

Data Management, Statistics and Dog Bones???©

O.M. Little, PhD, MBA
OML Associates

Data management and statistical analysis projects -- particularly in the pharmaceutical industry and therefore Contract Research Organizations (CRO's) -- are no small undertaking. Conversing with the 'uninitiated' on the fine points is usually a difficult undertaking -- almost a fool's errand.

To facilitate communication between all concerned a familiar shape -- that of Dog Bones is suggested. While the concept and the image may sound trivial, substantial validity exists. Having used the term for several years it is surprisingly effective. Most have at least seen a Dog Bone. The picture very accurately conveys the project management concept.

Everyone has seen multiple projects scheduled by "project management". Gantt Charts "Rule"... There is a difficulty. Those doing the Gantt charts are seldom those intimately familiar with the processes in the data management and statistical areas of companies. We know what happens:

- ▶ The first project is "contracted" then "planned" by an outsider. See Figure 1.
- ▶ The next thing that happens is the second project is "contracted" then is "planned" by the same outsiders.
- ▶ These folks don't know the steps involved, and think they have done this wonderful job of outlining the project. Think back to the typical Gantt outline you may have been given -- there are these series of "progress bars" across a page. It just happens that the "progress bars" are simply rectangular. Those Operations Research types (OR's) either:

- (1) Don't know the volume of front- and back-end work involved, or
- (2) Are so fascinated with their success with MS Project that nothing else matters!?

Taking a detailed look at the steps in the process may add understanding (see Figure 1).

Protocol Development and Study Design are a substantial bundle of work. These involve several different individuals. There are Data Managers and Statisticians involved these engage members substantially.

Case Report Form (CRF) Design also engages these same two groups (Data Managers and Statisticians) plus programmers and data entry supervisors.

Database Design occupies Data Managers and Programmers primarily, but also Statisticians -- to ensure the concepts of the data management is appropriate. Access to data points in a most efficient manner will be most useful.

Central Lab (data) Processing usually involves Data Managers, Programmers and Statisticians (slightly) to ensure the connections are appropriate and facilitated for later analyses.

Encoding is a matter largely between the Data Managers and the Clinical Staff. Data Managers typically create the "dictionary" in the datafiles -- these are the numeric codes that will appear on CRF's for expansion in the "proofing listings". Programmers are occasionally involved to ensure the data is exactly received and recorded for subsequent analyses.

CRF Review and Correction Processing engages the Data Managers / Data Entry Supervisors and mostly the Clinical Staff. CRF Review begins mostly with the Clinical Staff and begins before data entry gets underway. IN THEORY: Case report forms are "reviewed" before they reach data entry. Data recorded in the CRF must "match" the possible choices in the database. CRF's often must go back to the investigator for "correction". ((For the record, corrections require the investigator and/or the study nurse at the investigational site. These "corrections" must bear initials and dates. Exact details depend on the Standard Operating Procedures [SOP's] in place.)) When corrections are returned these are typically "doubly entered" in the database. The original entry must be replaced and the data file updated to "journal" the changes. Recent efforts at Remote Data Entry, places data may be being entered in the investigator's site. The concept of "data details" must match "allowed entries" in the database. ((Not represented on Figure 1, but represents a sizable involvement mostly for the clinical staff.))

Data Entry / Management involves mostly Data Entry Operators and Data Entry Supervisors. Data Managers and Programmers are occasionally involved when minor "modifications" occur. The details vary depending on local SOP's. This is a very time- and people-intensive part of the project. New technologies involve remote entry, but the essence is this is a big effort!!! If new technologies are involved, the difficulties attached are relocated from the "home office" to the "remote investigational site". This may sound like an improvement, the author suspects this aggravates the difficulties owing to differences in training levels at various investigational sites when multi-center studies are involved. The database still has to be validated and certified. Substantial testing must be done to ensure the "matching 'double entries'".

Draft Analysis / Report Writing are heavily reliant on Data Managers, Programmers and Statisticians. It seems almost impossible to start this event sufficiently early. Those who did not understand the process at the beginning seem to think they know about management. It is difficult to analyze data that doesn't exist, but that doesn't seem to persuade. The logic in analyzing non-existent data escapes the author.

Some of the difficulties are creating the sets and subsets that may be required. There is almost always an Intent-to-Treat (ITT) dataset. This involves all patients that satisfy both the inclusion and exclusion criteria. Additional datasets may be either specified or allowed according to the protocol. These datasets must be allowed. The ITT dataset is the premier set. The others are supplemental, but can certainly be very important. Tables and graphics are particularly important at this stage; it is extremely important to correctly represent the results. Among the very important steps here are spell and quality checking (ensuring the details of p-values) are correctly reflected and interpreted in the text. Simply put, this is very manpower-intensive. Very little can be accomplished by machine.

Draft reports get a lot of scrutiny. Typically, draft reports travel a complex orbit inside both companies -- the CRO first, eventually the sponsor. The trip inside the CRO occurs before the draft leaves for the sponsor. There may be several levels of revision before the draft gets to the sponsor.

A particularly difficult element is getting the draft report reviewed by the sponsor. ...and getting comments back in time to make "timelines". A two-week delay in this review because the sponsor has other priorities drives a lot of planning off track.

Given moderate success the report will return with coherent comments with only two or three weeks delay. This is the time intensive work begins again. The sponsor's comments must be incorporated. This typically requires changes in multiple definitions, datasets and countless tables. This overhaul is virtually a new report. One of the difficulties is there is almost always increased "time pressure" attached. The sponsor has a submission deadline eminent -- nevermind the sponsor was the cause of major delays in the first place.

Final Report Writing is typically another complete iteration of the draft report. The CRO management cannot antagonize a sponsor with untoward comments or results. The same quality- and spelling checks are required. The maximum levels of clarity and explanation are also necessary. This is the report that will go to the Food & Drug Administration.

Documenting and Backing Up all files begins once the Final Report has been sent to the sponsor. This is another level of

work: All programming and data files are all included in this action. Good judgement says that exact copies of ALL DOCUMENTS will be copied and retained. All files are provided to the sponsor and to "off site storage" for the CRO's records.

Having completed all these steps it is easy to see there are collections of actions that amass at both ends of a project. These accumulate at the two ends of projects. The problem is -- all of these are high-pressure, high-quality elements that cannot be sloughed over. Defining and Installing the Database, Writing the Draft(s) and Final Reports all are extremely intense.

To treat all the steps in these projects as routine is misrepresentation to those involved. It seems inappropriate to treat the planning cycle Gantt charts and simply rectangular. These front- and rear end pieces are extremely critical and taxing. With all this it seems appropriate to re-examine the planning cycle by having the planning modules more appropriately represent the "bolus" of activities on the two extremes of projects.

Examining a collection of "planned projects" illustrates the issue. All stages of all projects are treated similarly. While this may be acceptable for many departments. It is the author's opinion this does not work for data management and statistics projects. Review Figure 1: Project Planning Template. Notice the volume of work beginning with Develop Protocol and ending with Database Design that occurs early in the project. Also, notice the concentration of work in the Draft Analysis through Closeout that occur at the end of the project.

Figure 2: Department Project Summary presents six projects. These are of varying lengths. The darker text on top indicates the planned "window" for each project. The lighter text below indicates progress to date. All looks workable – especially to the OR's.

What to do? Take a look at the concept of the picture with a Dog Bone representing the project. This becomes especially important when multiple projects occur. The solo project is not much of a problem. Adding the second project amplifies the problem. Consider the accumulations of work that occur at the "front end" and "rear end" of the project in Figure 1, Project Planning. Review Figure 2, Department Project Summary.

Figure 3 is believed to better reflect the process of project planning for Data Management and Statistical shops. The Dog Bone Gantt Chart better represents reality. If we look again at Figure 2 there is a very busy last half of 2003. Attaching the "bolus shapes" to the ends of projects helps understand the extent of work anticipated. What is eminently clear in Figure 3 is the workloads on the department. There are too many databases being started and too many projects being concluded at the same times. The "bolus" on each Dog Bone helps reflect the real workload on the department staff. Data entry – while important and not to be overlooked – planes out for these projects. The peaks occur at the start and conclusion of all these projects.

```

Protocol Name / Sponsor's Identity...

Task Name / Title

Develop Protocol / Design study      XXXXXXXXX
-Develop specific objectives         //
-Define inclusion criteria           //
-Define exclusion criteria           //

CRF Design                           XXXXXXXXX
-CRF pages                          //
-Unique CRF pgs                     //

Database Design                      XXXXXXXXXXXXXXXX
-Customize design                   //
-Validate tools                     //

Central Lab processing               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-Comm'l lab                        //
-Freq'y of cks                     //

Encoding                            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-Select dictionary / implement     //
-Spell checks                      //

Data Entry/Mgmt                    XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-Double entry                      //
-Conduct range cks                 //
-Logical & consistency cks         //

Draft Analysis / Report             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-Develop draft code                //
-Write draft rpt                   //
--Spell QA                        //
--QA Tables/P-values              //
-Revise draft rpt                  //
-Print & copy / Package rpt        //
-Ship to sponsor / client          //

Final Analysis / Report             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-Incorporate client's comments     //
-Revise, incorp't QA input         //
-Final Review                      //
-Final Report                      //
-Print/copy/package rpt             //
-Ship final to client              //

Closeout / File Xfer / Archive      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
-Finalize documentat'n            //
-Final version of data files       //
-Backup final versions/programs    //
--Data files                      //
--Program codes                   //
-Ship SAS/ASCII files to client    //
-Ship to offsite storage           //

```

FIGURE 2: Department Project Summary Template

Data Management, Statistics Project Summary										27 Oct 2003			
Project \ Date	Aug '03	Sep '03	Oct '03	Nov '03	Dec '03	Jan '04	Feb '04	Mar '04	Apr '04	May '04	Jun '04	Jul '04	Aug '04
Spons ABC; Proj 001 Progress:	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX //////////												
Spons DEF; Proj 002 Progress:	XXXXXXXXXXXX ////////												
Spons DEF; Proj 003 Progress:	XXXXXXXXXXXXXXXXXXXXXXXXXXXX ////												
Spons GHI; Proj 001 Progress:	XXXXXXXXXXXXXXXXXXXXXXXXXXXX ////////												
Spons ABC; Proj 005 Progress:	XXXXXXXXXXXXXXXXXXXXXXXXXXXX ///												
Spons JKL; Proj 003 Progress:	XX ////////												

FIGURE 1 : Department Project Summary Template

