Robert Temple, MD
Deputy Center Director For Clinical Science
Center for Drug Evaluation and Research
Food and Drug Administration

ISCTM
Sept 26, 2016

MISSING DATA: US REGULATORY CONCERNS:

PROBLEMS OF INTERPRETATION; GOALS AND CONSIDERATION FOR ADDRESSING PROBLEM

Main Issues

In almost any study, some patients will leave the study before completion for good reasons (e.g., adverse effects) or bad ones (tired of coming to clinic) and the analysis of study results must somehow deal with them, preferably by using the data that they did provide in as unbiased a way as possible (although what that means can be debated).

From a regulators viewpoint, there are two main issues:

1. The analysis of whether the drug was effective; generally, was there a statistically significant difference between treatments.
2. (Less critical, in my view), the magnitude of the treatment effect.

What Can Missing Data Do?

1. Suppose it is missing "at random," i.e., bears no relation to the assigned treatment, to the favorable or unfavorable effect of the drug. Then

- Number missing should be similar in both groups
- Dropouts should be similar with respect to favorable or unfavorable drug effects
- Randomization is in no way compromised

What Can Missing Data Do?

There should be no bias in the results

- Even if the missing data give no result for the missing patients and there is therefore some loss of power, there should be no bias, although there could be a change in calculated effect size.
- If early data in the dropouts can be credibly used, either as is, or extrapolated using modeling, power can perhaps be restored (there ARE data on everyone) and, again, there should be no bias.

What Can Missing Data Do?

2. Suppose data are not missing at random, but are more common in the treated group because of an adverse effect (dry mouth, headache, etc).

If the dropout reason bears no relationship to whether the patient would have responded favorably to the drug, results should not be distorted. That is, the people who stayed in, and the data they produced, should be similar for both groups, except for shortened study time (and maybe less data). But the 2 groups should have a similar likelihood of response to drug or to placebo. That is, the randomization should still hold.

What Can Missing Data Do?

But suppose the dropouts are in fact related to the response being assessed, what I long ago learned to call “informative censoring.” For example, maybe the dropouts from the drug group are more or less likely to respond favorably to the drug. This could bias results.

Classic informative censoring case: RECORD study of rosiglitazone.

Eliminating Sicker Drug-Treated Patients

The RECORD CV safety outcome study of rosiglitazone vs placebo posed a problem:

Problem: rosiglitazone worsens heart failure. People with heart failure have increased risk of death from various CV causes.

Eliminating patients with worsened heart failure (cause by rosiglitazone) in the rosiglitazone group would make that group “healthier” and decrease the CV events of interest, perhaps hiding an adverse CV effect.

In a clinical trial of a beta blocker in CHF, there could be a similar problem. Sicker (worse CHF) patients might worsen, dropout of the BB group & make it healthier. Remedy for most of the studies was to assess tolerability, THEN randomize.

Is Censoring Often Informative?

Very hard to know, but some examination is possible. You could, for example, in a depression trial, see whether patients who dropped out were similar, during their time on therapy, in response to the patients who stayed in the whole trial. Of course, if an important reason for leaving was a poor response, they might not be the same as the remaining patients, but dropouts in both groups might be similar, even if more numerous in the placebo group.

Present Practices
Not Uniform

The approach that was most common in depression, psychosis, anxiety studies was “last observation carried forward” (LOCF), simply using the last available value. If dropouts in both treatment groups occur at similar rates over time, this should not distort results too much, but there is one problem in these trials, a very large effect in placebo group [Typical effect in a 6-week study of patients starting at a HAMD of 25 is a fall of 10-12 points on placebo and 13-15 on drug treatment]. That is, the “placebo response” is far larger than the drug effect, i.e., difference between drug and placebo. Patients who leave at, say, week 2, lose most of the placebo response, potentially reducing the effect in the early dropout group, either reducing or enhancing the drug effect.

We have therefore moved from LOCF to modeling approaches, using data from completers to determine what effect to attribute to the dropouts. I am not aware of systematic comparisons of the two approaches but I am sure they exist.

Present Practices

A few years ago analgesics were analyzed using a "baseline carried forward" approach. As opioid analgesics cause numerous dropouts because of constipation, and as there is considerable spontaneous improvement in trial patients, the analgesics, with more dropouts than were seen on placebo, given baseline (no change) values, looked worse than placebo, not the right outcome at all. More recently, randomized withdrawal studies have been utilized.

What Is the Fix?

Don't Have Dropouts

(or have fewer)

In 2010 the NAS/NRC published: The Prevention and Treatment of Missing Data in Clinical Trials

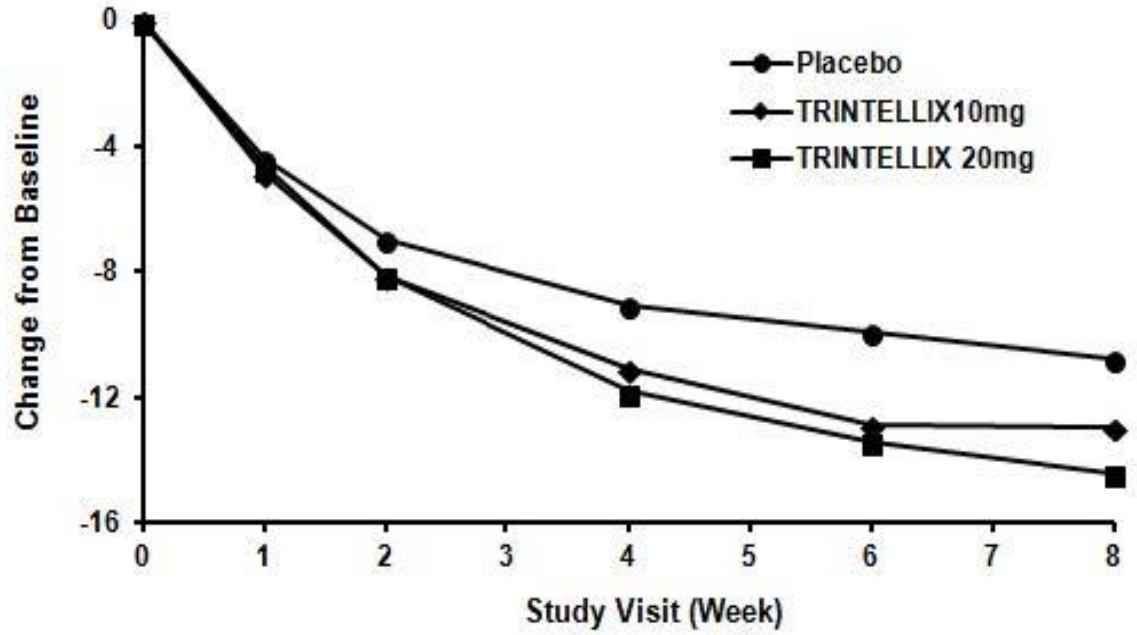
The report considers LOCF and a range of imputation models, generally finding none wholly satisfactory and it strongly emphasizes the need to minimize dropout rates (not so easy). They specifically endorse

- Tolerability run in periods (not OK for acute depression studies but standard for placebo-controlled maintenance studies)
- Randomized withdrawal studies for long-term effectiveness

One more possibility: Shorten the follow-up period for the primary outcome.

One More Possibility
Vortioxetine

Figure 4. Change from Baseline in MADRS Total Score by Study Visit (Week) in Study 5



Most effect is present by 4 weeks and all is there by 6 weeks. Study primary endpoint was 8 weeks.

Shorter Studies
(or shorter primary endpoint)

Years ago it was not uncommon to have, for depression trials, a primary endpoint at 4 weeks. It is now common to use 6-8 week values.

It is also clear that dropouts increase in frequency after 4 weeks. The possible use of earlier primary endpoints should be examined, possibly choosing 4 or 6 based on dropout rates. An analysis by our biometrics group showed lower success rates at 4 weeks than 6, so such a change needs care and could be damaging, but even a move to a 6 week endpoint could create an advantage over 8 weeks. Whatever the primary, of course, the shorter (4 weeks) and longer (8 weeks) endpoints could also be examined.

Shorter Studies
(or shorter primary endpoint)

The NAS/NRC report describes many approaches to improving retention, in some cases assessing off-therapy status if possible. That is sometimes done for outcome studies and can protect against informative censoring, but it will clearly reduce effect size (off-drug cannot benefit from drug). Nonetheless it might be reassuring regarding evidence of effect, even if the on-therapy analysis gives a better estimated effect size. That approach does not seem at all appropriate in symptomatic conditions, where patients on therapy will generally lose effect and worsen or move to another therapy, neither of which is informative about the test drug effect.


FDA
U.S. FOOD & DRUG
ADMINISTRATION