FDA invited the sponsor to submit a new drug application.

(Medical Review, July 25, 1988, 2-3). The medical officer concluded by finding efficacy on the basis of the sponsor's report of 57 patients treated with this drug, using each patient as his/her own control. (Medical Review, 24-25).

125. Tranexamic acid – Cyklokapron

In this December 1986 approval for treating hemophilia patients "to reduce or prevent hemorrhage and reduce the need for replacement therapy during and following tooth extraction," all the efficacy evidence came from 6 published literature studies that were all conducted "in the late sixties and early seventies" (that is, more than a decade and a half before the approval) and only one of these studies was placebo-controlled, randomized and double-blind, with two others open and retrospective and the remaining 3 uncontrolled. (Medical Review, Nov. 6, 1985, 18).

126. Treprostinil sodium - Remodulin

This May 2002 approval to treat pulmonary arterial hypertension was based on the results of two concurrently run, identically designed trials, both of which were double-blind, randomized, placebo-controlled with a primary endpoint of the 6 minute walk test of exercise capacity. The sponsor and FDA had agreed in advance that a positive result would be either: (a) both trials having a p value of < 0.05 on the primary endpoint; or (b) one trial having a p of <0.05 plus the pooled result having a p value of < 0.01. The primary endpoint results of each of the two trials were p values of 0.0607 and 0.0550, while the pooled result was 0.0064.

127. Trientine - Syprine

This November 1985 approval for treating Wilson's Disease was based on a summary of results obtained by two different investigators in a total of 41 subjects, in which there were no concurrent controls. Particularly, there was no placebo control as the FDA medical reviewer observed that: "The sponsor did not initiate and/or subsidize the [two] clinical trials reported herein. They were carried out independently by two recognized experts in the field. The sponsor was able to obtain the detailed records of the cases and to transfer the data to case report forms for inclusion in this NDA. Placebo-controlled studies were not done because they would be flagrantly unethical in this disease." (Medical Review, April 9, 1984, 2).

128. Trimetrexate Glucuronate – TMTX, Neutrexin

This December 1993 approval to treat pneumocystis carinii pneumonia (PCP) in AIDS was based on a single randomized, active-control (trimethoprim/sulfamethoxazole or TMP/SMX) trial. According to the medical review:

> The stated objective of this study was to attempt to show that TMTX was superior to TMP/SMX with respect to survival of the PCP episode, as assessed at day 56. Clearly the data do not support such conclusion [because the risk of death in the TMTX group was roughly twice that in the TMP/SMX group]. However, from the regulatory perspective, this was not the appropriate objective. From a scientific and regulatory perspective, the objective should have been to attempt to show that TMTX was 'equivalent' to the approved therapy, TMP/SMX. The treatment groups were equivalent with respect...to the percentage of successful responders [which] was 50% for each...group.

(Medical Review, Aug. 9, 1993, 38). Further, "The reasons for failing to respond to therapy were, however, different for the two treatments. TMTX patients were more likely to fail due to lack of efficacy, while TMP/SMX patients were more likely to be failures due to treatment limiting toxicity." (Medical Review, 38).

129. Vaccinia Immune Globulin (Human) Intravenous - VIGIV

This February 2005 approval to treat severe complications from the smallpox vaccine was based on two studies in healthy volunteers, and without any controlled studies showing benefits such as decreased mortality or severity of smallpox. One study was an open-label safety study in 33 healthy volunteers and the sole evidence of efficacy was an open-label study in 78 healthy volunteers in whom the sponsor showed serum neutralizing antibodies for vaccinia 5 days after drug, which "were not less than those expected following a similar dose of" an approved therapy. (Label, 6).

131. Vigabatrin - Sabril

This August 2009 approval for treating infantile spasms was based on "studies...that are principally derived from published reports." (Cross-discipline Team Leader Review, [July 20, 2009]). There were three controlled studies submitted: Study FR03, of which the cross-discipline leader stated, "would not normally meet the criteria as a pivotal trial" (Cross-discipline Leader Review, 11); Study 1A "does not meet the normal standards for the FDA for reasons described above (e.g. lack of a predefined protocol, interim statistical plan, questions regarding the completeness of the blinding...)...nevertheless the primary endpoint analysis would suggest a positive effect" (Cross-discipline Team Leader Review, 10); and Study WO19 whose prespecified primary endpoint was change in average spasm frequency as measured over a 2-hour window ($p = 0.562$). However, "this endpoint was generally considered inadequate by Dr. Sheridan as it provided a very small sampling of seizures and therefore was likely to result in a larger variance...this combined with the small size of the study was unlikely to provide adequate power to detect a treatment effect. One of the secondary endpoints in Study W019 included a 24-hour...observation window. When this is examined a large and statistically significant ($p = 0.03$) difference is observed with a 68.9% reduction in the vigabatrin group and a 17% [reduction] in the placebo group. Thus, while the primary end-

point of this study was negative, the endpoints, which were also not optimal, suggested an effect." (Cross-discipline Team Leader Review, 11).

132. von Willebrand Factor/Coagulation Factor VIII Complex - Wilate

In this December 2009 approval for treating "spontaneous and trauma-induced bleeding episodes in patients with severe von Willebrand disease", the results of four open-label, non-randomized, non-controlled trials in a total of 70 subjects were pooled for analysis (and several subjects participated in more than one of these trials, raising also issues of patient selection bias). It is observed that at the time of these trials there were two other FDA-approved therapies for this condition, Alphanate and Humate-P, and therefore, the possibility of a non-inferiority trial without exposing subjects to the risk of randomization to a placebo arm was a possibility. However, the FDA statistical review of this application stated that these "efficacy data of Wilate are considered as secondary and are derived from [4 studies] which were open-labeled and uncontrolled" and therefore a "PK study…is the pivotal study for the basis of the product approval." (Statistical Review, 16). This quantum of efficacy evidence, while entirely appropriate for this orphan condition, illustrates an FDA exercise of judgment in its review of therapies for rare conditions.

133. Zalcitabine - Hivid

In this June 1992 approval for treating MDS, FDA relied upon "2 small studies. The first was a Phase 1/2, open-label, dose-ranging study…the second study was a randomized Phase 2 study designed to evaluate the virologic and immunologic effects of the combined administration of two nucleoside analogues (zidovudine combined with either [zalcitabine] or didanosine.) Both studies used an experimental regimen of zidovudine…and neither was designed to assess the clinical efficacy of the combination." (Label, 3).

APPENDIX 2: SUBPART H AND FDAMA 115

In its May 1998 Evidence Guidance, FDA describes nine different circumstances in which a single trial may provide the statutorily-required effectiveness evidence. Often this guidance has been misread to mean that only the last of the nine circumstances represents a situation in which a "single" study may be adequate. The last circumstance is a situation in which a highly persuasive statistical finding (a p value of less than 0.01 and often even "more persuasive" than that) in a single trial with some other indicia of the study's reliability (e.g., multicenter with no center driving the results) out of a potpourri of possible factors that may provide such additional credibility to the primary endpoint finding and where it is likely unethical to conduct a second study.

However, it is critical to observe that FDA lays out eight other circumstances in this same guidance in which a single study may be adequate for meeting the statutory standard. However, of the other eight circumstances of "single study" approvals described in the May 1998 Guidance, only one is relevant to a new chemical entity. Therefore, for purposes of this analysis of orphan drugs approved as new chemical entities, there are only two circumstances for a single study approval applicable to a new chemical entity described in the May 1998 Evidence Guidance.

At the same time that FDA was developing what later became its May 1998 guidance, Congress was enacting an amendment to that 1962 effectiveness standard that created a new alternative statutory standard for establishing a drug's effectiveness. This new alternative statutory standard is: "one adequate and well-controlled study and confirmatory evidence." This provision of the law is referred to as FDAMA 115 (after the section in the FDAMA that inserted this statutory standard into the law.)

The nine types of circumstances that FDA described whereby a single study may be sufficient to prove a drug's treatment benefit had been based by FDA on its 36 years of collective experience and set forth in its May 1998 guidance. These nine types of circumstances can be seen as ways for implementing the FDAMA 115 "one adequate and well controlled study and confirmatory evidence" alternative statutory standard. In this way, the May 1998 Evidence Guidance and FDAMA 115 can be seen as fundamentally similar policies that were fortuitously issued almost simultaneously. One must, however, guard against a commonly-held misconception which is that the ninth of those nine circumstances in the May 1998 Guidance is the sole method for approving a drug based on a single trial. There are eight other circumstances described in the May 1998 Guidance itself. Moreover, the breadth of the FDAMA 115 "one adequate and well-controlled study and confirmatory evidence" statutory standard extends beyond these nine circumstances described in the May 1998 Guidance. For instance, Dr. Russell Katz of FDA at an FDA orphan drug conference in October 2010 presented the approval of tetrabenazine for Huntington's disease as an example of FDA employing the FDAMA 115 standard in approving this orphan drug.