PhUSE 2009

Paper DH05

"This study is essentially just like that other one"

Louise Mazzeo, Quanticate, Hitchin, UK

ABSTRACT

A data handling slant on why 'this study is essentially just like that other one' is the most erroneous statement ever to be spoken in the CRO environment.

Within a CRO, in order to work quickly and efficiently to meet client demand, a lot of effort goes into avoiding reinventing the wheel where possible for each study received. Therefore when presented with a clinical study which appears to follow the same structure and protocol as a previous one, it is expected that those involved in data handling should rejoice with glee at the prospect of being able to reuse SAS code and procedures that are 'tried and tested'. However having experienced this situation I believe that this glorious scenario is as rare as hen's teeth. This presentation will discuss the issues and illustrate how even the addition of one question to a CRF page can affect database structure, data consistency/edit checks and the design of test data for this process. It will also give examples of how an additional option to a question might cause formatting and mapping difficulties, amongst other difficulties that have been personally encountered.

INTRODUCTION

The purpose of this paper is to highlight that seemingly minor differences between studies can have a significant impact on how these studies are handled within Data Management (DM) and DM Programming teams. Where one study precedes another similar study, it can result in a reduced time and resource allocation for the second study based on the assumption that they are 'essentially the same'. When this occurs it is considered appropriate to reuse much of the code and procedures, sensibly reducing the time and resource and therefore cost of the study to the CRO and client.

However it can be surprising to discover that the addition of one question to a CRF, different table structure or format can have a significant impact on the developed tools and procedures. In such circumstances where databases/SAS code (procedures) cannot simply be transferred, they would need to be redeveloped. This requires that following the full development cycle, including specifications, testing and the QC process be completed once more.

Unawareness of such instances could result in unfair expectations to be placed on the CRO personnel to complete tasks under-resourced and require client-facing personnel to deal with the frustrations of the client should time lines also be affected. Inability to explain the rationale for such circumstances would result in difficulty justifying the requirement for increased cost and resource to the client. Additionally it could result in the perception of inefficiency and a lack of resourcefulness amongst all personnel and directly affect working relationships.

This paper serves to identify seemingly minor differences that may occur between 'similar' studies and to consider circumstances where databases/SAS code (procedures) cannot simply be transferred but would need to be redeveloped. It discusses how key factors of Personnel, Protocol and CRF design contribute to these differences and therefore should contribute to decisions during the planning stage.

PLANNING FOR A STUDY

PERSONNEL

It would make good sense that the same personnel would be used to work on a study that is 'essentially' the same as that other one. CROs are however constrained by timelines set by their client, which may require them to develop different teams. This would result in different people working on two similar studies within the CRO. This can be further confounded by the Pharma personnel differing between studies, which can also have an effect on the content of the deliverables. With so many individuals, it is not hard to conceive how four different groups can affect the deliverables in such circumstances - even if there were no decisions to be made during communications!

T9-Mā ' TŌ ... T9-MÄ ' TŌ

In addition to time zone differences and geography the data issues should be planned for and considered when studies are run in Europe and the US. Lab data units particularly will differ (imperial versus metric) and if overlooked could result in issues arising in range checks or potentially at the statistical reporting stage. Date formats differ between these regions and so would need to be recognised and specified for example in an OC database and if manipulated in SAS programmes. Such information on transfer protocols and for personnel should be unambiguous and identified in order to avoid such mistakes.

PROTOCOL AND INDICATION

Factors such as: variations in number of visits; types of data collected at each visit and specific primary/secondary end points affect the design of the study and therefore any variation in these would result in differences between studies. Whether all labs are to be collected at every visit, or a select few depending on the visit would affect the database design, format catalogue (DVG) or variable type.

LABORATORY DATA

There are occasions where multiple Lab datasets are produced since different labs are collected at different visits. The naming conventions should be consistent across the study, but this does not happen by accident! The effects of this includes the recreation of a new format, CRF page layouts, new data edit checks for different CRF pages, new validation listing programme and possibly the creation of a new format for summary purposes.

A different protocol could change how the lab data is collected. The addition of time points to a page may not at first appear to change much, until a new CRF page layout results in a different dataset structure in order to make data entry simpler. Data validation code used which assumed one test per visit would now need reviewing in order to ensure merges were intact.

Compare Table 1 and Table 2 and visualise the dataset that would be created. Notably setting these SAS datasets together would require the creation of a new format to encompass all time frames and would require extra time to be allocated for subsequent data mapping. For a similar purpose lab test fields that are prefilled on a database would also require a format. If prefilled values differ at each visit a different format for that field has to be created. Much frustration can occur when multiple lab datasets are created which cannot be set together as the Lab test name variable has different formats.

The following tables for Lab data look very similar and it could be expected that the resulting datasets would be identical apart from the time of sample variable. This example was chosen as in this case database design resulted in 10 variables (1 for each field) present in the dataset produced by Table 2. If the database structure had collected the data as 5 variables indexed by time frame the same edit checks as Table 1 could have been applied. If this is to be achieved, manipulation of the Table 2 dataset would be required prior to running relevant edit checks or validation listings. This would have an impact on the transferability of the code for this process and data mapping.

Table 1

LABORATORY DATA (Blood collection for laboratory tests including, PSA, D-Dimers and Troponin)											
Date of Sample									**Procedure**	**Time Frame**	**Not Done**
									Laboratory Data	Visit 3 T_{end} + 24 hours	☐
D	d	m	m	m	y	y	y	y			

Table 2

LABORATORY DATA				
Date of Sample	**Time of Sample**	**Procedure**	**Time Frame**	**Not Done**
d d m m m y y y y	h h : m m	Laboratory Data (Hematology, Chemistry, Coagulation, and D-Dimers)	Visit 2 T_{end} + 4 hours	☐
d d m m m y y y y	h h : m m	Laboratory Data (D-Dimers and Troponin)	Visit 2 T_{end} + 8 hours	☐

CRF DESIGN

Database structure relies on page layout of the CRF and is designed particularly for data entry and not analysis. It is understandable that the CRF should be made as simple as possible to complete for the investigator and clear that data entry will be faster if following a perfect copy of the original script. As a general comment, however, where database consistency is affected detrimentally a tolerance to minor alterations to the data entry screen should be expected. Especially since in a typical CRO, expensive OCR machinery is not available and so data entry has the advantage of being conducted by humans.

PAGE LAYOUT (1)

There can be instances of separate page layout which affects the structure of the dataset created. See the following example of a vital signs page and two separate page layouts (Table 3 and Table 4). These different layouts allow for a question difference and the addition of two variables. However edit checks could not simply be copied from one study to another, since the key question regarding vital signs being taken has a different format: yes/no question [1/2] versus not done ['X']. This is sufficient change to require editing for all checks involving this question. Albeit not extremely time consuming but would require the development cycle to be followed and therefore budgeted for.

These two tables (Table 3 and Table 4) are an example from the same study resulting in data manipulation to create one large dataset and programming to determine the presence and integrity of data relating to the particular page it came from. This would in turn affect any subsequent sister studies that may or may not repeat these layouts and require personnel to programme keeping in mind the transferability of their code across studies.

Table 3

VITAL SIGNS	
Were Vital signs taken?	☐$_1$ No ☐$_2$ Yes (if Yes, please complete below)
Expected Time Frame:	Day 0 Prior to Anaesthesia
Time of collection:	h h : m m
Systolic Blood Pressure:	☐☐☐ mmHg
Diastolic Blood Pressure:	☐☐☐ mmHg
Pulse Rate:	☐☐☐ bpm
Position:	☐$_1$ Sitting ☐$_2$ Supine ☐$_3$ Standing
Oral Body Temperature:	☐☐.☐ ℃
Weight:	☐☐☐.☐ kg

Table 4

VITAL SIGNS							
Expected Time Frame	**Not Done**	**Time** (hh:mm)	**Oral Body Temperature** (°C)	**Blood Pressure** (mmHg)		**Pulse Rate** (bpm)	**Position** (1 Sitting 2 Supine 3 Standing)
				Systolic	**Diastolic**		
T_0 - 5 minutes	☐	h h : m m		☐☐☐	☐☐☐	☐☐☐	2
T_0 Start of injection	☐	h h : m m		☐☐☐	☐☐☐	☐☐☐	2
T_0 + 5 minutes	☐	h h : m m		☐☐☐	☐☐☐	☐☐☐	2
T_{end}	☐	h h : m m		☐☐☐	☐☐☐	☐☐☐	2
T_{end} + 15 minutes	☐	h h : m m		☐☐☐	☐☐☐	☐☐☐	☐

PAGE LAYOUT (2)

The following Tables (Table 5 and 6) are a further example of how CRF design affects the structure of the dataset in which information collected. Consider the datasets that would be created for these tables. It is clear that for Table 5, clinical significance field information would be specific to the examination (PR interval etc); however, the clinical significance would not be specific to the examination for that data produced from Table 6.

Although a standard check for listing 'Abnormal, not clinically significant' would select relevant records from each of the datasets, the format of the output would be sufficiently confusing. The differing format of output produced would confound the interpretation of the data for statistical reporting, queries and data validation. In effect, it would be unsuitable to transfer the code for database structure, data validation and safety reporting tables. Therefore as mentioned previously the development cycle would need to be employed for all tasks involving the collection, validation and reporting of this information.

Table 5

ELECTROCARDIOGRAM (12-LEAD)				
Was an Electrocardiogram performed?	☐1 No ☐2 Yes			
Electrocardiogram Date:	d d m m m y y y y	**Time:**	h h : m m	
Examination	**Result**	**Normal**	**Abnormal, Clinically Significant** (Record on Adverse Event page)	**Abnormal, Not Clinically Significant**
PR Interval	☐☐☐ ms	☐	☐	☐
QRS Interval	☐☐☐ ms	☐	☐	☐
QTc(B) Interval	☐☐☐ ms	☐	☐	☐

Table 6

ELECTROCARDIOGRAMS							
Expected Time Frame	**Time ECG Obtained** (hh:mm)	**PR Interval** (ms)	**QRS Interval** (ms)	**QTc(B) Interval** (ms)	**Normal**	**Abnormal**	
						CS*	**NCS**
T_0 - 12 minutes	h h : m m	☐☐☐	☐☐☐	☐☐☐	☐	☐	☐
T_0 - 5 minutes	h h : m m	☐☐☐	☐☐☐	☐☐☐	☐	☐	☐

TEACHING YOUR GRANNY TO SUCK EGGS?

It is unlikely that any of the points outlined in this paper are 'news' to those experienced in working in multiple studies. It should however help to increase 'healthy' cynicism of those repeating tasks on similar studies in order to evaluate possible discrepancies.

In an example of three 'identical' studies, it was fortunate that preliminary analysis work by a statistician helped to identify an important discrepancy. Again code/procedures could and were transferred across the studies to save time and cost. Study set up picked up any incongruities that would prevent code working between studies and once corrected allowed the following 2 studies to begin with minor delay. However six months after data entry had started it was noted that a variable had different format attached to it across studies [0-3 versus 1-4]. The identification of this could have been timelier, but this serves as a valuable example of the importance of ensuring studies do not receive diluted time and resource living in the shadow of their older brother.

When there is a study that 'is a repeat of a previous one' and has numerous personnel attached to it, inevitably ownership of design and data integrity is diluted. It is imperative that this dilution is limited and not to the detriment of the study integrity. This critical point is one that should be considered when estimating time and resource.

CONCLUSION

The nuances of each study should always prompt programmers to critique the true transferability of their code. A large Pharma has the advantage of identifying multiple projects from the same pipeline that would benefit from global and local libraries of study code and procedures. The effort involved in ensuring transferability and trying to envisage all scenarios for a group of 2 or 3 studies is not always cost-effective for the CRO.

In a CRO environment, two studies could be 'essentially be the same' if there is good and thorough planning. Ideally, the same personnel (CRO and Pharma) would be involved to ensure consistency is achieved in a lesser time. A perfect repeat of a study with the same CRF design for the same protocol and indication would be an unrealistic hope. However an understanding and appreciation of the effect of the minor differences and allocating appropriate planning time for a sister study would avoid placing counter productive constraints on time and resource.

CONTACT INFORMATION

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

Louise Mazzeo
Quanticate
Bevan House, Bancroft Court, Hitchin, Hertfordshire, UK, SG5 1LH
Work Phone: +44 (0) 1462 440084
Fax: +44 (0) 1462 440 086
Email: louise.mazzeo@quanticate.com
Web: www.quanticate.com