Writing a Briefing Book for a CDER Advisory Committee

1. Introduction

The meeting of a CDER Advisory Committee is typically concerned with a reviewing Division's consideration of a Sponsor's pending NDA. Before the meeting, members of the Committee have received a Briefing Book ("**BB**") from the Division and one from the Sponsor (in this document, **"you"**). At the meeting, corresponding Presentations are made by the Division and the Sponsor, after which the Committee discusses the available information, guided by a set of Questions prepared by the Division. Often, the Committee formally votes on some or all of the Questions, including (usually) a Question that effectively asks the Committee to recommend (or not) that the NDA be approved.

Several years ago, many FDA Divisions did not prepare BBs or give Presentations. Instead, the Sponsor's BB was vetted by FDA until there was agreement as to the facts. Of course, the Sponsor always included a final Discussion section whose recommendations were the Sponsor's own.

Now, when there are two BBs covering much of the same ground, members of the Committee may — although they shouldn't — elect to read only one. Other things being equal, the FDA's BB is the one that they are more likely to read, because its overall spin is less predictable. FDA's reasoning is sometimes poor, but it may be foolishly favorable or foolishly unfavorable. Your reasoning, the Committee knows, is not likely to be foolishly unfavorable.

You'll have a leg up if — as sometimes happens — the FDA's BB is poorly written, poorly organized, and illogically argued. Don't count on it. If yours is the BB that is sloppily produced, difficult to navigate, evasive, or unscientific in style, then it is less likely to be read.

1.1. scope of this document

This is not about How to Run a Drug-Development Program, How to Write English Prose, or How to Make a Slide Presentation. We have opinions about those matters, but we have restrained ourselves.

1.2. a Briefing Book is not an ISE or ISS

Composing a Briefing Book is not like composing some combination of an Integrated Summary of Efficacy ("**ISE**") and an Integrated Summary of Safety ("**ISS**"). The ISE and ISS are non-public documents, exhaustively providing the results of all your studies. The only alternative treatments they mention are those that were used as controls

in the studies. They make no recommendations. FDA reviewers expect to spend hundreds of hours with the ISE, ISS, and associated study reports, replicating the analyses and drawing their own conclusions.

In contrast, your BB will be a public document, and it should be organized so as to require only a few hours of each Committee member's time. It can be much shorter than the ISE/ISS, because the Committee members are entitled to assume that your facts have been checked by FDA. It must provide context: Why is your treatment worth considering, vis-à-vis the alternatives? Are the observed benefits and adverse effects clinically important? And so on.

1.3. a Briefing Book is not a PI

Composing a Briefing Book is not like composing a Package Insert ("**PI**"). A product's PI is meant to serve as a set of instructions for using the product, with little space devoted to the data and reasoning that led to those instructions. Alternative therapies are rarely mentioned, while some commercial details (*e.g.*, NDC codes) are always present. The format of the PI is rigidly specified by regulations, and the wording of the PI is sometimes restricted by regulations or by precedent.

In contrast, the organization of your Briefing Book should be that of a (large) scientific paper, with sections of background, methods, results, and discussion (not necessarily so labeled). Why did you do what you did? What did you do? What did you find? What conclusions and recommendations do you draw from those results? Especially in the discussion section, your BB must provide the sort of understanding of your product that you would, away from any regulatory context, try to give to a junior colleague. Your recommendations in the BB (unlike the instructions given in the PI) must be explicitly connected to the data and reasoning that preceded them. Where there are gaps in your data (*e.g.*, demographic groups of interest that were either excluded by protocol, or included but not successfully recruited; human overdoses), what are reasonable expectations as to what would happen in those populations or situations? Some of these extrapolations may be of the sort that can be described in the PI, but others will not be. You are not here discussing them for the PI. You are discussing them to show that you understand your drug.

1.4. why is this meeting being held?

Sometimes, there is no disagreement as to what action should be taken — neither within the Division nor even between the Division and you — but the issues presented by your application are novel, and the Division wants to use the Advisory-Committee meeting to publicize its approach to those issues, before they arise in other applications.

At other times, the Division may be preparing to take an action that might be publicly misconstrued. At such times, the Division may look to the Advisory Committee for protective cover. If the contemplated action is a non-approval, then an Advisory-Committee meeting will be the Division's only legal channel for explaining its reasoning.

If the Division is uncertain what to do, it may actually be going to the Advisory Committee for . . . advice.

If you ask the Division why they are bringing your NDA to the Advisory Committee, they may tell you. **Figure out what the Division cares about, and focus your Briefing Book and your Presentation on those issues.** Your Executive Summary should make the meeting's purpose explicit; the pivotal issue should not make its first appearance deep in the Discussion section of the main text.

Most of your NDA's regulatory history — when you began development, when you met with the Division, and so on — will not be of interest to the Committee, and it would be extravagant to spend Presentation time on it. In the Briefing Book, it deserves to be exiled to an appendix, if it is described at all. But sometimes, regulatory history may help clarify the meeting's purpose. For example, the Division might have been satisfied with your efficacy data long ago, and your latest studies may have been designed and executed only to resolve some lingering safety questions.

2. Worst-Case Scenarios

Your idea of the worst-case scenario may not be the same as FDA's. A Committee recommendation of non-approval might be your worst-case scenario, but if the Division's reviewers had made the same recommendation, then FDA wouldn't be disappointed at all.

For FDA, the worst-case scenario is one in which a drug is recommended for approval and approved, but later developments cause the drug to be restricted or withdrawn. Episodes of this sort have led to storms of public and Congressional unrest that have been much more stressful to FDA than any of the mutterings heard after allegedly unjust non-approvals.

If your trials were not in some sense successful, you wouldn't be coming to the Advisory Committee. Now, you have your desired proved result, but your trials provided much more information than that simple success. Some of those data need to be explored in your BB and Presentation, with the goal of reducing the likelihood of an unhappy public surprise much later.

3. Style

3.1. precision

Precision requires data. If 2 of 92 patients experienced some event, the reportable incidence was 2%, not 2.1739%. Some entries in a given table may be more precise than others.

3.2. the Briefing Book

3.2.1. using Word

You're probably using Microsoft Word to prepare the BB. RRF has comments elsewhere[1] about using Word, and many of those comments arose from experience with Briefing Books. We here mention only the issues that seem to arise most frequently.

Spellchecker is your friend. You should care enough about the BB to chase down all of the wiggly red underlines. If you don't care about spelling, why should the Committee believe that you've proofread the numbers?

3.2.2. things most readers skip over

Your text will probably be rich with a variety of parenthetical material, including citations to a bibliography, statistical details, and so on. Any given one of these detours will be of interest to some readers, but surely not all. Consider

> Observational data demonstrate a continuous increase in mortality with increasing resting heart rates above 70 bpm, and accumulating interventional data demonstrate a reduction in cardiovascular events with heart-rate lowering (Greene et al, 2013; Pittaras et al, 2013; Castagno et al, 2012; Habal et al, 2012; Fosbøl et al, 2010; Ho et al, 2010; McAlister et al, 2009; Kolloch et al, 2008; Fox et al, 2007; Pocock et al, 2006; Diaz et al, 2005; Gullestad et al, 2005; Metra et al, 2005; Lechat et al, 2001). This effect was statistically significant ($p < 0.002$, chi-square with 2 degrees of freedom) in one study (Bogus, 2003), but weaker in others (and so on).

(adapted from a real example). Its parenthesized parts will be no more than speed bumps to most readers. Material of that kind is best relegated to footnotes:

> Observational data demonstrate a continuous increase in mortality with increasing resting heart rates above 70 bpm, and accumulating interventional data demonstrate a reduction in cardiovascular events with heart-rate lowering.[2] This effect was statistically significant[3] in one study,[4] but weaker in others.[5]

---

[1] See http://www.fenichel.net/pages/Professional/subpages/collaboration.htm [accessed 2015-01-25].

[2] Greene et al, 2013; Pittaras et al, 2013; Castagno et al, 2012; Habal et al, 2012; Fosbøl et al, 2010; Ho et al, 2010; McAlister et al, 2009; Kolloch et al, 2008; Fox et al, 2007; Pocock et al, 2006; Diaz et al, 2005; Gullestad et al, 2005; Metra et al, 2005; Lechat et al, 2001).

[3] $p < 0.002$, chi-square with 2 degrees of freedom.

[4] Bogus, 2003.

[5] And so on.

3.2.3. paper & electronic copies

Distribution to the Committee of electronic copies of your BB will be appreciated. Those copies are easy to carry (if laptops and tablets are already being carried), and they facilitate pursuit of internal cross-references. If you distribute Word copies (as opposed to PDF files), then readers will be able to add their own electronic annotations. Still, some members will prefer to read hard copies, so you should be sure to provide them, too. If you use color in the document, are the colors adequately distinguishable on your hard copies? An internal cross reference like

> these results are described in more detail in Appendix A *[a clickable cross-reference]*

is fine on the screen, but

> these results are described in more detail in Appendix A (page 97)

will be better for those using the paper copy.

3.3. the Presentation

The central purpose of the whole exercise is to convince the Committee that you understand your drug. Accordingly, don't choose your primary presenter to be a high-ranking Suit who was not intimately involved in this development program. You will be embarrassed if your primary presenter responds to every scientific question with a deer-in-the-headlights stare and a quick referral to someone more knowledgeable.

Before you start making slides, read Tufte.[6]

4. Common Minor Errors

4.1. comparable & similar

Any two things of the same dimensionality are *comparable*, but only things of the same dimensionality and approximately the same magnitude are *similar*.

4.2. sex & gender

Distinguish *sex* and *gender* in the same way that those words are distinguished in reputable medical journals. *Sex* (male/female) is biological and anatomic; sex is an explanatory variable in the epidemiology of HIV/AIDS because of sexual differences in the distributions of mucosa and skin. *Gender* (masculine/feminine) is social; gender is an explanatory variable in the epidemiology of HIV/AIDS because extreme promiscuity has been characteristic of some gay subcultures. Two people of the same sex may not be of

[6] Tufte ER, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 1983) is a good place to start.

the same gender. In almost all clinical areas, what you know about your subjects is their sex, not their gender.

4.3. scales

Some of your results may come from the Glasgow Coma Scale, the Hamilton Rating Scale for Depression, the Kansas City Cardiomyopathy Scale, the European Quality of Life Questionnaire 5D, or similar instruments. Not all members of the Committee will be familiar with the scales of your choice. In your text, in your tables, and on your slides, use of any such scale deserves some explanation:

- Is it better to have a high score or a low score?
- How much must two scores differ for the difference to be considered clinically significant?

4.4. that

Taken from a real example,

> It is of note that the baseline characteristics were different between the regions.

typifies the irritating use of introductory phrases that add nothing but length. ~~We believe that our experience has shown that~~ before writing a sentence that contains the word *that,* ask yourself whether all the words up to and including that *that* serve any useful purpose.

4.5. positive & negative

Some ambiguities can be easily avoided. In

> fraggeltine had a negative effect on cardiovascular outcomes,

did the writer mean

> fraggeltine had no evident effect on cardiovascular outcomes

or

> fraggeltine had an adverse effect on cardiovascular outcomes?

5. Your Disease

Demonstrate comprehensive knowledge of your chosen disease, but don't overstress its prevalence. Your drug will be approved (or not) because of its expected effect on **each patient** with the disease, whether those patients are few or many. The more you impress the Committee with how widespread your chosen disease is, the more likely they are to conclude that your trials should have been bigger.

Similarly, don't waste the Committee's time talking about your disease's impact on the U.S. economy. Their job is not economics.

If your chosen disease is rare, it may be useful for you to point out some of the difficulties of recruitment: your trials may have included some non-trivial percentage of all the eligible patients in your recruitment areas, so the idea of much bigger trials might be preemptively refuted. The downside of a rare disease in the Advisory-Committee setting is that Committee members may have no clear idea of how much of a burden the disease, and its available treatments, impose upon those who suffer from it. Don't hesitate to address this.

For example, the Cardio-Renal Advisory-Committee meeting of 9 August 2001 considered a treatment for pulmonary hypertension. The treatment (subcutaneous infusion of treprostinil) was not shown to provide any advantage over existing therapy with respect to exercise tolerance or survival; the statistical strength of the reported studies was marginal; and the infusions had been associated with infusion-site pain, sometimes controlled only with opiates. In recommending approval, the Committee seemed to have been swayed by hearing that patients vociferously preferred the infusion-site pain of subcutaneous treprostinil to the life-changing restrictions associated with the available alternative treatment (continuous intravenous epoprostenol).

Existing alternative treatments for your chosen disease are part of the context in which your NDA should be considered. Not-yet-approved treatments are part of this picture, to the extent that you know anything about them: Some treatments widely used elsewhere may not be approved in the US, but they may — undisclosed (of course) by FDA — be on their way to approval. Even treatments known to be inferior may be relevant.[7]

6. Fact and Opinion

The BB is a public document, but it is not a marketing document. In Briefing Books that we have seen, the most common infelicities have been inappropriate minglings of fact and opinion.

Again, the overall structure should be that of a normal scientific paper. **Don't let your opinions leak into your reporting of Results.** Even in small details, the sections describing your results should be free of conclusory words and phrases.

> BLEEP was a *large [emphasis added]*, randomized, double-blind, 832-patient trial comparing fraggeltine to placebo in the treatment of porcellosis.
>
> The efficacy and safety of fraggeltine have been evaluated in an *extensive [emphasis added]* clinical development program.
>
> A *comprehensive [emphasis added]* nonclinical safety testing program was performed.

[7] For example, there is ample evidence that warfarin is superior to aspirin for preventing stroke in patients with atrial fibrillation ("AF"). In presenting a new anticoagulant to be used instead of warfarin in AF patients, it is pertinent that many US physicians still use aspirin in these patients, in part because of the perceived difficulty of using warfarin. If your new anticoagulant is easier to use than warfarin, then it is pertinent that its availability might spur some of those physicians to get away from using aspirin.

are not good — in the first case, it's only your opinion that 832 patients were sufficient to qualify the trial as "large." On the other hand, there's nothing wrong with including **facts** that are to your credit:

> BLEEP was a randomized, double-blind, 832-patient trial comparing fraggeltine to placebo in the treatment of porcellosis. To our knowledge, BLEEP was the largest trial ever conducted in the treatment of this condition.

is fine.

Words and phrases like *large*, *extensive*, and *comprehensive* in the bad examples above are common in marketing-influenced briefing books, and it takes only a few of them to give the document an air of puffery. In the real examples from which those statements were adapted, the immediately-following text presented specific information that presumably led the authors to their self-congratulatory conclusions. While reading the BB sections devoted to Results, the reader should be free to draw his or her own conclusions. You'll get your chance.

Not all of the fact/opinion zoning violations are as subtle as the examples above. Everything in

> Fraggeltine did not affect fertility in male or female rats, but it caused cardiac teratogenicity in rats and reduced embryo-fetal survival in rabbits at exposures close to therapeutic doses. A small number of fetuses with ectrodactylia in rabbits were found at exposures 15 to 34 times higher than therapeutic doses. Fraggeltine also caused intrauterine or post-natal mortality in rats at exposures 22 to 66 times higher than therapeutic doses. In the proposed prescribing information, women will be advised not to become pregnant when taking fraggeltine. If a patient becomes pregnant while taking fraggeltine, she will be apprised of the potential hazard to the fetus. Fraggeltine should be used during pregnancy only if the potential benefit justifies the potential risk to the fetus.

(adapted from a real example) might be part of a proper briefing book, but not in the Results section. The opinion starting with "In the proposed prescribing information" should be somewhere else. Similarly, much of

> The evidence of benefit in porcellosis comes from a 9999-patient large, international, multi-center, randomized, double-blind, placebo-controlled, pivotal phase 3 outcomes study, BLEEP. BLEEP was the only fraggeltine study designed to evaluate outcomes in porcellosis. In view of the unmet medical need in this condition, the rigorous design and conduct of the study, and the significance ($p < 0.0001$) and robustness of results, BLEEP is sufficient as a single pivotal study to demonstrate efficacy.

(again, adapted from the Results section of a real example) comes across as misplaced special pleading.

7. Hearsay

Don't try to sway the Committee with opinions that are not even your own. If your drug is already on the market in 45 countries, then you may have useful data from post-marketing adverse-reaction reports. But the 45 approvals *per se* are not helpful; authorities don't like to hear arguments from authority.

8. Unreviewed Content

In general, the only data discussed in your Briefing Book and Presentation should be data that have been reviewed by the Division. If you have late-breaking results that you feel obligated to include, you should warn the Division before the meeting, and you should accompany every mention of any such results with notice that those results have not been reviewed.

New presentations of old data need no special permission.

9. Results

You must, of course, first describe the results of your prespecified hypothesis tests.[8] But then, you need to explore some of the other information that is buried in your data.

Post hoc data exploration has a bad reputation, because it can provide tempting false positives,[9] and sometimes these are grasped at in an attempt to resuscitate an adverse result:

> Fraggeltine was ineffective in the overall trial population, but it seemed to work well in women over 60.[10]

On the other hand, your successful result in your protocol-specified population doesn't mean that every patient randomized to your treatment benefited. Probably some were harmed, but your trial was successful because more were helped. If those two subpopulations can (in hindsight) be convincingly distinguished by baseline characteristics, then you may be able to refine the target population to make a good drug even better.

9.1. avoiding false negatives

9.1.1. due diligence

Most exploratory results are negative

---

[8] Probably at least one of those was positive — otherwise, you probably wouldn't be preparing to face the Advisory Committee. But there can be exceptions, when non-primary findings are statistically compelling and clinically important (for example, see http://www.ncbi.nlm.nih.gov/pubmed/10027498 [accessed 2015-01-20] for the story of carvedilol).

[9] See Wittes J, On looking at subgroups, *Circulation* **119:** 912-915 (2009).

[10] Hearing a claim like this, FDA and the Committee may be reminded of the two PRAISE trials of amlodipine. In the first of these, the overall prespecified result was neutral, but there was a strongly positive result ($p = 0.004$) in a clinically-meaningful subgroup. In the second study, limited to that subgroup, amlodipine provided no benefit. (See Pfeffer MA and Skali H, PRAISE and criticism, *JCHF* **1:** 315-317 (2013) for the overall story and the primary references.)

The observed effect was independent of baseline blood pressure.

The data do not suggest any need for dosing adjustment in patients with compromised renal function.

Efficacy was similar in men and women.

The data do not suggest any need for dosing adjustment in older patients.

so your problem is usually more that of false negatives than false positives. When you make statements like those just quoted, you are asserting that you have exercised due diligence. That is, you have made a competent, good-faith effort to find relationships between drug effect and blood pressure, renal function, sex, or age, and no such relationships appeared.

9.1.2. interval variables treated categorically

Some Briefing Books do not provide good evidence of due diligence. For example, suppose that your drug was studied in crossover trials. Then every patient was his or her own control, and the data can be presented easily. Suppose further that you wanted to show that drug effect was unrelated to baseline blood pressure. You might present one view of the data in a two-line forest plot:

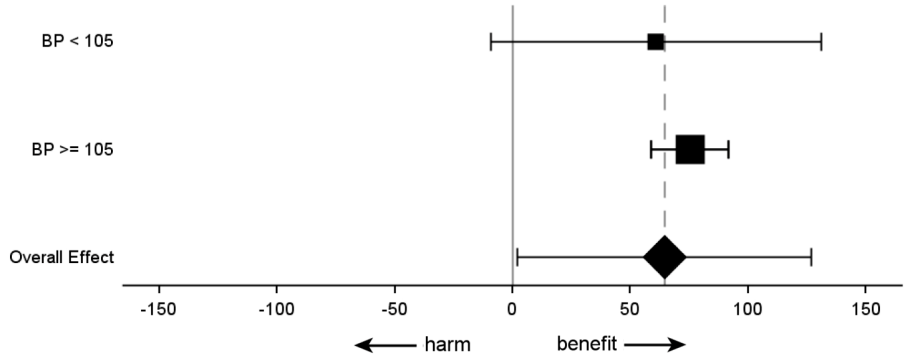

Figure 1, a successful trial

Plots like this are commonly produced, especially for consideration of the effects of age, renal dysfunction, or (as here) blood pressure. Here, the subsets separated by the arbitrary BP = 105 cutpoint look pretty similar, and the reader might conclude that all is well.

Figure 1 **is a typical failure of due diligence.** The results in the 400 simulated patients behind the above figure were

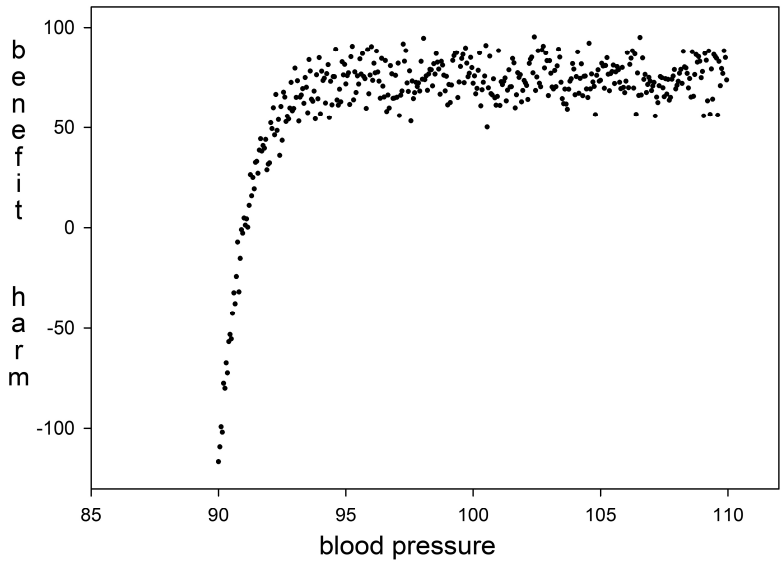

Figure 2 (same data as previous figure)

The scattergram is more informative than the two-group forest plot because the scattergram makes use of all the information. When — as in Figure 1 — one uses a cutpoint to separate a continuous range into just two sets, one is discarding information. True, the adverse results in patients with low baseline blood pressures, seen clearly in the scattergram, might have been evident if the forest plot had used a different cutpoint:

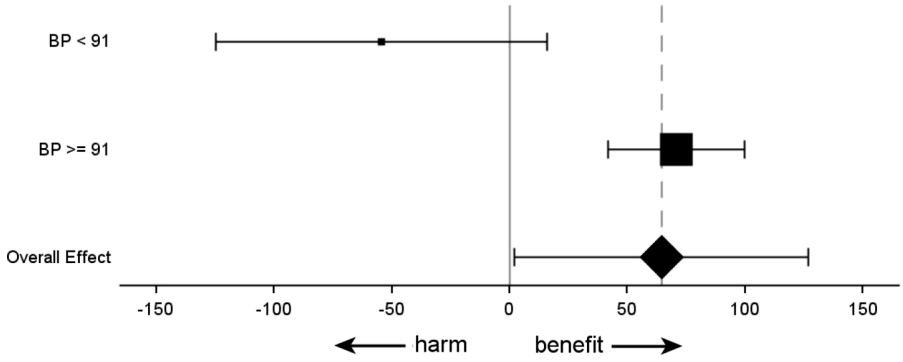

Figure 3 (same data as previous figures)

but many readers would dismiss the message of Figure 3 as a false positive, noting that the outlier group consisted of just 20 patients. Scattergrams are not always feasible, but they are not the only way to avoid loss of information. For example, here are the same data again, now presented using a forest plot with smaller bins:

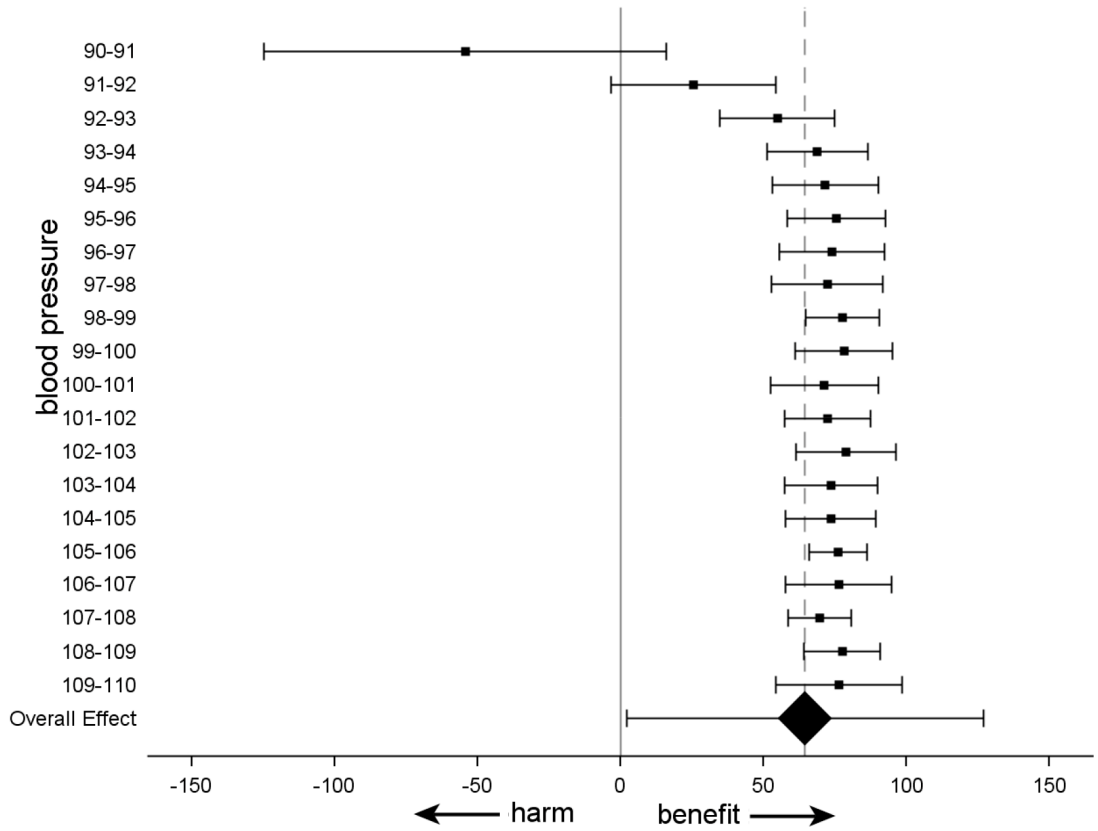

Figure 4 (same data as previous figures)

The increased sensitivity of Figure 4 is not bought at the price of greatly reduced specificity. If, for example, one of the **inside** bars of Figure 4 had stood apart from the others, but its neighbors were unchanged, readers would not hesitate to conclude that that result was spurious.

9.1.3. use of cumulants

The imposition of arbitrary cutpoints is not the only common way to lose information from interval data. Given data comprising a set of (x,y) pairs, one sometimes sees displays in which what is plotted against a given x_i is not y_i, but rather the average of all y values corresponding to x values greater than or equal to x_i. At the low end of such a display, any effects present only at that end of the x range can be swamped into insignificance. From the same dataset used above, a characteristically misleading graph of that kind is

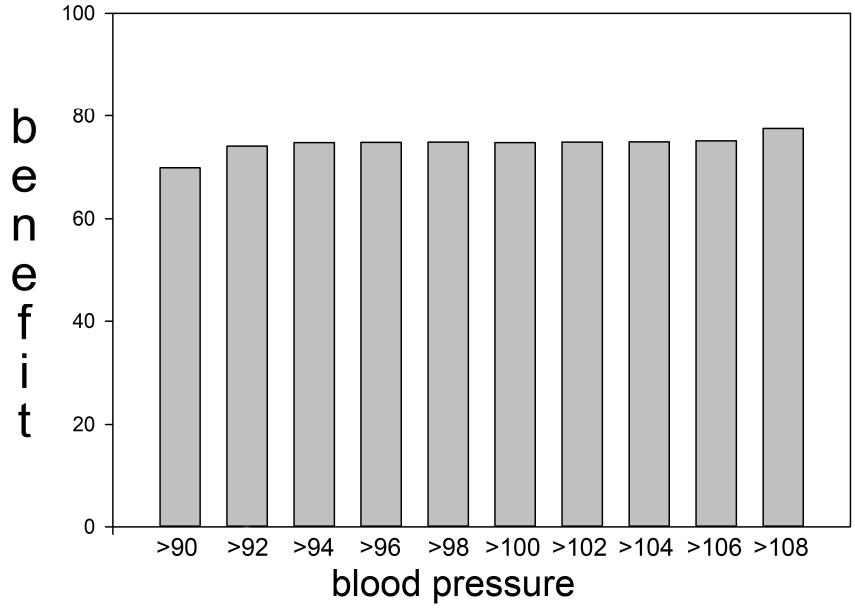

Figure 5 (same data as previous figures)

9.1.4. categorical variables

One way to describe the error leading to Figure 1 was that an interval variable (blood pressure) was treated as if it were categorical.[11,12] Categorical variables are always problematic: if the two data lines of Figure 3 had been labeled "male" and "female," then the reader might again wonder whether the apparent difference was a true or false positive, but no fine-grained analogy to Figure 4 would have been possible.

---

[11] Blood pressure is a continuous variable in nature, but by the time you get the data, it has been quantized into an interval variable. Also, the scales called "categorical" in pharma circles are called "nominal" in the wider literature. See http://www.ats.ucla.edu/stat/mult_pkg/whatstat/nominal_ordinal_interval.htm [accessed 2015-01-20].

[12] FDA encourages this sort of categorization by, for example, expecting patients over 65 to live together in their own category. We don't.

Sex is irretrievably categorical, but some subsets often treated as categorical can be examined on ordinal or interval scales. For example, inter-national differences in drug effect shouldn't be said to be patternless until they are seen to be patternless when the countries are arranged in per-capita income order, and again in order by pertinent demographics (certainly patient weight, maybe other, drug-specific factors).

When a categorical cofactor is expected to be noncontributory, the results in different categories shouldn't be expected to be identical. They should, however, have ordinary random properties. For example, you should expect smaller categories to vary more from the pooled result than bigger ones; if there are enough categories, a graph of per-category result against size should provide a typical funnel plot.[13] When such a plot does not have the expected taper, a true-positive interaction may have been revealed. For example, in the MERIT trial of metoprolol in systolic heart failure,[14] the overall relative risk of mortality in the active-treatment group was 0.66, but it was 1.05 in the US subpopulation. Some reviewers of these data thought that the finding in the US patients was a false positive,[15] but others (including FDA) were persuaded by a funnel plot that, as soon as the US was included, stopped looking like a funnel:

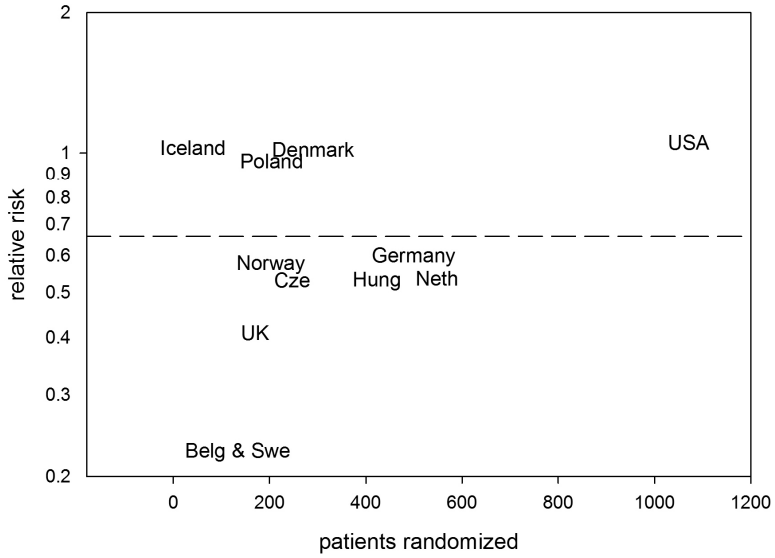

Figure 6, the MERIT trial

9.1.5. what if you find something?

Your *post hoc* discovery and description of an apparent relationship between a drug effect and a cofactor does not obligate you to believe or assert that the relationship is real. It does obligate you to discuss what you make of the finding. If the Division or a

[13] For funnel plots, see http://handbook.cochrane.org/chapter_10/10_4_1_funnel_plots.htm [accessed 2015-01-20].

[14] Effect of metoprolol CR/XL in chronic heart failure, *Lancet* **353:** 2001-2007 (1999).

[15] DeMets DL, Statistical issues in interpreting clinical trials, *J Int Med* **255:** 529–537 (2004).

Committee member finds a relationship you missed, and if they believe the relationship is credible, then your *post post hoc* improvised refutation may not be effective.

9.2. avoiding false positives

9.2.1. statistical false positives

Statistical false positives are the false positives that your mother warned you about. If you draw the target around the bullet hole (that is, if you claim victory on the basis of a non-prespecified endpoint), then your victory is likely a false positive. If you enter a pistol competition with a shotgun (that is, if you unfairly give yourself many different ways to win), then your victory is likely a false positive. Janet Wittes' paper[16] goes over this ground.

9.2.2. biological false positives

Statistical false positives arise because if one looks at enough **uncorrelated** measurements of biological variables, it is inevitable that one of them will be beyond the end of its typical range. Conversely, biological false positives arise when an inactive variable (say, sex) is wrongly given credit for the achievements of a **correlated** variable (say, weight). If you must include the uncorrected analyses — because you (foolishly) committed yourself to them in your Statistical Analysis Plan, or because FDA (even more foolishly) required them — then don't fail to have the corrected analyses nearby.

An observed effect may be statistically related to several different factors, but the contributions of some factors may be clinically negligible.

9.2.2.1. apparent sex effects are often really the effects of weight

Don't ever say something like

> The observed effect tended to be greater (or smaller, or about the same) in women.

Instead, say

> After adjustment for weight, the observed effect tended to be . . . .

[16] See footnote 9.

9.2.2.2. apparent age effects are often really the effects of renal function

Similarly, don't say

> The observed effect tended to be greater (or smaller, or about the same) in elderly patients.

Instead, say

> After adjustment for renal function, the observed effect tended to be . . . .

9.2.2.3. apparent effects of country or ethnicity are often biological false positives[17]

And don't say

> The observed effect tended to be greater (or smaller, or about the same) in <ethnic or national designator>.

Instead, say

> After adjustment for weight and renal function, the observed effect tended to be . . . .

9.3. data you want but may not have

9.3.1. patients admissible per protocol but not recruited

It's probably easy to identify patient groups that were not excluded by protocol, but are not represented in your database. For example, you probably don't have any Inuits. Some omissions might derive their importance only from political correctness, or because FDA requires attention to them (we partially repeat ourselves), but others might be important if patients from those groups (women, the elderly, African-Americans,[18] poor

---

[17] In one dataset (Walpole SC, Prieto-Merino D, Edwards P, *et al.*, The weight of nations, *BMC Public Health* **12:** 439-444 (2012)), the mean adult weight worldwide was 62 kg, but it was 82 kg in the US. Only a few Pacific-Island populations had higher mean body weight than the US.

In some recent datasets (unfortunately, all proprietary), there appear to be unexplained, similarly large, international differences in glomerular filtration rate, at least as estimated by conventional formulas.

[18] African-Americans are not just blacks by another name. The ancestors of African-Americans came overwhelmingly from West Africa, so they are genetically different from East Africans and others. If you have reason to wonder about an effect often seen in African-Americans (*e.g.*, an increased liability to drug-induced angioedema), then data from a population of European blacks (likely to be mainly of East African origin) may not be helpful.

Janet Wittes says

> You probably won't believe this, but how could I make it up? I was reviewing interim data from a large Pharma company that was doing a trial in sub-Saharan Africa, including South Africa. The racial groups were "White," "African-American," and "Asian." I said, "these aren't 'African-American'; they are AFRICAN!" And, I added

CYP2D6 metabolizers) are known to be different in their responses to therapies similar to yours.

Every important omitted population should be discussed in the Briefing Book, possibly drawing on experience in other indications or even experience with related drugs. If your drug is thought to have a specific mode of action (say, inhibition of an enzyme E), then data about the phenotypic variation of E in the population(s) of interest might be helpful.

9.3.2. patients excluded by protocol but likely to be given your drug

Your drug will sometimes be given to patients who shouldn't receive it at all. For example, your treatment for supraventricular tachycardia will mistakenly be given to some patients who have ventricular tachycardia. If that ever happened during your development program, then

- you **know** it will happen after the drug is launched, **whatever you do** in labeling and risk management; and
- the Committee will want to know what happened after those mis-administrations. Did your drug seem to be beneficial, neutral, or harmful?

Similarly, suppose you successfully excluded patients with blood pressure less than 90 from your development program. It is reasonable to expect that a few patients with pressures in the 80s will sooner or later receive it. You are not going to promote such use, but how worried are you about its (inevitable) occurrence? Was your safety margin still pretty good as your patients' pressures approached 90 from above, or was it mediocre, and sinking fast with declining pressure?

9.3.3. treatment of overdose

Overdoses happen. Sometimes they are the results of errors by prescribers, pharmacists, nurses, or patients; sometimes they are the results of untoward drug-drug interactions; and sometimes they are suicidally deliberate. You are unlikely to have randomized data with respect to any of these scenarios, but again, the Briefing Book is your opportunity to reassure the Committee and the FDA that surprises are unlikely. For example, is the drug dialyzable?[19] Is activated charcoal likely to be useful in an oral

---

> that the "Asians" were basically all from India, because the only country that had Asians in the study was South Africa. I was told that the company couldn't change the racial groups because that's how they were programmed, and it would require a major change to modify the programs.

Don't let yourself be victimized by Dilbert's Mordac, the Preventer of Information Services.

[19] If you don't now know whether your drug is dialyzable, you could know from *in vitro* data by the

overdose, or is absorption of the drug so rapid, and enterohepatic circulation so minor, that activated charcoal won't be useful? Are there data to support the use of other general-purpose antidotes?[20]

9.4. data you got but didn't want

9.4.1. outliers

In some of your tables and (especially) some of your figures, the Committee will see that a few subjects did much better or much worse than all the others. Either way, discussing these subjects is part of demonstrating that you understand your drug. Some outlier data may just come from data-entry errors that somehow weren't caught before the data lock. If you think the data are real, you must be prepared to discuss them a little more. Did the outlier subjects have any distinguishing baseline characteristics? Each one of the outliers may deserve a paragraph or so of description, stored away in an appendix but accessible by a reference — in the electronic version of the BB, by an internal link — from the subject's first anomalous appearance.

9.4.2. protocol violators

Suppose that your protocol excluded patients with creatinine clearance less than 60 mL/min, but a few randomized patients turned out to have baseline clearances in the 45 mL/min range. That doesn't make your investigators look good, but it may — depending on what happened to those patients — be good for you. If those patients did well, then you may still want to recommend that the drug not be given to patients with clearances < 60, but you might no longer feel obligated to go to great risk-management lengths about it. If they did poorly, then your instructions will ultimately need to be more careful, but you're still better off with the information: Now you understand the drug better.

9.5. dosing

Even if you studied only one dosing regimen, your patients received a wide range of drug exposure. The drug concentration in the bodily compartment(s) of interest was probably higher in patients who were smaller, or who had impaired metabolic or renal function, and it was probably lower in the bigger patients with better-functioning metabolism and excretion. The fact that your overall results were favorable doesn't prove that all of your patients were correctly dosed, any more than Figure 1 proved that the underlying facts were not those revealed in Figure 2.

---

end of next week. You really have no excuse not to know this.

[20] For example, if your drug tends to induce bradycardia, responders are likely to administer atropine. In dogs or pigs given overdoses of your drug, does atropine work? Or, does your drug seem to interfere with the mechanism through which atropine is antidotal in other overdose-induced bradycardias?

FDA believes in pharmacokinetics. You may have pharmacokinetic data from your Phase 3 patients, but even if you don't, you probably have enough Phase 2 pharmacokinetics, Phase 3 demographics, and Phase 3 laboratory values to predict the C_{max} and AUC of your drug in each Phase 3 patient. Show your work. Were efficacy and safety preserved across your range of drug exposures, or do the forest plots look like Figure 4?

9.6. adverse events

9.6.1. standard vocabularies

Standard vocabularies (*e.g.*, MedDRA) for the description of clinical events are valuable, but some standard terms are ambiguous, and there may be international variation in the way in which the putative standards are applied. These ambiguities can be preemptively resolved during the design of a trial's Case Report Forms, but if that was not done, then you'll need to do some repair work for the Briefing Book.

9.6.1.1. dizziness

For example, *dizziness* is a standard term in some vocabularies. Patients complain frequently of "dizziness," sometimes referring to vertigo (the false sensation of motion) and sometimes referring to lightheadedness (the sensation of impending faint). These sensations are different, and patients can usually distinguish them if challenged. Vertigo (reflecting ototoxicity) and lightheadedness (reflecting hypotension, hypoglycemia, transient arrhythmias, or other difficulties) are different AEs. Third-year medical students know that a patient complaining of "dizziness" must be questioned further. If you somehow forgot about that when designing your CRFs, and if you have described more than a few treatment-emergent events as "dizziness," then you are obligated to explain why you think these events were vertigo, or why you think they were lightheadedness, or why you think they were some of each.

9.6.1.2. anemia

Anemia can easily be defined in terms of hematocrit, hemoglobin, or need for transfusion, but as a drug effect it needs more description. Did you do anything to convince yourself that the anemia was non-progressive, like the minimal anemia associated with ACE inhibitors? Why were your patients anemic (?marrow suppression ?hemodilution ?bleeding ?hemolysis ?renal effects)?

9.6.1.3. kidney problems

In the same vein, some standard vocabularies include one or more of *abnormal renal function, renal dysfunction,* and *renal failure.* Experience has shown that these terms are often applied haphazardly. In some trials, investigators have implausibly reported a

higher incidence of “renal failure” than of “renal dysfunction.” If you have actual lab data, those data should be the center of any discussion of renal AEs.

9.6.2. presentation sequence

Some BBs begin their discussion of adverse events by going methodically through the MedDRA dictionary, filling pages with tables most of whose entries are zero. This is a way to lose readers. Start with what the readers probably want to hear about, using the ordering of the next five sections.

9.6.2.1. deaths, dropouts, and unscheduled hospitalizations

Every death, most dropouts, and every unscheduled hospitalization arose after some more specific AE, but you must first consider deaths, dropouts, and unscheduled hospitalizations as lumped entities. At the very least, patients probably thought that these events were the most important ones.

There may be very few events of this sort, so the only possible accounting may be a series of individual clinical anecdotes. If there were enough events, then they need the usual exploration: Were there differences in mortality between your treatment groups? Were there differences in dropout rates? Were there baseline characteristics that predicted unscheduled hospitalization in patients randomized to certain treatments? Do these baseline characteristics make any biological sense? And so on.

9.6.2.2. AEs inevitable from the drug’s mechanism of action

These events (bleeding from antiplatelet agents, headache from organic nitrates, and so on) are indeed adverse, but you’d be worse off without them. Their absence would raise readers’ suspicions that you didn’t understand the drug at all, or that you had systematically underdosed it, perhaps achieving less efficacy than you should have. The Committee may be interested to see how the incidence of these AEs associated with your drug compares to their incidence in association with other drugs in the same class; when you provide such data, refrain from conclusory comments.

9.6.2.3. AEs of pre-Phase 3 concern, but not thought to be inevitable

Some of these AEs will be ones for which there were worrisome but flimsy signals in early trials (hepatotoxicity commonly shows up here); others will be AEs that are established for other drugs in the class, but are of uncertain incidence with your drug. For example, because aplastic anemia was by then an established adverse effect of ticlopidine, aplastic anemia was of special interest in discussion of later thienopyridines.

With luck, you will use most or all of this section to describe dogs that did not bark in the night.

9.6.2.4. observed off-target AEs

Here is where you discuss adverse events that you didn't anticipate, but that happened anyway. For example, this is where the original sponsors of ticlopidine might have discussed aplastic anemia.

Events in this category should be grouped by putative mechanisms of action, even if you don't understand how your drug could trigger those mechanisms. The groups need not be mutually exclusive. For example, observed instances of lightheadedness might be grouped with events suggestive of arrhythmia, and then again with events suggestive of vasodilation.

9.6.2.5. the slush pile

This is the catchall, for all those events that you think are background noise. For reasons similar to those of Section 3.2.2, these data should be exiled to an appendix.

10. Discussion

This is the most important section of your Briefing Book. Everything before it, after all, should be only subtly different from the corresponding sections of the FDA's Briefing Book. Here you are on your own.

Matters that took up many pages in the Results section may, if they are truly non-contentious, be here dispatched in a paragraph or two. If the data support

> In the studies described in Section 6.3 (pages 53-78) above, fraggeltine was robustly shown to be an active antihypertensive. Except in poor CYP2D6 metabolizers, but independent of other demographic variables and independent of baseline blood pressure, the studied regimens of fraggeltine were consistently associated with blood-pressure reductions of 6-8 mmHg.

then this might be your whole discussion of efficacy, assuming that what was seen in the outlier CYP2D6 group was not an efficacy problem. Similarly, the Results section may have provided the necessary foundation for disposing of hepatotoxicity with

> Hepatotoxicity was of special concern, primarily because of the known hepatotoxicity of wiggeltine, a chemically similar compound. In aggressive preclinical and clinical studies,[21] the hepatic effects of fraggeltine were found to be indistinguishable from those of placebo.

[21] Section 4.2 (pages 32-37) and Section 7.5 (pages 84-89), respectively.

These examples illustrate some of the difference between this section and the Results section. Words like *robustly*, *consistently* and *aggressive* would have been out of place there; here they're just right.

Not every issue will be as easily taken care of. To begin the discussion of a truly contentious issue, describe the problem impartially:

> Almost 30% of the patients in the BLEEP trial dropped out of the trial before their planned final evaluation. Although the dropout rates were similar in the various treatment groups, it is appropriate to wonder whether fraggeltine might be efficacious in only a subset of the target population.[22]
>
> Only 832 patients have been exposed to multiple doses of fraggeltine in clinical trials. This is a smaller population than that usually seen in the development programs of drugs to be used for chronic diseases.
>
> Although fraggeltine convincingly reduced overall mortality in the BLEEP trial, it appeared to have only equivocal effects on any of the signs and symptoms of porcellosis. That is, the mechanism of action thought to have been established in preclinical trials does not appear to have been confirmed.

It's hard to give generic advice that's applicable after that. For example, some issues lend themselves to discussion in terms of sensitivity analysis (What is the best-case possibility for the patients who were lost to follow-up? What is the worst-case possibility?). In other cases, the central issue will be benefit/risk trade-offs.[23]

Your discussion should not be limited to the interpretations that you favor. If there are different, adverse interpretations that cross your mind, you can be certain that they've crossed the minds of FDA staff or Committee members. Spell them out here, with such rebuttal as you can muster. Pay particular attention to interpretations that you expect FDA to favor in its BB and Presentation. Don't let the Committee conclude, when one of those interpretations is brought forward, that you hadn't thought of it, or, worse, that you had, but were hoping that they had not.

10.1. proposed Package Insert

Your desired approval is, after all, the approval of a Package Insert, so it might seem to be obvious that the Briefing Book should include, and defend in detail, your proposed Package Insert. Certainly many Briefing Books do this, but we don't recommend it. You should be able to cover the substantive issues behind the PI — but not its wordsmithing — elsewhere in the Briefing Book.

After the Advisory-Committee meeting, if FDA later decides to approve your NDA, then the Division will meet with you to thrash out the Package Insert. The Division will, in general, take as much language as possible from the PIs of related drugs, to reduce the

---

[22] Dropouts and missing data can be showstoppers. If you think that you might have a problem in these areas, you may need specialist statistical help.

[23] For example, that between the benefit (moderate, non-obtunding analgesia) and the risk (occasional irreversible liver failure, even at therapeutic doses) of acetaminophen.

chance that you (or competitors) will try to impute meaning to linguistic variations that are actually synonymous. Committee members are rarely familiar with much of this preexisting art, and they are rarely acquainted with the many other labeling conventions. Asking the Committee to write or comment on PI language is like asking them to team together to write a sonnet.

The final approved PI **will be different** from any version that might have appeared in your Briefing Book. The difference may show, to your potential embarrassment, that FDA thought that you were over-reaching.

10.2. risk management

A portion of your Discussion section may need to be devoted to risk management.

The reviewing Division to which your NDA was submitted tries to consider both efficacy and safety. In contrast, the Office of Drug Safety — which will oversee your risk-management program — looks only at risks. This incentive structure is like that of the Department of Homeland Security, and it can lead to analogous security theater. You may need to recruit the Advisory Committee to bring reason to the process.

A good way to engage the Committee is to discuss several different alternative schemes of risk management, not just your favorite one. You should include — and criticize — some schemes that you think would be inappropriate. You may be able to preemptively establish consensus as to the range of schemes from which reasonable people should choose.

10.2.1. disproven risks

As described above in Section 9.6.2.3, you may have gone into Phase 3 worrying about certain drug-related adverse events. If you believe that you've now proven that those events don't happen, don't offer to waste time looking for them in some scheme of nonrandomized surveillance.

10.2.2. symptoms vs. *irreversible harm*

The proper goal of risk management is to prevent **irreversible harm**. Don't offer to waste time trying to keep people from having headaches that stop when they discontinue taking your drug.

10.2.3. detection vs. *prevention*

The proper goal of risk management is to **prevent** irreversible harm. Patients receiving clozapine undergo surveillance because it's been proven that if the drug is discontinued upon the appearance of early signs of hematologic toxicity, then progression to agranulocytosis is less likely.

Unfortunately, surveillance is not always this fruitful. For example, moderate overdoses of acetaminophen are sometimes without sequelae, but they sometimes cause liver failure. Antidotal treatment should be started before any change in liver function tests, because it's been proven that by the time the LFTs are abnormal, treatment is much less likely to be effective. Similarly, the connection between smoking and lung cancer is clear, but monitoring with serial chest x-rays is known to provide no benefit.[24]

In other words: If your drug occasionally causes irreversible harm, surveillance is appropriate only if it might plausibly lead to reductions in that harm. Detecting harm is not good enough.

10.2.4. evaluation

Part of your risk-management plan should be a means of ongoing evaluation, perhaps justifying relaxation to a less onerous scheme, or perhaps leading to escalation to a scheme that is more rigorous. The results of any evaluation will be most credible if the evaluation scheme is spelled out before approval.

You might be able to compare two or more risk-management schemes for your drug: If it is being launched in several different countries, then you might be able to evaluate the relative effectiveness of the several (probably different) risk-management schemes.

Some goals of your risk-management scheme may be similar to those used with unrelated drugs. For example, if you propose program P1 to keep your drug from being given to pregnant women, you might be able to find that a program similar to P1 is used by the sponsor of some other drug D1, and that a more aggressive program (P2) is used with respect to another sponsor's drug D2. Has there been any difference in the reported rates of fetal exposure? If not, that's an argument against the hassle of P2.

10.3. future work

You probably have some ideas about other indications, other doses, other populations, and so on. You should expect the Committee to speculate along similar lines, and to want to hear about some of your plans. Much of what you might say under this heading will be implicit in earlier parts of the Briefing Book and in the FDA reviews that will be publicly available if your application is approved.

In particular, go back to all the gaps (missing populations, iffy doses, AEs of uncertain causality) that you blustered past in your earlier sections. Here is where you should describe ongoing studies. At least briefly, you should also describe future studies, including those planned and — if you can — others under consideration.

---

[24] See Eddy DM, Screening for lung cancer, *Ann Int Med* **111:** 232-237 (1989). More recently, surveillance using axial tomography has turned out to be lifesaving (see Aberle DR, Adams AM, Berg CD**,** Reduced lung-cancer mortality with low-dose computed tomographic screening, *N Engl J Med* **365:** 395-409 (2011)).

Other possible studies may be bad ideas that you considered, but have rejected, perhaps because they would be impractical or, even if completed, inconclusive. If a Committee member proposes one of these designs, you'll be much better off if you can demonstrate — with detailed backup slides, if not a paragraph in the BB — that you thought along similar lines, but then rejected the idea for plausible reasons.

With respect to some of these matters, you might genuinely be undecided, or you might for commercial reasons not be ready to make certain decisions public. Even the most non-committal discussion can demonstrate to the Committee that your thinking is not purely commercial.

11. Questions for the Committee

Writing Questions for the Committee is difficult. If (the extreme case) there's only one Question:

> Should fraggeltine be approved for the treatment of porcellosis?

then Committee members are effectively invited to describe their various lines of reasoning in long, discursive remarks. Without biasing the decision-making, a good set of Questions breaks the necessary decision-making into a sequence of simple steps. Good Questions can function like a judge's impartial instructions to a jury.

It's perfectly appropriate for you to suggest, in discussion with the Division, the tenor, and even the wording, of the Questions that will be put to the Committee. Few Sponsors do this; only certain Divisions are receptive to such suggestions; and some Sponsors propose Questions like

> Is fraggeltine the best thing since sliced bread, or what?

If you can get out of that mindset, your suggestions may be helpful to the Division and to yourself. It's worth a try.

The best Questions allow Committee members to explain their thinking without having to come up with their own impromptu explanations. Such Questions are usually, therefore, a longish sequence of focused, quickly-dispatched forced choices. For example, the Cardio-Renal Advisory-Committee meeting of 24 October 1997 discussed clopidogrel, which had been compared to aspirin in the CAPRIE trial. Approval of clopidogrel hinged on its comparison to placebo, which had never been directly studied. There were fifteen Questions, the first of which was

> In the overall CAPRIE population, clopidogrel appeared to be superior to aspirin. This apparent finding is
>
> (a) not meaningful, because of heterogeneity among subgroups of patients.
>
> (b) probably attributable to the play of chance.
>
> (c) a plausible finding, but weaker than that of a typical successful trial.
>
> (d) as persuasive as the finding of a typical successful trial.
>
> (e) as persuasive as a package of two or more typical successful trials.

This was a successful Question.[25] Most Committee members refrained from oration and answered with a few words, or even with a single letter.

12. Acknowledgements

Earlier versions of this document were discussed with David B. Clissold, Raymond J. Lipicky, and Janet Wittes; some of the graphs were artfully improved by Vijay Vaidya. The document's faults remain ours.

[25] RRF wrote it, and he is proud of it. Go forth and do likewise.