PharmaSUG2011 - Paper DM02

The Illusion of Good Data,

or

How Not to Think of a Blue Elephant

Kim Truett, KCT Data, Inc., Alpharetta, GA

ABSTRACT

When someone says, "Do not think of a blue elephant," what is the first thing that you think of? A blue elephant. The suggestion that has been put into your mind makes it impossible not to think of blue elephants. The same collection bias will occur with improperly designed CRF questions. 'Real time' edit checks can keep sites from entering 'wrong data' during EDC data entry and therefore, reduces queries, but improper use of edit checks can bias data collection.

This paper will present common ways that data collection can create bias, for both paper-based and EDC studies, and will illustrate methods to minimize and avoid data biases that can be created during data collection and will present some ideas for using SAS® software post-entry to look for some of these biases.

INTRODUCTION

In the 1800's in London, the prevailing medical opinion was that cholera was transmitted via air. When the "Committee for Scientific Inquiries" (led by William Farr) sent out surveys to learn more about the transmission of cholera, the questions were all related to air (air pressure, moisture, etc.) and the report they issued "Report of the Committee for Scientific Inquiries in Relation to the Cholera-Epidemic of 1854" detailed these air-related findings and modes of transmissions. But others, particularly, John Snow, disagreed. He mapped out cases of cholera, and showed that they clustered around water pumps. When the handles from suspected pumps were removed, the cholera rates dropped. Had the "Committee for Scientific Inquiries" not limited their questions to what they "knew" was correct, the cholera epidemic might have ended sooner. So, as we conduct clinical trials today, we must still be careful to guard against building in assumptions and asking leading questions in our research.

TYPES OF BIAS

I. Using Leading Questions: questions that suggest the answer or contain the information being sought.

"Are you going to the grocery store?" compared to "Where are you going?" "How do you feel?" compared to "Have you had any headaches?" Which question will lead to over-reporting of headaches?

Which drop burned more?
- o Right eye
- o Left eye
- o Both the same

How often do your peers use alcohol?
- o Often
- o Rarely
- o Never

When was the last time that you saw a blue elephant?
- o Yesterday
- o Last Week
- o Last Year

Why It Occurs and Possible Pitfalls (or "Why we reported seeing the blue elephant")

How a CRF question is asked can influence the data that is collected. This occurs on television courtroom dramas. The questions are asked in a way that ensures that the information the lawyer wants to hear is contained in the answers that the witness gives. This can be a good thing in the courtroom, but, in clinical research, we should be searching for an answer, not directing the sites towards the answer that we expect to hear.

For example, “Have you used any medications in the last thirty days?” may not elicit a response from all subjects, while “What medications have you used in the last thirty days?” suggests to the subject that they should try very hard to think of medications that might have been used in the recent past.

Sponsors may be trying to get more information about an expected event or outcome, because they need to know about a specific event that they judge as a given. For an eye drop, if ocular burning upon instillation was reported in previous studies, the sponsor might ask about ocular burning at each visit. However, astute study subjects, knowing they will get asked this at each visit, begin to watch for this symptom, so they can be sure to report it at their next visit. Therefore, soliciting symptoms actually increases the likelihood that you will see those symptoms.

Questions can have unspoken implications that may not be accurate. For example, the question, “How often do your peers use alcohol?” << Never, Rarely, Often >> assumes that you know the alcohol use patterns of the people around you. It includes an unasked question, “Do you know the alcohol use patterns of the people around you?” If your answer to this is “Yes,” then the question is valid, and can be answered. If you do not know about the alcohol use of the people around you, then the question is not valid and you cannot answer it.

Sometimes, sites start re-phrasing questions as a study progresses. If something appears more frequently in the data that is recorded later in the study, then you might want to query the earlier subjects, e.g., if every med history recorded in May includes aspirin as a prophylactic for heart disease, but none of the med histories recorded in January include aspirin, this might indicate a change in the way that the site is collecting the data.

Suggestions for Improvement (or “How we can learn that the blue elephant didn’t exist”)

The easiest way to fix this is to ensure a more complete set of responses that do not presuppose anything.

Which drop burned more?
- o Right eye
- o Left eye
- o Both the same
- o Neither burned

How often do your peers use alcohol?
- o Often
- o Rarely
- o Never
- o I don’t know

Or split the questions into two questions:

Have you ever seen a blue elephant?
- o Yes
- o No

  If yes, when was the last time that you saw a blue elephant?
  - o Yesterday
  - o Last Week
  - o Last Year

Using SAS to see this bias:

1. Use SAS to look for questions with numerous missing responses, to see if an underlying assumption exists in the questions or answers.

2. Use SAS to analyze frequency of responses for collection bias. If something appears more frequently in the data that is recorded later or only at certain sites such as the appearance of allergies in medical history, then you might want to query the earlier subjects.

II. **Limiting Acceptable Reponses:** Restricting data entry in EDC systems via drop down lists, radio buttons and range checks or on paper studies via choice boxes.

What color was the blue elephant that you saw?
- o Green
- o Red
- o Blue

Record Resting Pulse (should be between 50 – 150) |__|__|__| bpm

Why It Occurs and Possible Pitfalls

Limiting choices makes data *seem* clearer, because all responses are standardized and unambiguous. It makes it easier to reuse SAS code from previous studies, if standardized categories are used. But limiting choices on CRFs can result in data bias as seen from the cholera questionnaires. It forces data into the classifications of interest, but that classification may not reflect the actual data.

It can reduce queries. For example, many are confused about the differences between race and ethnicity. If race is reported as 'Hispanic', this will be queried. A drop down list avoids this problem. But limited choices forces sites to make subjective categorizations. Sites do not always receive standardized and unambiguous answers from subjects. For example, having a checklist of body systems seems like a good idea, but when you start asking the sites to classify symptoms, different sites may classify the symptoms differently.

Adding range checks to continuous variables can reduce data entry errors, but oddly variant data may not be enterable. Although there is a range of expected responses, sometimes the actual responses that are obtained in a clinical practice may differ from the expected responses. If systems are designed to only accept expected responses, then you may be missing important deviations from those expected responses.

Suggestions for Improvement

On paper studies, the site can always enter a marginal note, and those notes entered into a separate dataset. Data management should not simply discard these notes as 'extra information' rather they should be entered and catalogued to allow further review.

In EDC systems, use an 'other' category, to allow the sites to enter the data whenever they are not sure. DM should review the 'other' data and re-categorize if needed, but the site is not forced to make a categorization when they are not certain.

For range checks, having confirmatory checks on out of range data is good, to prevent data entry errors, but no system should disallow entry of out of range data – the actual data collected, no matter how illogical or out of range, should be enterable into the database.

Record Resting Pulse |__|__|__| bpm

What color was the blue elephant that you saw?
- o Green
- o Red
- o Blue
- o Gray
- o Other: _____
- o I didn't see the blue elephant

Using SAS to see this bias:

1. Always enter and review marginal notes on a paper study, looking for out of range data, unexpected replies or unusual explanations that may indicate a CRF issue.

2. Look for data clustered towards one end of a range. This may indicate that there could be values that fall outside the range. Values that cluster at the high or low point of an expected range could indicate that that range was not appropriate for the study.

3. Look for data where an 'other' category would have made sense.

III. Requiring data: EDC data collection instruments that require that data fields must be completed, before respondent can move to the next question.

The MaxPerks account registration requires a business name – even on a personal account. So, my MaxPerks mail comes to Kim Truett; Not A Business; 123 My Street; Alpharetta, GA because business name was a required field and I could not navigate away from the page without providing a business name.

When you registered for the PharmaSUG conference, the registration page required the following fields be complete: Job Title, Department, and Company. But if you are an independent SAS programmer, what do you enter for "Department"?

Why It Occurs and Possible Pitfalls

In clinical research, the same data must be collected from all subjects, and requiring that fields are completed is a way to ensure that. For example, for demographics, at a minimum, age and gender are needed. Primary efficacy data must be recorded; otherwise the associated study could be inconclusive. This is the driving force behind requiring fields to be completed.

Sometimes the data simply is not attainable. For example, companies, in order to calculate an accurate age, may require a day, month and year of birth. In third world country studies, the day and month may not be available, so sites may be forced to make up responses if these fields are required. Some US and EU sites are reluctant to provide precise birthdate information because it can be used as identification. Therefore, they will approximate data, in order to move past the 'required fields'. This is also true for medical histories and medication histories where true accurate dates 'to the day' are rarely attainable. Therefore, if the fields are required, the site is forced to fabricate data.

Visual acuity measurement offers another example. Vision is often recorded as 20/20, 20/40, 20/200, where the larger denominator indicates a worsening of vision. However, at some point, a subject's vision can become so poor that the measurement switches to "counting fingers", "hand motion", and "light perception". A data entry system that only allows for data in the form of 20/xx, would not allow the site to record these low vision results.

Suggestions for Improvement

Allow overrides. The system can to be set up to allow a code for 'not applicable' or 'unknown' as acceptable responses and should allow gaps to remain when the data is actually missing. This will allow quick differentiation between inadvertently missing, and unknown / not applicable data.

Avoid overly precise data collection, where the level of precision required will result in inaccurate data. For example, in many cases, an age generated from year of birth, would be sufficient. Allow the data to be entered in a looser format. For example, dates for events that occur during the study can be precise, but dates for historical medical history events should allow for partial dates.

Using SAS to see this bias:

1. Use SAS to look for sites where all the birth months and days are the same.

2. Use SAS to look for cases where history dates are all the same (every medical condition for a subject started on June 30th.)

3. Use SAS to look for missing data where data should exist (like missing Acuity) to see if the data was simply not enterable.

IV. Using page layout to define “expected replies”

If you provide a list with seven lines and ask, "List your most recent blue elephant sightings," some respondents will see the seven lines as representing the magic number of sightings that you want listed. Respondents, who have seen more than seven blue elephants recently might list their most recent seven sightings. Respondents who have seen fewer than seven blue elephants, might list purple and aqua elephants, in an attempt to fill in the empty lines.

Likewise, a medical history collection page that presents 5 lines will lead to a different medical history collection paradigm than a medical history collection page with 20 lines.

Why it occurs and Possible Pitfalls

Sponsors do not want an unwieldy number of responses. For example, medical history should only contain significant or relevant data, but significance and relevance are both subjective. The more space that you provide, the more people feel compelled to fill up that space.

If you limit the amount of space, then you encourage the sites to focus on the data that they think is most significant and relevant (which may not match sponsor expectations). If you limit space too much, you will not collect all the relevant information, because the sites will feel compelled to stop when the space is filled.

Suggestions for Improvement

For paper studies, allow a pre-determined number of rows, but also include extra pages with additional room for entry. It removes the sense of obligation, if those extra rows are included on a separate page, but still allows additional information to be submitted if needed.

For EDC studies, the best option is clearly defining what needs to be collected, training the sites on the materials, and ensuring that completion guidelines clearly define the level of depth of collection need.

Using SAS to see this bias:

1. Use SAS to count the number of entries by sites, to look for sites what always have the number of medical history entries that exactly matches the number of medical history rows.

2. Use SAS to look for variation across sites in the number of AEs and history reported. While some sites may have a less healthy population, this review may also indicate that some sites are over or under reporting.

3. Use SAS to look for clusters of symptom responses seen at one site but not at others. Clusters of responses could indicate collection bias. If a subject states, "I've got allergies," this might cause site personnel to start prompting subjects by asking "Do you have any allergies?".

CAUTION: HIDING BEHIND STANDARDS:

Standards are good, but be careful about introducing bias or bad science through enforcing standards. Below are four examples, where, by hiding behind standards, companies inadvertently introduced problems into their data.

Example 1: A US based company required the use of “African-American” for non-US studies because that is their “standard race codeset”. Standards need to be adaptable to be accurate for different circumstances.

Example 2: In ophthalmology trials, OD or OS (right or left eye) must be collected for ocular history and adverse events. When a non-ophthlamic company collected ocular information, they did not have a standard variable for location. Instead of expanding their standards, location was collected at the beginning of the AE term. This led to sites not recording the location, extra queries to collect location, and the analysis programs had to parse the location out of a text field. Use opportunity to expand standards, rather than trying to shoehorn the information into standard forms in which it does not fit.

Example 3: A company's standard is that they do not allow protocol exceptions, so the only option for eligible at baseline is yes, but if an underlying condition that makes the subject ineligible that is only discovered at the 6-month Visit, the subject cannot be un-enrolled. In this case, there is no available answer for the question 'eligible at baseline'.

Example 4: Companies vary as to whether they define AEs as after informed consent or after first study drug. Some countries require that the AEs that start after IC but before drug be collected, particularly if there is washout during that period. Hiding behind the standard of 'AEs only after start of study drug' may result in neglecting to collect information that may be required by regulations of certain countries.

Suggestions for Improvement:

Standards groups, enforcement departments and others responsible for standards within a corporation need to blend standards enforcement with the flexibility that accurate clinical trials data needs. They must recognize that exceptions are needed, and improve judgment of when to allow deviations from the standards.

Examine the use of validated questionnaires, particularly in quality of life (QoL) areas, that are much better than each company developing its own unique set of questions. These standardized questionnaires have been tested in large groups and been tweaked until they allow for accurate collection of the data of interest.

Looking at standards developed by groups such as Clinical Data Interchange Standards Consortium (www.CDISC.org) and Society for Clinical Data Management (www.SCDM.org) - where many companies are contributing to make data collection standard and consistent, and because of the expansive involvement, these standards tend to be more general.

CONCLUSION

As SAS programmers, the data collection seems predetermined as we are the 'recipients' of the data. But actually we have the benefit of hind-sight. We get all the data. We can see how the sites responded to all the questions. And we can use this unique view of the CRF data collection, to examine the collection practices, and look for bias in data collection that could affect the outcome of a study.

Therefore Data Management, more than any other group, has the appropriate data to provide feedback to CRF designers and clinical trial teams that show possible steps that can be taken to clarify forms, to minimize data collection bias, to avoid unexpected responses, all of which can improve clinical trial results and reduce variation between sites within a single clinical trial.

REFERENCES

Lu Z, Su J Clinical data management: Current status, challenges, and future directions from industry perspectives *Open Access Journal of Clinical Trials* 2010:2 93–105

"Report of the Committee for Scientific Inquiries in Relation to the Cholera-Epidemic of 1854", London, 1855

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Kim Truett
KCT Data, Inc.
11877 Douglas Rd; Ste 102-146
Alpharetta, GA 30022
Work Phone: 770.372.0989
E-mail: KCTData03@kctdm.com
Web: www.kctdm.com