JOHNS HOPKINS
UNIVERSITY

Center for Clinical Trials

Department of Biostatistics
Department of Epidemiology
Department of International Health

Department of Medicine
Department of Ophthalmology
Oncology Center

(Monday) 20 October 2014

Memorandum

To: Trialists

Fr: Curtis Meinert

Re: Data collection mistakes

Starting data collection before forms have been tested

If ever the old saw "haste makes waste" is true, it is in the haste to get started with enrollment in trials. It has taken months to get the protocol written and through IRBs. The only thing standing in the way now for starting enrollment is a few data collection forms. How hard can they be to produce?

Even old dogs like me can get pushed to start before data collection forms are ready. Starting before things are ready will be one of those mistakes you will make repeatedly over your career.

Poor wording

Assuming one does not set out to produce poorly worded data collection forms, how is it that we have so many wording issues when analyzing results? Two reasons. Because of haste to get started and failure to stress data collection forms before use.

A basic fact to keep in mind when framing questions is that questions phrased as negatives are harder to understand than positives. "What is the highest grade you completed in school?" "What is the last grade you completed in school?"

It takes longer to process positives that are negated than positives alone. For example, odds are it takes you a millisecond or two longer to grasp the meaning of negated right or left turn signs than those signs without the negating diagonal line through the arrow.

Consider Brewster Higley's refrain in *Home on the Range* "and the sky is not cloudy all day". What was Higley trying to say? That the sky is cloudy part of the day or that the sky is clear all day?

Double negatives can make your head spin. As a Midwesterner, if I say something is not too bad, I mean it is good. Meeting with a gradual student I told her that her dissertation was not too bad.

After that meeting Watson (aka Betty Collison) asked what I did to the student because she was in tears. I asked why? Watson said you told her dissertation was bad. It was then I realized that I had to stop talking "Midwestern" to students.

Interpreting the absence of a response as a negative or "no" response

This mistake arises in relation to "skips", e.g., as in "Skip to item 19 if no history of pregnancy". If no history is reported it is due to the absence of a history or failure to answer the item. Avoid the confusion by a "Yes-No" filter to indicate whether the person has a history of pregnancy.

No time windows for followup visits

You have to specify the time intervals within which followup visits are to be made in order to be counted as done within the time interval.

The visit schedule in the UGDP was set at 3 month intervals with each visit consisting of a regular followup visit and an eye, heart, kidney, or peripheral vascular examination in the 1st, 2nd, 3rd, or 4th quarter of each year of followup. Visits were numbered sequentially by quarter, i.e., FU 1 the 3rd month after enrollment, FU 2 the 6th month after enrollment, etc.

Well and good, save for what clinics did when people missed a visit. Suppose a person does not show up for the month 6 visit, but does for the month 9 visit, i.e., the second followup visit for the patient, but the 3rd required visit in the idealized schedule. Does the clinic fill out the data collection form designating the visit as FU 2 or FU 3? Some did it according to the idealized schedule and others did it by the number of visits. Needless to say, counting visits to produce performance statistics by clinic was near onto impossible without hard and fast rules as to when a visit was to be counted as missed.

The solution was to construct contiguous time windows that indicated the limits within which a visit was to be done. Visits not done in the specified time interval were counted as missed.

Using local labs for key study data

If data are key to the trial use a laboratory where you can specify and control the collection and laboratory protocol. Generally that means even if the trial is single-center, specimens are sent to a central laboratory.

Persons enrolled into the UGDP had to have a sum glucose tolerance test (GTT) of $\geq$500 mg/100ml (consisting of four glucose values as measured after an overnight fast and then again at 1, 2, and 3 hours post-glucose challenge) to be eligible for enrollment. Glucose determinations were to be done locally on whole blood. There had been discussion of requiring determinations to be done in a single central laboratory but that approach was rejected because of logistical problems in shipping specimens and in getting determinations back to the clinics in a timely manner.

The only issue was whether to require clinics to perform determinations on whole blood or serum. After a fair amount of discussion, investigators settled on whole blood. Things proceeded uneventfully until around three years after the start of enrollment.

An investigator made an offhand remark regarding their method for determining glucose levels during a meeting of study investigators. Since the method cited was one requiring use of serum, another investigator wondered aloud as to how the method could be used on whole blood.

“Whole blood? We use serum.”

“You do? The protocol specifies whole blood.”

"It does?"
And so unfolded the "glucose story" with the ultimate discovery that four of the 12 clinics used serum instead of blood. When the smoke settled, the mistake affected determinations for 280 patients.

That mistake required converting serum values to whole blood equivalents. Since serum glucose values are higher than for whole blood, the conversion resulted in 57 of the 280 patients having corrected sum GTTs below the diagnostic cutpoint of ≥500 mg/100ml.

Changing data collection forms without changing form identifiers

Typically, the version number or version date of forms is part of the data system. An inviolate rule of thumb is to change version number or date when anything on a form is changed. Failure to do so means analysts will not be able to map changes into analysis datasets.

Making promises that cannot be met

In general, anything promised to patients after the trial ends should be scratched from consents. This is especially true of promises of care or treatment after the trial ends. A vexing issue in trials involving life-limiting diseases and drugs not yet approved for marketing. If the treatment proves effective patients have expectations of continued treatment after the trial ends but that will be precluded until the drug is approved for marketing.

Being too explicit in consents as to when a study ends

IRBs want consents that are explicit as to length of followup whereas investigators want to keep options open. Investigators know plans can change over the course of the trial. It is not uncommon for data monitoring committees to recommend extension of followup beyond the original plan. If consents indicates a specified length of treatment or followup, persons will have to be reconsented if the length of followup is longer than indicated in consents.

Use of an eligibility screening variable as a baseline variable

Variable used for screening are subject to "regression to the mean". If a variable is used to determine eligibility, it has to be measured again after eligibility has been determined to provide a reliable baseline for assessing change.

Failure to collect essential baseline data

One of my 19 universal criticisms of trials is failure to collect some "important" baseline variable. You can always think of one that was missed. Never mind that they are rarely useful in explaining treatment differences. The fact is the criticism plays well to the gallery.

You can only collect so much data, so failure to collect a particular baseline variable is only a mistake if the variable is a known risk factor or possible contributors to the disease or condition being investigated.

We took a lot of flack from critics in the UGDP for failure to have baseline smoking histories. Critics argued that the mortality difference against tolbutamide was due to a baseline imbalance in smoking behavior. Investigators did, in fact, make an effort to rectify the oversight via a retrospective collection of smoking histories late in the trial. However, the results were never published because of obvious questions involved in constructing baseline smoking histories long after patients were enrolled and then with the use of surrogate respondents for deceased patients.

Self-administered randomization schemes

A self-administered randomization scheme is one where the clinic is able to obtain randomization assignments without any external controls or checks. The most common form of such schemes is one where clinics are given numbered sealed envelopes containing treatment assignments with instructions to open the envelopes in order as numbered. The difficulty with such systems is that there is no way to ensure concealment of assignments until issue or any reliable way to ensure the envelopes are opened in order as numbered.

Avoid self-administered randomization schemes.