The following are comments on the draft "Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials," released for comment on May 23, 2006. The medical device regulatory-industry partnership has a lot to gain from adaptation of Bayesian methods, so I am enthusiastic to see this guidance being developed.

Comments

I will organize my comments according to the section numbers given in the draft guidance.

3.1 What is Bayesian statistics?

This section identifies the method of "the" Bayesian approach with Bayes Theorem. However, it is much more to the point to say that Bayesian inference makes probability statements about scientific hypotheses (based on observed data), and Bayesian decision theory compares potential decisions (allocations of resources) with respect to expected costs and benefits. In contrast, frequentist methods form probabilities about past or future data, assuming the hypotheses. Frequentism does not formally support probabilities about hypotheses, or assessments of decisions. The ubiquitous ignorance of this fundamental paradigmatic difference[1] is, I am convinced, the main reason for the prevalence of frequentist methods in applications of statistical inference.

This section claims that, in "the" Bayesian approach, prior and current information is combined "throughout both the design and analysis stages of a trial." Such language might be changed to "throughout both the design, conduct and analysis stages of a trial," to reflect the "adaptive" uses of Bayesian methods mentioned in the following section.

In this introduction to Bayesian methods, a greater stress on decision analysis would provide greater clarity as to why these methods are so important and useful. In particular, once it is explained that the risks of decisions are calculated using probabilities of hypotheses (rather than, say, frequentist error rates), the need for these probabilities and, hence, for Bayesian methods, receives greater emphasis. In other words, it then becomes clear why bayesian and not frequentist inference provides the probabilities about hypotheses that are needed for principled decision-making.

3.2 Why use Bayesian statistics for medical devices?

The first paragraph of this section raises several questions:

- Of reaching "the same decision" with a "smaller-size" or "shorter-duration" trial (reiterated in §3.8): It is doubtful whether "the same decisions," presumably the same as those supposedly supported by frequentist analyses, are categorically even desirable. Frequentist analyses generate probabilities of data assuming hypotheses, rather than statements about those hypotheses or about the expected risks and benefits associated with any particular regulatory action. It is, therefore, difficult to envision how frequentist analyses can be shown to lead to approximately optimal decisions, or even defensible ones. In other words, Bayesian methods should be promoted not because they (more efficiently) arrive at decisions associated with frequentist analyses, but because they assess decisions at all. Granted, some frequentist results appear to be inferential[2] or even decision-theoretic because they mimic certain Bayesian ones; e.g., a one-sided p-value may approximate the posterior probability of the tested hypothesis. However, in other cases, these paradigms can yield grossly disparate results.

---

[1] Michael Oakes (1986). *Statistical Inference: A commentary for the social and behavioral sciences.* Chichester: Wiley.

[2] Inferential: Extending sample results to larger populations.

- Of "smaller" or "shorter" trials: The overall objective of sample size selection is not to minimize trial size and duration, but to balance optimally between expected costs and benefits. Because a sample size requirement aiming for a particular frequentist statistical power does not account formally for the expected costs and benefits associated with recruiting and treating study subjects, it may be either excessive or deficient. A Bayesian sample size calculation, though, accounts for these costs and benefits, so it may lead rather to a smaller/shorter trial, or a larger/longer one. Also, the Bayesian "negative borrowing" mentioned in §5.6 leads potentially to sample size increases. This comment affects also the mention, in §5.5, of potential decreases in sample sizes.
- Of "good prior information:" Often, such information is taken to signify information, based on the results of other experiments, that, independent of a trial's results, concentrates the probability distribution of unknown parameters in relatively small ranges. However, actually, even a flat or non-informative prior based, usually, on a lack of such "good prior information," is good in that it allows Bayesian inference to proceed. Therefore, the existence of "good prior information" may accentuate the advantages of Bayesian approaches, but does not create them. Even with a (say) flat prior distribution, or a maximum entropic one, Bayesian inferences and risk statements are inferences and risk statements, whereas frequentist results are statements about data.
- The distinction between different types of information that should affect regulatory decisions is not so much between "prior" and "intra" study, as between "extra" and "intra" study. Information may affect the formation of priors even if collected at the same time as a study, as long as it is not affected by the results of the study.
- Consequently, this paragraph might be changed as follows: "Bayesian approaches may enable FDA to reach conclusions that are more consistent with all evidence, whether or not obtained from a trial." One could, of course, argue for putting "better supported by" instead of "more consistent with" here, to match the language of ICH E3 and E9 at several points; however, there is a philosophical debate concerning whether empirical data can in any way "support" a scientific conclusion. This debate is arguably outside the scope of this guidance.

On the fifth paragraph: As mentioned above with respect to the first paragraph, the primary reason for promoting Bayesian methods should not be to lessen the burden of demonstrating regulatory claims. Rather, these methods should be advanced because they have the following effects on scientific beliefs and decisions:

- Direct their formation according to mathematical rules, rather than convention (e.g., $\alpha = 0.05$) and speculation (concerning, e.g., what a probability of data indicates concerning the credibility of a hypothesis).
- Formalize them and make them more defensible.
- Ensure their mutual coherence and consistency.

3.3 Why are Bayesian methods more commonly used now?

As the draft guidance states, improvements in computational technology have, undoubtedly, made complex Bayesian methods more accessible. However, the strongest reservation among statisticians to using these methods has been the Bayesian need for prior distributions which, to many statisticians, would seem to "contaminate" analyses with subjectivity. A modernist, positivist view of science prevailed until approximately the 1960s, portraying the scientist as a detached observer objectively investigating the world around him or her. Such a scientific ideal made it difficult to acknowledge and accommodate the need for subjectivity in science.

However, the 1970s and beyond witnessed a greater willingness among scientists to regard subjectivity as the foundation for objective research. As described in the 1980s in http://www.ibri.org/RRs/RR020/20nature.html, increasing numbers of scientists view science as a "participatory" activity, meaning that scientific investigation consists of, or is at least grounded in, interactions between knowing subjects and known objects. In other words, science is always done in the context of a network of "pre-scientific" beliefs and commitments. The motto of Augustine (354-430 AD) summarizes well this sentiment: "I believe so that I may understand." One might adapt Augustine's words to statistics by saying, "The prior makes the experimental observations meaningful."

In summary, to this section of the draft guidance might be added something like "In addition, contemporary philosophy of science is more accepting of science as an iterative learning experience, involving an interplay between scientific observers and empirical observations."

3.6 What resources are available. . .

Undoubtedly, the excellent 1999a and 1999b papers by Steven Goodman have been instrumental in leading many (including me) to prefer Bayesian methods. Nonetheless, they are less about Bayesian statistics, and more about the

- misuses and irrelevance of frequentist tools
- hybrid between the mutually incompatible Fisherian and Neyman-Pearsonian forms of testing
- "likelihood ratio" as a measure of statistical evidence.

3.8 What are the potential benefits of using Bayesian methods?

As I mentioned with respect to §3.2 above, the fundamental benefit of Bayesian inference is the ability to make statistical statements about hypotheses based on observed data, which may support assessing the expected utilities of potential decisions. In other words, Bayesian inference is inferential.

Contrary to the first paragraph of this section, a Bayesian approach may require more study subjects than a frequentist one; however, if the sample size target is selected using a Bayesian decision-theoretic framework, then the costs of recruiting, treating and assessing responses on subjects will be balanced with the benefits of greater information about device effects and probability of regulatory approval. In this way, the overall expected (or other) utility will be maximized.

3.9 What are the potential difficulties in the Bayesian approach?

Few statisticians would agree with the claim, in the first paragraph of this section, that the planning, conduct and analysis of a trial are only *important* (in contradistinction to *crucial*) for non-bayesian approaches.[3] Perhaps the claim intends to say that, for frequentism, the gathering and quantification of prior information is only important. If so, however, even this does not characterize significance testing and hypothesis testing as these frequentist tools were originally promoted. R.A. Fisher, Jerzey Neyman and Egon Pearson, the ones primarily responsible for popularizing these tools, emphasized that the latter could not be applied without prior information. Fisher argued for a post-experimental, Delphic "mediation" process in which the statistician brings together p-values in the mind with "what seems not unreasonable" concerning tested hypotheses, all in the context of intimate knowledge of the scientific subject matter. Neyman and Pearson, on the other hand, advocated a more formal "integration" process

[3]In contrast, the claim, in §5.1, that "the basic tenets of good trial design are the same" would more easily gain agreement among most readers.

of selecting critical regions by combining pre-experimentally what ICH E9 §3.5, more recently, calls the "prior plausibility of the hypothesis under test" with "the desired impact of the results."

Admittedly, some frequentists speak of the "blind statistician," who may analyze data regardless of its origins and potential uses. They do not regard the careful specification and use of prior information as crucial to significance testing and hypothesis testing. However, in this respect, they have broken with their heritage and, possibly, with ICH statistical principles. Consequently, it is inaccurate to imply (as does this section) that the need to account for prior information is a "difficulty of the Bayesian approach" not shared by frequentist approaches.

Also on the topic of prior information: It may be worthwhile to mention the merits of demonstrating, in a Bayesian analysis, that the posterior probability (or probability density) or measurement of expected utilities are substantially unaffected by the choices of pessimistic, neutral and enthusiastic priors. This type of sensitivity analysis is superfluous for analyses supporting a company's internal scientific investigations and decision-making, but may indeed be crucial in a regulatory context.

Clinical trial evidence, as it accumulates, increasingly dominates the priors (a.k.a. the "stable estimation principle"). The utility (or loss) function, however, remains influential regardless of the sample size: Any statistical decision will be optimal given certain utilities, and sub-optimal given others. Therefore, it is a little odd that, at several points, this section emphasizes difficulties accompanying specifying and using priors, but does not at all mention challenges of working with utilities. Admittedly, utilities are not always integrated explicitly into a decision analysis. However, when regulators or sponsors use a posterior probability (or, for that matter, a frequentist outcome) to make a decision,[4] they are, implicitly at least, assuming sets of utilities.

The "Checking calculation" subsection remarks that certain features of Bayesian analyses create greater possibility (again, assumably, when compared to frequentist approaches) for misunderstandings. This remark is almost comical, in light of the ubiquitous misunderstandings of frequentist outcomes.[5] A p-value, for instance, is commonly (usually?) interpreted as the post-experimental probability of the tested hypothesis H; alternatively, it is interpreted as the probability, once one has rejected H, that one is wrong (which is, incidentally, still the probability of H). P-values,[6] even those from "simple" analyses, are much more difficult to understand inferentially (i.e., as statements about hypotheses given data) than are Bayesian results.

---

[4]Even when deciding on the basis of a posterior probability, as stated in §4.1.

[5]The misunderstandings of p-values are arguable the most thoroughly documented reflections of the ways statistics are interpreted. See, e.g., Bandt, C. L., and J. R. Boen. 1972. A prevalent misconception about sample size, statistical significance, and clinical importance. *Journal of Periodontics* 43:181-183; Beck-Bornholdt, H.-P., and H.-H. Dubben. 1994. Potential pitfalls in the use of p-values in the interpretation of significance levels. *Radiotherapy and Oncology* 33:177-178; Berger, J. O., and T. Sellke. 1987. Testing a point null hypothesis: the irreconcilability of *P* values and evidence. *Journal of the American Statistical Association* 82:112-122; Boring, E. G. 1919. Mathematical versus scientific significance. *Psychological Bulletin* 16:335-338; Brewer, J. K. 1985. Behavioral statistics textbooks: sources of myths and misconceptions? *Journal of Educational Statistics* 10:252-268; Campillo, A. C. 1996. [Erroneous interpretation of p values.] [Spanish] *Atencion Primaria* 17:221-224; Capone, C. A., Jr., and S. L. Seaman. 1989. Uses and misuses of hypothesis testing. *Journal of Business Forecasting Methods and Systems* 8:18-27; Chew, V. 1980. Testing differences among means: correct interpretation and some alternatives. *HortScience* 15:467-470; Cowger, C. D. 1984. Statistical significance tests: scientific ritualism or scientific method? *Social Service Review* 58:358-372; . . .and those papers are selected from only the "A," "B" and "C" primary authors.

[6]The same can be said for hypothesis test results.

4.1 Introduction

The first paragraph of this section identifies several "unknown quantities of interest" about which Bayesian methods could make inferences. The second and third bullets are certainly such quantities. However, the bullet "clinical safety and effectiveness endpoints" creates more confusion than clarity because, in device and pharmaceutical trials, the word "endpoint" is used to refer to a variety of concepts, none of which represents a primary "unknown quantity of interest" in a clinical trial:

- A datum from a study subject, already observed in the course of the study. A probability about one such datum is $\Pr(\text{Subject}_t = \text{"success"} | H_o)$, where t is some previous point in time and H_o is the tested hypothesis. This is a frequentist probability, a probability about a previous observation assuming a hypothesis. It is not an inferential probability about an "unknown quantity of interest."
- A datum from a single subject yet to be observed, about which one might make a prediction or predictive probability, as in Bullet 2 of this section. While such a prediction is used in certain adaptive Bayesian trial designs, it is not a quantity about which sponsors or regulators wish to make inferences at the end of a trial.
- An objective (as in the language of §5.5) or goal for which the trial has been (will be) conducted, e.g., "the endpoint of this trial is to assess the efficacy of medical treatment *T* in preventing occurrence of event *E*." Such an endpoint is, obviously, not a quantity at all; however, to use the same word in this context may nonetheless confuse the reader (for whom this guidance is written) unfamiliar with Bayesian methods.

One primary quantity of interest in many trials is, on the other hand, the average[7] degree of improvement in a specific clinical endpoint to be afforded to eventual users of a medical treatment now being investigated, should the treatment be approved by regulators. Perhaps a reference to such a quantity could replace the first bullet of this paragraph in the guidance.

The subsection "Prior distribution and non-informative prior distribution" several times attributes priors selection to "the investigator." However, in clinical trials parlance, "investigator" usually refers to a medical person administering treatments and measuring and recording responses and other data from study subjects once a trial has begun. Sponsors and regulators are usually the parties who assign priors, despite the fact that principal investigators, when they serve on expert panels, may provide advice on priors.

This same subsection speaks about "non-informative" prior distributions. However, the pursuit of the "holy grail" of non-informativity in priors has proven problematic, and many Bayesian philosophers have abandoned it. In the first place, as even Laplace noticed, if non-informativity is achieved according to some criterion for one continuously-valued parameter, say Θ, it will still not be achieved for non-linear transformations such as $\log(\Theta)$ or Θ^3; the best the analyst can do, then, is to choose the function of Θ with respect to which some type of impartiality is to be achieved. Secondly, even so-called non-informative priors convey very specific information about a parameter; a flat prior, for example, for Θ indicates that any interval of length $\partial\Theta$ contains Θ with exactly the same probability. Such information can hardly be considered non-informative.

Claiming (here and in §5.5), therefore, that "absolutely nothing" might be known (which may even be an epistemological impossibility) about a quantity, and that a non-informative prior expresses that lack

[7]The phrase "mean or median" might replace the less specific word "average" here; however, it would do so at the expense of verbosity as well as technicality.

of knowledge, is apt to raise insurmountable difficulties. It may be better to portray certain priors, such as flat ones, as an attempt not to state the parameter lies in one region with more certainty than that it lies in another. Even this description is, however, far from bullet-proof.

4.2 What is a prior distribution?

Describing priors both here and in §4.1 may be a less efficient use of space in this guidance. The guidance could be shortened by consolidating together these two discussions. Alternatively, perhaps, parts of them could be replaced by a brief description of developing priors through, say, elicitation or prior studies.

Some readers may experience difficulty in this section trying to think of the symbol "**x**" as representing an unknown statistical parameter, because **x** is more often used to signify observable data. Clarity on this topic is of paramount importance, since many of the misinterpretations of frequentist test results are the result of thinking of the (deductive) likelihood $f(\Theta|x)$ as if it was a (inductive) posterior density $\pi(x|\Theta)$.[8] Even Frank Yates and R.A. Fisher, at times, confused statements about data given hypotheses with statements about hypotheses given data[9]; since they got it wrong, what hope is there for most of the rest of us, especially when we are asked to assign unfamiliar meanings to familiar symbols such as **x**? Advantages do accrue from keeping Greek letters out of a guidance for an audience which includes non-statisticians. However, use of the symbol Θ may be warranted here (perhaps saying "Θ (theta)"), so as to allow **x** to continue, in the reader's mind, to represent data.

4.3 What is the likelihood of the observed data?

While it is true that "the covariates" affect the relations between data and parameters of interest, they are not material to introducing the Bayesian approach as such. Therefore, they might be mentioned in §6 rather than §4.

As in §4.1 above, "endpoint" is taken here to refer to an unknown quantity of interest, **x**. "Endpoint" might better be replaced by "unknown quantity of interest" or, simply, "parameter," because of the variety of other meanings more often attached to "endpoint" in clinical trials contexts.

It is indeed to be hoped, and even categorically true for conjugate models, that "as more data are obtained, . . .the less uncertainty there will be about the posterior distribution for **x**." However, for other models, more data can on occasion lead to greater posterior variance $E[(\Theta - E(\Theta))^2|\text{data}]$ or uncertainty measured in other ways. Furthermore, the possibility that conjugate models may mask gross differences between prior moments and moments of the likelihood (e.g., if the sample mean differs markedly from the prior mean) is actually a criticism of such models. It may be better, then, to temper the claim that more data implies more certainty, with the qualifiers "usually" or "in most cases."

---

[8] A fortiori, frequentism survives as a popular statistical paradigm precisely on account of these misinterpretations; for, according to them, frequentist results provide the inductive inferences required by the applied sciences. If scientists realized these results are instead statements about data, bayesianism might quickly come to dominate the discipline of statistics.

[9] For instance, both Yates and Fisher referred to "testing the significance of a hypothesis." This cannot be correct; data, not hypotheses, may be significant. Also Fisher claimed the (Neyman-Pearson, prior) probability that a confidence interval contains Θ cannot differ from the (Bayesian, posterior) probability that Θ lies in a realized CI. Actually, however, the difference between these probabilities lies in the different propositions on which they are conditioned.

4.8 How do the Bayesian and frequentist approaches differ?

As I stated with respect to §3.1, Bayesian inference makes probability statements about scientific hypotheses (based on observed data), whereas frequentist methods assume the hypotheses, and do not formally support probabilities about them. The second paragraph of §4.8 almost states this fundamental paradigmatic difference; however, its two instances of the words "based on" are too weak and even obscure the meaning.

5.2 Selecting the relevant endpoints or parameters of interest

This section, as do others, uses the word "endpoint" in a variety of confusing ways, indicating an endpoint is both a parameter (heretofore in this guidance unobservable, in principle) and "directly observable." It propagates the ever-present confusion between probabilities about data and probabilities about parameters. This confusion props up frequentism as an ostensible system of inference.

It would be most helpful in this guidance to consistently refer to "parameter" as an unknowable statistical quantity about which inference is sought, and "endpoint" as a quantity, observable in a study subject, which can provide evidence about a parameter.

5.3 Collecting other important information: covariates

The first paragraph of this section equates covariates and confounding factors. These entities are indeed related; however, in two respects at least, they are not the same. Firstly, covariates are continuous prognostic factors or other independent variables, whereas confounding factors may be continuous, ordinal or nominal. Secondly, covariates may be related to the study subjects' responses but unrelated to their treatment assignments. Ideally, randomization will cause important covariates to be distributed similarly between treatment groups, preventing them from biasing unacceptably the analysis. Analytic adjustment, on the other hand, can take advantage of covariates so as to empower testing, narrow confidence intervals and concentrate posterior probability distributions. Confounders, in contrast, are related to both treatment and response so as to obscure treatment-response relationships. Not all covariates are confounders, and not all confounders are covariates. Some improvements to analyses reduce bias by causing confounders to become covariates or classification variables.

This first paragraph also indicates "a Bayesian trial in which other trials are used as prior information" will (or may?) suffer from imbalances in covariates. How imbalances will occur, or are more likely than in a frequentist trial, should be clarified. If the Bayesian trials referred to here are those lacking a randomization strategy (and hence more potential for imbalances in covariates), this reference needs to be made more explicit. If, however, the chances of imbalances, or ways of minimizing them or reacting to them, are no different between frequentist and Bayesian trials, then this section does not clearly belong in a guidance specifically on Bayesian methods. §5.1 of the guidance already alerts readers to the clinical trials analysis principles common to frequentist and Bayesian methods; these principles need not be re-emphasized excessively.

5.4 Choosing a comparison: controls

As do other sections of this guidance, much of this section explains concepts and promotes principles contained in previously published guidances, such as ICH E10 on choices of control groups. If such discussion is minimized, this guidance will more effectively concentrate attention on Bayesian analyses.

5.5 Initial information about the endpoints: prior distributions

This section offers a third "introduction" to priors. Instead of repeating much of what has already been said, the material of this section could be shortened and combined with that of other sections. Since,

however, improper and reference priors are mentioned (at the end of the section), they should be explained briefly.

Also, to speak of "the" appropriate prior may be overly prescriptive, because a multiplicity of priors may represent adequately the beliefs of a multiplicity of concerned parties or assumptions about how information from a variety of sources should be combined. This language is also inaccurate, because regulators may call for a demonstration that inferences or decisions are sufficiently invariant to the choices of pessimistic, neutral and enthusiastic priors.

This section mentions reasons that existing valid prior information may be unavailable. To these reasons could be added a few remarks concerning possible techniques for substantiating such information when it is unavailable, and incorporating it into analyses.

The last few words of the sentence beginning "Approval of a device could be delayed" contain an erroneous "or:" replace "evaluators or do" with "evaluators do."

This section's claim that a prior may be "too informative," because many subjects have already been studied in previous trials, is puzzling. It seems to demand an unjustified departure from the "least burdensome" criterion mentioned in §3.2, as well as from the depictions, in §4.4 and 6.1, of the posterior as the summary of all available information about the unknown parameter. If in fact "the prior probability that the pivotal study is a success will be excessively high," then some aspect of the device development program has been mis-conceived:

- If, due to that excessive prior probability, the sample size requirement of the pivotal is too low to allow accumulation of sufficient safety data, then the objective of the pivotal has been misstated as a solely efficacy rather than a (possibly partly) safety objective (as in §5.7).
- If that prior probability is too high purely because it exceeds some conventional frequentist power such as 80% or 90%, then, as in §5.8, 6.2 and 6.5, an arbitrary and mathematically irrelevant frequentist probability has been allowed to contaminate a (relatively) philosophically and mathematically principled bayesian formulation of the sample size decision problem.[10]
- If that prior probability is too high, from a proper decision theoretic perspective, to demand that the pivotal be conducted, then the pivotal should not be conducted. The expected net benefit is non-increasing in the sample size, starting with the first subject recruited.

The criteria with respect to which previous studies should resemble the study being planned, in the subsection "Similarity of previous studies to current study," have already been stated in §5.5. They need not be restated. Further, as they may be nominal or ordinal, they are not necessarily "covariates," and would better be called "factors." What is more, although one of the items is "objectives," the previous and planned studies need not be similar with respect to objectives, except insofar as objectives have determined subject characteristics, treatments, endpoints and other aspects of study design and conduct.

5.7 Determining the sample size

This section might also mention the need to adjust the final posterior estimates of treatment effects, should a frequentist or Bayesian flexible (data-sensitive) trial stopping rule be implemented. Unfortunately, throughout statistics, but especially in sequential designs, emphases on testing have overshadowed the proper treatment of estimation.

---

[10]According to http://www.cs.ubc.ca/~murphyk/Bayes/economist.html, such an impure union between Bayesian and non-bayesian approaches is the reason the Microsoft "paper-clip" helper offers, or used to offer, help too often.

The subsection "Special considerations. . ." mentions "effectiveness endpoints." Many have argued, however, that clinical trials examine the efficacy, not the effectiveness, of investigative medical treatments. As explained in
http://www.unc.edu/~uwolt2/cepor/v2n2.htm#s-2,
"efficacy [is] the probability of benefit to individuals in a defined population from a medical technology applied for a given medical problem under ideal conditions of use. Effectiveness can be defined similarly, though it refers to average conditions of use."

6.1 Summaries of the posterior distribution

The language "*based solely on* the posterior distribution" is ambiguous, reminiscent of a child excusing himself for striking another in the playground: "It was *mostly all* his fault." Does this language signify that the posterior determines the results and conclusions (in which case, "based solely on" should change to "is determined by"), or that it only influences them (in which case "solely" is superfluous), or something else?

6.4 Predictive probabilities

The subsection "Deciding when to stop a trial" defines predictive probabilities. Such definitions would better be placed in a section within §4. §6 should focus on considerations regarding the derivations, presentations and uses of such probabilities.

This same subsection mentions the possibility that a predictive probability of success may be high enough to stop a trial. This raises the question, however, concerning "how high is high enough?" which has no principled answer without considering a utility function. In other words, inference is incomplete without decision analysis.

6.5 Interim analyses

More could, of course, be said about any of the topics in this draft guidance. However, the important topic of decision analysis should receive much greater attention, and not only within the section "Interim analyses." Decision analysis can be useful at any stage of device development, and even at any stage of developing or conducting a pivotal trial, even though the complexities of doing so must be weighed against the trial efficiencies gained.

Consider, for instance, the reference in §6.4 to the minimum predictive probability of success required to stop a trial for efficacy. For the thoughtful statistician, this reference raises questions more than answering them, because such a minimum predictive probability can be determined rationally only through use of a utility function. Some important questions would be:

- What types of biological, societal, economic, etc. costs and benefits should be considered when developing a utility function for potentially stopping a device trial?
- What, if any, sensitivity analyses should be run to assure a decision can be supported with alternative utilities?
- Should costs or benefits be realizable within a certain timeframe, to be included in a utility function?
- From what perspective(s) should costs and benefits be calculated: insurers', regulators', consumers',. . .?

A few words concerning these topics might be helpful in this guidance.

Conclusion

I thank the authors for bringing together the draft guidance, and hope that my remarks are of use to them.

Andrew M Hartley, PhD
Wilmington, North Carolina
910-772-7147
Andrew.hartley@wilm.ppdi.com