Regulatory Implications of the Draft ICH F9 (R1) Guidance

Robert Temple, MD
CDER Deputy Director for Clinical Science

ISCTM
February 20, 2020

Overview

- The question raised by the session is whether the guidance urges a fundamental rethinking of the way we plan and think about clinical trials. We have long been aware of the need to identify before we review data, of all critical aspects of the analyses.
  - Choose study end points (precisely described events, time) primary, secondary
  - Statistical test for the primary and secondary endpoints
  - Handling of missing data

  This is well and extensively discussed in ICH E-9, which strongly emphasizes the need to specify and make choices in advance of looking at data (time to first event, any occurrence of event during a interval, rate of events). Hard Choices, but you have to make them to avoid multiplicity and “ data-dredging”.

Multiplicity - My Eye - Opener

We have been conscious of the statistical risk of choosing among mulitiple plausible end points with data in hand for decades but I recall my first experience/revelation: the Anturane Reinfarction Trial (ART) in 1980.

ART

Sulfinpyrazone Vs Placebo

Sulfinpyrazone lowers U.A , treats gout (U.A. values hidden) from investigators

Patient with AMI 25-35 days before entry (enzymes, ECG; no CHF, cardiomegaly)

Endpoint: total mortality but the reported study finding was on SD (early discussion on whether to even count SD)

- First outcome trial by a drug company
- Prominent, supportive NEJM editorial and FDA A.C. support
- Distinguished panel/consultants
- The final analysis and classification of deaths led to major questions

Initial Presentation (excluding 9, 8 on Ant)

	Placebo	Sulfiupyrazoue	P
Patents	783	775	
Cardiac deaths	62	43	0.058 (primary)
Sudden	37	22	0.041
AMI	19	17	-
Other	7	4	-
All Cardiac			
0-6 Mos	35	17	0.021
7- 24 Mos	27	26	-
Sudden			
0-6 Mos	24	6	0.003
7-24 Mos	12	16	-
Non-Sudden			
0-6 Mos	11	11	-
7-24 Mos	14	10	-

[Only one non-cardiac death patient on sulfinpyrazone inclusion (in sulf, group p= 0.076 for total mortality]

Initial Presentation Cont.

- There was much excitement about the finding on early sudden death, but this was one of many analyses, none pre-specified. That was the starting problem, but it gets much worse. On examination, we found 9 post-study exclusions of death and gross mis-classification of deaths, all detailed in 1980 paper by Temple and Pledger [NEJM]
- Exclusions: there was no stated plan. Probably, if excluded BEFORE any event (and presuming randomness) there would have been no major effect (e.g., exclusion in month one). Butt they excluded:

  3 – poor compliance (pills found in dead patient's room) – 2 on Sulfinpyrazone and one placebo
  6 – found AFTER DEATH to have been ineligible: 6 on Sulfinpyrazone, none on placebo

- Most reasons for exclusions were trivial, e.g. patient entered a day too late.
- Classifications of cause of death: generally made more deaths on placebo sudden and on Sulfinpyrazone MI, supporting the "desired" outcome.

ART - Final Results

	Placebo	Sulfiupyrazone	P
Original	62	43	0.058
Add poor compliance	1	2	-
Add “ineligible:”	0	6	-
Total	63	51	>0.2

Led to strong focus on planned analysis

Strong reluctance to exclude anyone with outcome event

Caution on cause – specific mortality in CV trials (beyond CV vs Other)

And, more generally, PLAN EVERYTHING

E- 9 Other Points

It has a very good discussion of NI trials, especially

- how noise/bad performance gives a bias toward success
- Need to specify AND support the NI margin based on past well designed and conducted studies and use a similar design in the new study (consistent with FDA NI guidance).
- It encourages blinded sample size adjustment if event rates are lower than expected or if variances are greater than expected, a good idea.

Finally, it discusses briefly the potential value of subgroup analyses, even where NOT part of the planned confirmatory analyses and is reasonably positive about such exploratory analyses after success on the primary analysis, always with “cautious” interpretation.

E 9 (R1) What's New?

The main message I get is that you MUST plan and describe how to incorporate a wide variety of "intercurrent events"(dropping out, adding another therapy, stopping the drug but staying in the study, etc.) And sometimes the choices are not at all easy to anticipate or to make, and can depend on the nature of the effect being examined.

Historically, we have not liked to exclude from an analysis people who stop the drug or have poor compliance with the planned regimen, plausible as doing that may seem, and because those behaviors will almost surely lead to underestimation of drug effect, and because we worry about the effect on randomization. A potential remedy, cited in the ICD guidance, is enrichment designs.

Enrichment Designs

- E9 (RI) says
- “Certain estimates may necessitate or may benefit from, use of trial designs such as run-in or enrichment designs, randomized withdrawal designs, or titratiomdesigns. It might be of interest to identify subjects who can tolerate or respond to (RT) a treatment during a run-in period (prior) to randomizing.

A really good suggestion, even if, as we know, those procedures overestimate the “real-world” effect size. And they have been used:

VA cooperative HT Studies in the 1970’s Randomized only patients who took placebos prior to randomization. The placebos contained riboflavin which made urine glow under a florescent light.

- Beta-blocker CHF trials randomized only tolerators
- Opiate Trials used randomized withdrawal trials in patients who tolerated and appeared to benefit from opiates
- CAST trial of encainide/flecainide in post-infarction patients with frequent VPBs enrolled only responders (70% decrease in VPBs (study failed)
- PARADIGM-HF trial (sacubitril/valsartan vs enalapriol with prior demonstrations of ACEI tolerance.

Can What to do depend on endpoints

- A particular difference, I think, is in endpoints that are symptoms vs outcomes. Consider treatment of depression vs prevention of stroke. In the former case we would not expect to see much change in effect after a few weeks so when a patient drops out of a depression trial it seems Ok to use the result up to drop out (if, e.g. more than 3 weeks) either carried forward or adjusted by group data (MMRM), This while on treatment, strategy, which is what we do, seems reasonable for drugs that have a symptomatic effect but don't affect the underlying disease. On the other hand, if a treatment is intended to alter the underlying disease (anti-inflammatory, anti-platelet), it may be critical, after a patient stops the drug, to collect the post-treatment data.

One choice I do not make

- Although I understand the interest in understanding the consequences of treatment choice (the treatment policy strategy), I think ITT can be overdone, specifically when patients receive additional treatment (s) that have effects similar to the test drug. In that case the added treatments presumably most likely in a placebo group, can reduce (or obliterate). The effect of the test drug. So I think assessment should stop at the point of added treatment.

Other Analyses

- We are well aware of the need to order endpoints for analysis and reserve specified alpha for the analyses after the primary. But after winning on a pre-specified endpoint, can we look at other aspects of THAT endpoint? I believe NOT doing so loses important information. I am particularly interested in 3 other analyses of the endpoint the study “won” on
  - 1. Distribution of results
  - 2. Effects in subpopulations
  - 3. The time course of the effect
- I should note that our guidance on the content of section in 14 of drug labeling notes two of these analyses for possible inclusion. (distribution of results and time course) and that the multiple endpoints draft guidance (2017) specifically states

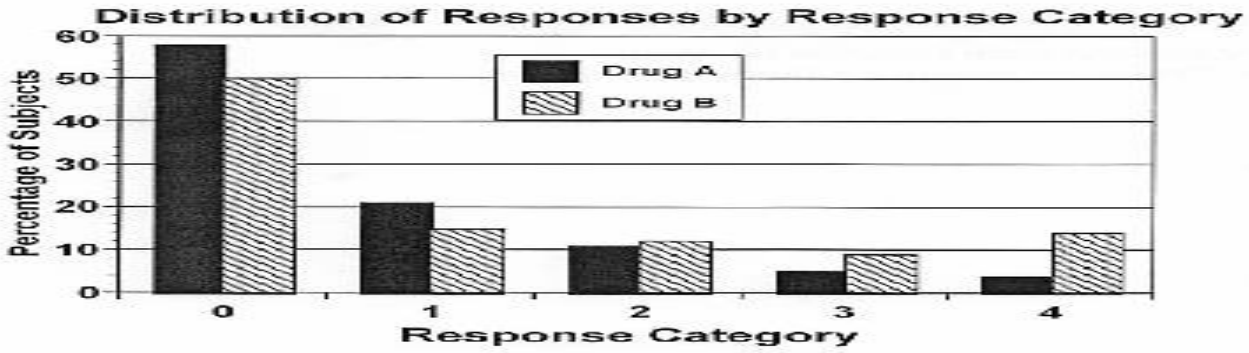
Other Analyses Cont.

- “Once a trial is successful (…”wins” on the primary endpoint (s), there are many other attributes of a drug’s effects that may be described (and) can be informative, (including) the time course of the treatment effect, the full distribution of responses among participants, and treatment effects amongst subgroups”

Distribution of Effect

- This can be shown as a cumulative distribution or as a bar graph, as illustrated in the guidance for section 14

Bar Graph



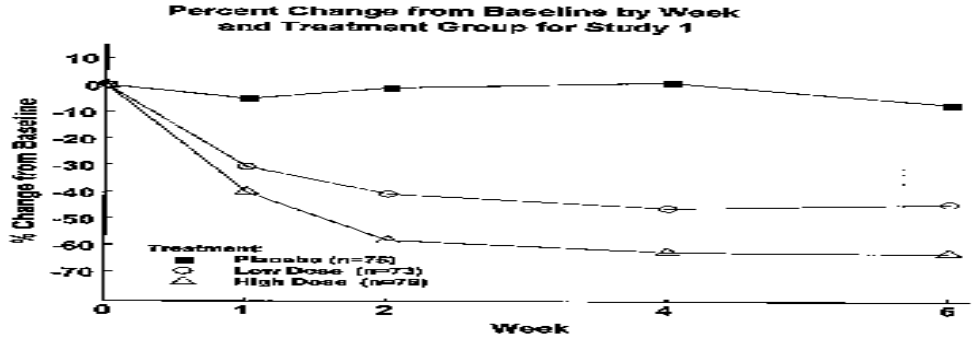
In a bar graph, the length of the bar (the y-axis) represents the group response for the outcome variable or the percentage or frequency of subjects exhibiting a categorical response. Similar graphs can be useful for comparing effects among subgroups. In this case, the response would appear on the y-axis, with various subgroups on the x-axis. In most cases, it is helpful to include error bars. A bar graph should not be used to illustrate just a few numbers that could be summarized better in a table. Graphs in 3-D should be avoided because the values for the bars are difficult to read. Stippling or other small patterns in bars should also be avoided because the bars can be difficult to differentiate after reduction or reproduction.

- Note that in the example, drug B has the larger effect on the more extreme responses. In a classic case, flibanserim for treatment of female sexual dysfunction had an overall effect of 0.5 more than placebo events/week, BUT 10% of treated patients had one event per week and NO one on placebo did.

Time course

- Widely used for symptomatic treatments recently, as illustrated in guidance for section 14. Of course, these are regularly shown in Kaplan-Meier plots for outcome studies.

Line Graph



A line graph most often illustrates responses (y-axis) over time (x-axis) where each line represents the data for a defined group of subjects (e.g., a treatment group, a subgroup). It is helpful in some cases to include error bars and number of subjects remaining on study treatment at each time point. Similar graphs can be used to show dose response with response on the y-axis and dose on the x-axis.

Forest Plots

- This is not in our section 14 guidance but should be. They are used universally in published reports of CV outcomes trials, but infrequently in labeling, (e.g., labeling for the new anticoagulants) but I believe they should be used more often. Classic case is ticagrelor, an anti platelet drug. In the PLATO study of ticagrelor vs clopidogrel in ACS, ticagrelor was superior in all subgroups, EXCEPT in the US, where it was 25% worse and in people on ASA of 300 mg or more (45% worse). In fact, the whole US effect was explained by the higher ASA doses used here.