

Good versus Better SDTM: Data Listings

Mario Widel, Eli Lilly & Co, Indianapolis, IN

Henry Winsor, Relypsa Inc., Redwood City, CA

ABSTRACT

What does SDTM have to do with data listings? The popular answer is not much, if all you have is "good" SDTM. SDTM (or its predecessor with FDA, Item 11 data sets) have been supposed to replace data listings for about 15 years now. However, data listings aren't going away anytime soon, so while they are still here why not make the absolute best of them? They are still a required component of an ICH compliant CSR. Too many people think of data listings as a simplistic dump of the data, but when properly developed by the right people they can be a very useful tool for data review and CSR preparation. The authors will design and prepare an Abnormal Lab Listing (which is actually referenced as a table in ICH compliant CSRs) and show some ways in which SDTM creation can be used to store lab metadata that improves the listing appearance and utility while simplifying the creation.

Disclaimer: The opinions expressed in this paper are solely those of the authors and not necessarily those of Eli Lilly & co, Relypsa Inc., CDISC, FDA or ICH

INTRODUCTION

Back in the dark ages (circa 1985), the first efforts to submit data to FDA in electronic form were begun. Originally, there was something called Case Report Tabulations (CRTs) that listed all the CRF data in tabular form, and the first efforts were on paper. Paper tabulations could be very large (in excess of 20,000 pages, four boxes of paper!), so the initial Computer Assisted New Drug Applications (CANDA) projects started with converting these piles of paper into electronic files. This worked somewhat well in small trials, but as study enrollment sizes increased, you had pdf files with thousands and thousands of bookmarks, rendering the files almost unnavigable. Can you imagine an 8000 patient trial, with 35-40 domains of data? It was simply not workable to expect reviewers to be able to manage and work with files of such size, even if sponsors could create them. This caused FDA to dump electronic CRTs (and CANDAs) and request the submission of the study data in standardized data sets, starting with Item 11 data sets in 1999 and later SDTM data, starting in 2005.

Some of you probably already think that you know the relationship between data listings and SDTM, specifically that data listings are one of the predecessors of SDTM data sets. Actually, they aren't. While data listings can appear to be similar to CRTs but their purpose is different; they are supposed to be listings of the data for a Clinical Study Report (CSR) that presents the data necessary to support the analysis conclusions reached in the CSR. This not all of the data in the clinical database, only the data necessary to support the CSR.

Sometime around 1995, the International Consortium on Harmonization (ICH) introduced the first Guideline on the content of a CSR, calling it a Common Technical Document, which has since morphed into eCTD, which contains all of the files submitted to a regulatory agency in a filing. But, the original guideline for the CSR is pretty much unchanged since 1995, with data listings still specifically identified as among the contents of a CSR.

Now, in a practical sense, not everybody makes a distinction between CRTs and data listings; they contain many of the same data fields, but the difference is that data listings need only contain a subset of the clinical data, and that they are not necessarily exclusive to one domain. However, very few companies have ever put much design effort into data listings. All too often they are just dumps of the data, pushed off on the junior staff to kick out quickly and not to spend too much time thinking about the contents, just get it done and let the end user get out of it what they need.

WHY DATA LISTINGS

The reality is that data listings have always been the illegitimate stepchild of Clinical Study Analysis and Reporting. They are almost always an afterthought, as the trial statisticians are focused on reporting results, not the data itself. This is by no means an unreasonable practice, as our employers tend to focus on the results and treat everything else as superfluous. However, our end clients are not necessarily just the people who give us paychecks (but thank you very much for them, we are most appreciative!!). As a regulated industry, we are responsible to three letter (four in the case of Japan) government agencies and their employees as well, who do not always have the same trust and faith in our abilities to accurately present the study results and require data in a reviewable format. Also, the path to final results can be a long and bumpy path, and there are times where being able to quickly present selected data to others in a readable format is a valuable skill.

While we have covered the formal need for data listings in submission documents, there is a use for listings in providing others with small, targeted looks at data elements as part of the workflow in conducting and reviewing clinical trials on an ongoing basis. Most reviewers of clinical trials do not have great computer skills in terms of extracting and formatting the data into useful layouts, and it isn't probably a great use of their time to take the time to acquire them (Should a Chief Medical Officer take a Basic SAS class?). That is why programmers are part of the study team, they can make everyone's job easier by being able to produce little snippets of the information for others on demand.

The CDISC sponsored initiatives can take credit for a lot of things, but providing a non-programming literate reviewer with an easy, intuitive way to look at the clinical data is not one of them. ADaM is a result focused output format, SDTM contains everything but the kitchen sink and is focused on being a repository for all of the data, often too broad and hard to view by someone not well-versed in the intricacies of the format. They are data sets and not everybody in the clinical world is sufficiently skilled in query languages like SQL to be able to produce a meaningful extract from the data sets, much less quickly. While tools useful to non-coders have been promised for twenty years, the shelf of non-vaporware offerings is quite bare.

NEED TO PROVIDE INFORMATION TO USERS IN A FORM THEY CAN USE

A common way which we provide to users is spreadsheet data (.xls or .csv data files) that is accessed through Excel, which is familiar enough of a tool for most users to be able to browse data, select subsets and otherwise work with the data in whatever fashion they desire. Alternatively, many of us are still printing out data dumps for users on paper, so they can spread the pages out and reorganize them to fit their needs. A third method is to sit down with the user, work up a clear set of specifications so they can be coded and prepare a custom report for the user. The problem with this method is that it can be time intensive and all too often, the time needed to prepare a report from scratch is too long to provide the user with something meaningful in the necessary time frame.

A fourth method is to have a set of program templates lying around, ones that have been designed in conjunction with the end user to contain meaningful data laid out in a compact fashion, so that the end user can see at a glance what is normally relevant for a topic. The template is easily modifiable, so that if the user needs something added that isn't part of the template, it can be done quickly or it's also easy to remove parts that aren't necessary.

While this sounds like a lot of work, you already have a reason to have a set of listing templates if you are in the business of providing content for Clinical Study Reports. The ICH Guideline E3, on the Contents of Clinical Study Reports, list several sections where the sponsor is supposed to provide listings of the critical data for the CSR. While some sponsors may provide the listings as data sets, they are provided as electronic inserts of tabular data by most, simply because the people working on the CSR need access to the clinical data for writing purposes, and they tend not to be people who do well when asked to view a clinical data set through the SAS Universal Viewer.

WHAT DATA LISTINGS DOES THE ICH E3 GUIDANCE SPECIFICALLY REQUEST?

In the ICH E3 Guideline, requests for listings appear in several places. Section 12.2.4 describes the expected contents of an Adverse Event Listing, although the listing itself is designated to appear in Appendix 16.2.7. The guidance is that the listing should be sorted by investigator and treatment group, with a number of variables to be included that do not appear in the SDTM AE domain. The complete list of other variables is age, sex, race, weight and height if relevant, last treatment at time of event, last dose of treatment, length of treatment prior to event, and any concomitant treatment for the event. One final comment is that any codes and abbreviations should be listed, preferably on each page. In other words, the request is for a complete exposition of the subject's status in the trial at the time of the adverse event, with additional non-AE domain variables included as necessary and the information presented as a contiguous report, so that a reviewer can see everything they need to know about the event in one place. No mention of hyperlinks to data sets, or pointers to other data is here, except a request for the physical location of the case reports if they are provided.

Because the AE listing is so important to a report (in most studies, even more important than efficacy), let's look at this request for a minute as to what it isn't. It's not a request for a dump of the AE data. There are derivations requested, so considerable manipulation of the data is necessary before rendering the Appendix. It also wants everything relevant in one place, not spread over multiple pages or locations in the document. It is meant to be read by a human being, and should be constructed with that in mind.

The next listing referenced is a listing of specific adverse events; those events that led to subject death, are categorized as serious and/or required "intervention, including withdrawal of test drug/investigational product treatment, dose reduction, or significant additional concomitant therapy". The same data points are requested to be included, which should trigger in any programmer's mind that they should construct an Adverse Event listing and reuse it for this request, including only those event that meet the criteria for inclusion. This listing is to be placed in Appendix 14.3.2, so it actually appears in the Appendices before the Listing of all Adverse Events.

Appendix 14 is suggested to be titled Tables, Figures and Graphs Referred to but not Included in the Text, and contains three sections; Demographic Data Summary Figures and Tables, Efficacy Data Summary Figures and Tables, and Safety Data Figures and Tables. Safety Data Figures and Tables is further broken out, to include Summary Adverse Event Tables, the Listing of Deaths, Serious and other Significant Adverse Events, Narratives of Deaths, Serious and other Significant Adverse Events and a Listing of Abnormal Laboratory Values (to which we will later return).

See a request for Medical History here, or Vital Signs? What ICH is doing is pointing to what it considers the “key” data in the trial and puts those summaries in Section 14, emphasizing its importance and not bulking out the section with everything under the sun. That there are Data Listings explicitly included in the Safety Summary items should be a clue that the review of Safety Data is handled differently than the review of Efficacy or Demographics, that Safety reviewers focus on individual cases rather than aggregations of the cases to do their work.

The next reference to data listings comes for inclusion in Appendix 16. While there are a few listings under 16.1 (randomization scheme, etc.) that might be produced by a programmer, most if not all of the content in 16.1 is usually produced by Clinical Operations, who holds much of the content and is in a better place to provide it. Section 16.2 is titled Patient Data Listings and contains eight items:

- 16.2.1 -- Discontinued Patients
- 16.2.2 -- Protocol Deviations
- 16.2.3 -- Patients Excluded From the Efficacy Analysis
- 16.2.4 -- Demographics
- 16.2.5 -- Compliance and Drug Concentration Data (if available)
- 16.2.6 -- Individual Efficacy Response Data
- 16.2.7 -- Adverse Events
- 16.2.8 -- Laboratory Measurements by Patient (if requested)

No guidance is provided for the content other than the title, although the expected content for the Adverse Events listing has been provided earlier. Appendix 16.3 contains the complete CRFs for selected patients, then there is Appendix 16.4, helpfully titled Individual Patient Data Listings, so as to differentiate them from Appendix 16.2, Patient Data Listings. Confused? Welcome to the club, as neither ICH nor FDA have ever clarified the distinction between 16.2 listings and 16.4 listings. Some companies put data listings from domains not listed in 16.2 in 16.4, others have interpreted 16.4 as the place to put Patient Profiles (a special listing that contains all data for a single patient, usually organized more compactly than a printing of the patient’s casebook). We suspect this lack of guidance is deliberate; the agencies are basically saying put what you think appropriate here, we might actually look at it if we need to do so.

Notice also that there are a lot of study domains that are not referenced here. While some sponsors have gone to extreme lengths to wedge all of the usual suspects into some place in this sequence (Medical History as an extension of Demographics, etc.), the authors doubt that this inclusion of everything does little more than increase the size of the CSR while adding little if any value. You may recall that you probably know people that back in college would bulk out their college term papers in order to conceal their scanty understanding of the paper topic. We suspect that reviewers pretty much think the same thing of unnecessarily huge CSRs, and probably review their contents more closely than they would a briefer, but well-composed CSR.

APPENDIX 16.2 PATIENT DATA LISTINGS

Let’s go into a little more detail about what belongs in Appendix 16.2. The first item is Discontinued Patients, which technically can cover everyone enrolled in the trial, as everyone does eventually terminate the trial. However, there are two emphases that need to be implemented here; one is to focus on those subjects who prematurely terminated from the trial, second is an overall listing that contains all subjects screened in the trial, including screen failures. The first is to allow a reviewer to quickly identify those subjects that may need further scrutiny, the second is to allow them to see any subject in relationship to all others from the site.

Protocol Deviations contains slightly different data from most of the rest of the listings, both SDTM data if your study has populated the DV domain and programmatic extracts from the other domains. As such, this listing will change content from study to study much more than the other ones, as what is considered a deviation in one trial may not be thought to be so in another. An ADaM data set for deviations could contain all deviations, not just those items that qualified for the SDTM data set.

Patients Excluded from the Efficacy Analysis should be exactly that, a listing of those enrolled subjects who for whatever reason, their data is not included in the efficacy analysis. The reviewer will also appreciate a clear rationale for each exclusion.

Demographics should contain more than what is in the DM SDTM domain. Besides the usual demographic data points, you should include baseline assessments, especially those that may influence treatment assignment. The idea here is that you should be able to provide the reviewer with a clear picture of each subject at the time of study enrollment and do it on one page, not scattered over a multitude of listings.

The compliance listing is another listing where a SDTM data dump will not be sufficient; this is not merely a listing of drug exposure. Compliance information is derived from the union of Exposure and Drug Accountability, and it is suggested that the derivations be mapped into an ADaM data set, so the data may be available for both summarization and the data listing. On the other hand, if you are reporting concentration data, a combination of the PK and PP SDTM domains is appropriate, only because the PP domain is a domain of data derived from the PK domain, so a separate ADaM data set is not necessary.

The Individual Efficacy Responses listing is also not a good candidate for a “standard” listing, as the responses are likely to change from study to study, even their importance will fluctuate. What is called for here is a generic style listing, which contains the programming necessary for the investigator/subject/timing/dosing information (hereinafter referred to the “left side of the page” information) and a blank canvas for fleshing out the rest of the specific listing.

The content of the Adverse Events Listing has already been addressed, but the importance of laying out a standard AE listing cannot be overemphasized. There’s a lot of data fields of importance here and a lot of interest in the contents, so this should be a priority for standardization. This listing begs for data fields to be combined in a column and lengthy text responses in some fields can result in text wrapping/page break issues. What is most important here is readability; you should do whatever is necessary to keep the presentation of the adverse events for a subject compact, so a reviewer can see all of a subject’s events in as few pages as possible.

For a listing that is nominally “by request only”, the Guidance goes into considerable detail as to how the Lab Response listing is to be laid out, something not done for the Adverse Event listing, which is the only other listing where specific content guidelines were provided. The reason for this is simple; there’s a lot of content here and the authors suspect that reviewers don’t want to be assaulted with every potential layout for such a big chunk of content. The guidelines request that the same “left side of the page” information be present as in the Adverse Event listing, that test results should be presented over time (in a column for each test) and lastly, that the tests “should be grouped logically, (e.g., hematological tests, liver chemistries, electrolytes, urinalysis).”.

Everyone should be familiar with the major lab test categories (Chemistry, Hematology, etc.) and have them as LBCAT values in their SDTM, as the lab sources for the results will supply them from the sample collection types, but the subgroups are a topic avoided by many. The grouping at this level is strictly a sponsor decision, as certain lab tests belong with others almost always (Liver Function Tests), while other groupings may be compound or even study specific. This classification should be provided to you, preferably in SDTM as a LBSCAT value. As for whether you should even do the work until requested, you should expect that the listing will be requested. Even though the SDTM data is provided to the reviewer, the likelihood of a reviewer wanting to see all of a subject’s lab tests on a single page without programming should not be considered “optional”.

DESIGNING A DATA LISTING

There is little guidance ever given on data listings, either by the appropriate regulatory agencies or by company management, mostly because the topic isn’t always considered important enough for “important” people to spend time on, or because “Everybody knows what goes in a data listing, just do it.” The reality is, although you may have a title that has “programmer” in it, the real value in your job is knowing how to organize and present clinical study information in a fashion so that others can quickly and easily glean what they need to know from it. We are word processors, only we use a fourth generation programming language to organize the information instead of typing it into a document.

If we go back to the ICH guideline for the adverse events listing, we noted a large number non-AE data points that are requested, such as basic demographics and treatment information. We think you should take this request as being applicable to all data listings, that the extra variables provide context that the reviewer will want on all listings. The key is to provide the context while not taking up too much space. The basic layout for a data listing in the 21st century is still a piece of paper, either in portrait or more likely in landscape orientation. This shouldn’t be construed to mean that data listings must be printed, only that they fit into the same layout paradigm as the rest of the report. For some reason, landscape is preferred to portrait, only because the extra columns on the page seem to allow you to pack more information on a single line.

A good layout is one that conveys as much study information onto the page that is desired to communicate the point of the listing, without being so compacted that lots of explanatory footnotes are necessary to decode content. Please remember that footnotes take up a line (at least), pull the reader’s eye away from the actual content and otherwise reduce the amount of study information that can appear on the page. Also, tall column headers take up lines and you should balance the need for explanation with the space needs. One thing that is far too often seen is a listing with

column headers 2-3 lines high except for one column with a 6 lines or higher column header, effectively wasting three precious lines per page.

Well, we've addressed the edges of the listing, it's time to talk about the middle. What goes into the body of the data listing? It needs to be relevant information, not every data point in the available data. As you may have noted, while SDTM data is consistent, it also contains a number of redundancies, while ADaM has even more redundancies by design. Your job is to pick and choose what variables need to appear that you know the reviewer will want to see in as compact of a way as possible. One of the most common things included is timing information, when an event occurred or an item was collected in the trial. One of the biggest space wasters in listings is the use of full dates when a calculated relative study day is available that communicates more clearly when something happened than a hard date. You also don't have to display the full chain of all values and calculations in a derivation, only the final value(s) of interest so long as your steps are properly documented someplace. You should depend on the SDTM/ADaM data sets to be comprehensive and not redo what is already there, so do your complicated derivation in an ADaM data set and document it there.

DATA SOURCES

ICH E3 precedes the existence of the SDTM/ADaM data standards, so there isn't an explicit reference to using them as sources of data. While this may imply you have the freedom to choose any data source you want, do you really think that is a good idea? Reviewers are now starting to expect traceability for data sources, you already have to supply SDTM and ADaM (if applicable) data, why would you want to give them a third, competing data source?

Most information in a data listing should be able to come from SDTM; recall that we are not focused on summarizing data but are displaying the data in order to make it possible for a reviewer to gain an overview of all the data. As derivations are pretty much forbidden in SDTM (other than the ones FDA has specifically requested), you may find it convenient to pull some information from some ADaM data sets; not only derivations, but also data that is being presented in an analysis context. With this, you also gain all the advantages of using standard data sets in programming, specifically familiarity, reusability and adaptability. The variables will be familiar to your colleagues so they can easily modify your programs for their needs, coming closer to fully reusable code. While not necessarily completely reusable, a great percentage of the program will be standardized and make someone else's life easier in the future. A standard layout becomes familiar to your readers, so they know what to expect and can help in modification suggestions to make it easier. Don't think of a listing program as a single item. Think of your listing programs as one in a chain of continually improving programs at your company, and suddenly your management may actually see the value of encouraging some real effort in doing the work.

EXAMPLE -- ABNORMAL LAB TABLE

Table One (below) is an example of a perfectly acceptable Abnormal Lab listing; let us look at the problems here. Starting with left side of the page, the starting dose is identified (study treatment groups are identified by starting dose), but there is no investigator identified. While the subject ID does contain the three digit site number, and the site number does uniquely identify the investigator, you may have noticed that there isn't any way to know what site number corresponds to which investigator here. The idea behind sorting by investigator is to allow someone paging through the pile to get to a particular site by searching alphabetically by Primary Investigator. That's hard to do if the data isn't presented on the page.

Next, you notice that the treatment information column is thirteen characters wide, when the meaningful information presented in the column only takes up four. Column headers are much more adjustable than data, care should be taken not to waste space on the page. The authors have little problem with the Subject ID column, but the absence of any age/sex/race/weight information is not a positive. The visit column is unnecessarily wide; while it's nice to repeat visit labels verbatim from what was used in the protocol, it is not a requirement. Some judicious editing of the text can preserve the meaning while utilizing fewer columns.

The Collection Date column is an example of the biggest wasters of space in data listings, especially when someone thinks that the Study Day isn't clear unless you surround it with parentheses and provide an explanatory footnote. The only meaningful information here is the Study Day; it makes explicit where in the collection happens in the progress of the trial. Dose level at Collection is a space waster like the first column; the meaningful data here only needs four display columns, the header should be smaller than 13 columns.

Test Name and Unit may be separate columns in the data, but this is one place where combining the two values into one can save space, especially if you abbreviate Blood Urea Nitrogen as BUN. If this was SDTM data, we wouldn't recommend abbreviating the label (changing Controlled Terminology is never a good idea); you should be using CDASH values in SDTM, no matter how long and verbose they can be. Here, an abbreviated value, especially one that doesn't need a footnote, is an option you can and should use. One bigger problem evident here though, is look at the order of the tests. Alphabetic sort may be quick and easy, but ignores the guideline of grouping tests in a logical fashion.

XXXXXXX, Inc.
Protocol XXXXXX-000

Date/Analysis: 30JUN2013/Final
Page 1 of 2

Listing 14.3.4.1
Listing of Abnormal Serum Chemistry Values
Safety Population

Randomized Starting Dose	Subject ID	Visit	Collection Date (Study Day) [1]	Dose Level at Collection[2]	Test	Unit	Source	Normal Range	Result	Alert	
7.5 g/day	51801	Screening	16APR2012 (-35)		Blood Urea Nitrogen	mg/dL	Central (XXLab)	7 - 25	30	High	
					Calcium	mg/dL	Central (XXLab)	8.5 - 10.3	8.3	Low	
					Creatinine	mg/dL	Central (XXLab)	0.8 - 1.5	1.9	High	
		Baseline: Day 1	21MAY2012 (1)	7.5 g/day		Blood Urea Nitrogen	mg/dL	Central (XXLab)	7 - 25	28	High
						Calcium	mg/dL	Central (XXLab)	8.5 - 10.3	7.7	Low
						Creatine Kinase	U/L	Central (XXLab)	0 - 235	280	High
						Creatinine	mg/dL	Central (XXLab)	0.8 - 1.5	2.5	High
						Glucose	mg/dL	Central (XXLab)	70 - 125	177	High
						Potassium	mEq/L	Central (XXLab)	3.5 - 5.3	5.59	High
						Potassium	mEq/L	Local	3.9 - 5.1	5.5	High
		TRT: Day 3	24MAY2012 (3)	15.0 g/day		Blood Urea Nitrogen	mg/dL	Central (XXLab)	7 - 25	31	High
						Calcium	mg/dL	Central (XXLab)	8.5 - 10.3	7.5	Low
						Creatinine	mg/dL	Central (XXLab)	0.8 - 1.5	2.5	High
						Glucose	mg/dL	Central (XXLab)	70 - 125	128	High
						Potassium	mEq/L	Central (XXLab)	3.5 - 5.3	5.65	High
		Potassium	mEq/L	Local	3.9 - 5.1	5.5	High				
		TRT: Week 1	28MAY2012 (8)	15.0 g/day		Blood Urea Nitrogen	mg/dL	Central (XXLab)	7 - 25	34	High
						Calcium	mg/dL	Central (XXLab)	8.5 - 10.3	8.1	Low
						Creatinine	mg/dL	Central (XXLab)	0.8 - 1.5	2.3	High
		TRT: Week 2	04JUN2012 (15)	22.5 g/day		Blood Urea Nitrogen	mg/dL	Central (XXLab)	7 - 25	31	High
						Creatinine	mg/dL	Central (XXLab)	0.8 - 1.5	2.4	High
TRT: Week 3	11JUN2012 (22)	22.5 g/day		Blood Urea Nitrogen	mg/dL	Central (XXLab)	7 - 25	34	High		
				Calcium	mg/dL	Central (XXLab)	8.5 - 10.3	7.8	Low		
				Creatinine	mg/dL	Central (XXLab)	0.8 - 1.5	2.7	High		
				Glucose	mg/dL	Central (XXLab)	70 - 125	126	High		
				Phosphate	mg/dL	Central (XXLab)	2.5 - 4.5	1.6	Low		

[1] Time (days) relative to the date of the first dose of investigational medicinal product.
[2] Dose level at collection is displayed when there is a corresponding date from the dosing data.

X:\XXXXXXXX\XXXXXXXX-000\Biostats\programs\TLFs\t-abnormch.sas 02JUL2013 19:09

Table 1 Normal Listing Design

The presence of Source at all is questionable. The only lab collected in this trial that was evaluated by a local lab was Potassium. To indicate that you need 21 columns? The last three columns have their own issues, but will be addressed later. The only issue to point out right now is take a look at the results column. Is it easy to see the individual values? Realistically, this column should be closest to the test name for readability, not separated from it by the lab name and the normal range.

Finally, take a look at the two footnotes provided. While we don't think either is necessarily inaccurate, we question the need for them to be present at all.

MAKING A GOOD LISTING BETTER

The good listing is 146 columns wide, let us see how much we can improve the layout, add necessary information and see whether we can increase available space for any additions that may be useful.

Take a look at Table Two (below). Going from left to right and top to bottom, the first addition is a line that contains the Primary Investigator information. This is inserted in the body of the listing using a compute block and so long as the data is sorted by the Primary Investigator name, the listing is manually searchable.

The Randomized Starting Dose column has been narrowed, with the column header abbreviated (with an asterisk added to reference a footnote), the unit moved to the header (they are all the same in this trial), and the data is still formatted to 4.1, so it all fits in four columns, with only an additional footnote line in exchange.

A new column is added, with a column header of AGE/SEX. The column width is three, so is subject age (2 decimal places) concatenated with the first letter of the subject sex. Some will argue that a dividing slash is necessary, is it? Does the content require an explanatory footnote? It may be argued that the column doesn't contain all of the requested information (Race, Weight, Height), we are cheating and considering those variables of minimal use in this trial, but they can be added if deemed necessary. If so, an explanatory footnote is probably helpful, as there isn't a clear abbreviation for Race, even with the limited categories requested by FDA. The investigational product dose is not weight dependent, so weight is probably superfluous. At most, the column may need to be widened further and an additional footnote may be necessary.

The Visit column width has been reduced from sixteen columns to eight, by implementing the following remapping:

- Screening => Screen
- Baseline: Day 1 => Baseline
- TRT: Day 3 => Day 3
- TRT: Week 1 => Week 1
- TRT: Week 2 => Week 2
- TRT: Week 3 => Week 3
- TRT: Week 4 => Week 4

Do you see any ambiguity introduced by shortening the Visit Label? This represents changes that can hopefully be introduced to the protocol before it's completed, but nobody will be confused by seeing these labels here with the full labels in the protocol.

Next, we have removed the complete date column, left the Study day with a column header of DAY, and made the same reduction to the Collection Dose column that was done to the Randomized Starting Dose column, and spanned the two columns with the word "COLLECTION". Thirty two columns has been reduced to ten, with no loss of meaningful information.

The Test/Unit combined column has gone from taking up twenty six columns to twenty two columns, not a significant change, but test names can be long (so can some units. The important thing here is the test names are clear and easily readable.

The remaining columns have been extensively modified, to make the presentation both compacter and more readable at the same time. First, the values have been individually formatted, to display the correct precision for the test, and also aligned so that the decimal place is a vertical column. This display is easier to view results from different tests in one column, and isn't that difficult to do, assuming that you've put some structure in place in your SDTM data. If you have properly formatted your LBSTRESC values (a very big if, we have yet to see anyone else do this), you can extract the consistent format for each test and determine the offset number of spaces that you need to precede each result to keep the decimal point aligned and the overall size of the character format that will display all results consistently. The rest of the field uses only one column to display the abnormal flag (H=High, L = Low is pretty intuitive), but only displaying the applicable high/low range value and an indicator for the local lab isn't so an explanatory footnote is added. Still, the overall width has gone from forty three columns to twelve columns.

XXXXXXX, Inc.
Protocol XXXXXXX-000

Date/Analysis: 30JUN2013/Final
Page 1 of 2

Listing 14.3.4.1
Listing of Abnormal Serum Chemistry Values
Safety Population

RSD* (g/d)	Subject ID	Age Sex	Visit	Collection Day	Dose (g/d)	Test	Value	F(NR)+
<< James, Dr. Claudia - Bethesda Hospital, Maryland. Enrolled 7 subjects. >>								
7.5	51801	47F	Screen	-35		BUN (mg/dL)	30	H(25)
						Creatinine (mg/dL)	1.9	H(1.5)
						Calcium (mg/dL)	8.3	L(8.5)
			Baseline	1	7.5	Potassium (mEq/L)	5.59H(5.30)	
							5.5	H(5.1)+
						BUN (mg/dL)	28	H(25)
						Creatinine (mg/dL)	2.5	H(1.5)
						Glucose (mg/dL)	177	H(125)
						Calcium (mg/dL)	7.7	L(8.5)
						Creatine Kinase (U/L)	280	H(235)
			Day 3	3	15.0	Potassium (mEq/L)	5.65H(5.30)	
							5.5	H(5.1)+
						BUN (mg/dL)	31	H(25)
						Creatinine (mg/dL)	2.5	H(1.5)
						Glucose (mg/dL)	128	H(125)
						Calcium (mg/dL)	7.5	L(8.5)
			Week 1	8	15.0	BUN (mg/dL)	34	H(25)
						Creatinine (mg/dL)	2.3	H(1.5)
						Calcium (mg/dL)	8.1	L(8.5)
			Week 2	15	22.5	BUN (mg/dL)	31	H(25)
						Creatinine (mg/dL)	2.4	H(1.5)
			Week 3	22	22.5	BUN (mg/dL)	34	H(25)
						Creatinine (mg/dL)	2.7	H(1.5)
						Glucose (mg/dL)	126	H(126)
						Calcium (mg/dL)	7.8	L(8.5)
						Phosphate (mg/dL)	1.6	L(2.5)

* RSD = Randomized Starting Dose of investigational Drug X.
+ F = High/Low Flag, (NR) = Applicable Normal Range, + indicates Local Lab Analysis

X:\XXXXXXXX\XXXXXXXX-000\Biostats\programs\TLFs\t-abnormch.sas 02JUL2013 19:09

Table 2 Better Listing Design

The net changes are one extra line in the column headers, the width of the body has gone from one hundred and forty six columns to eighty four columns and we have maintained the same number of footnote lines. The original listing was oversized when a readable non-proportional font is used, it fits into an acceptable size only with use of a proportional font. Such a font is no longer necessary in the updated listing, there is plenty of room for additions if they are desired.

EXAMPLE -- AN ADAM DATA LISTING

While most listings should contain data exclusively from SDTM domains, there are circumstances where derivations not part of SDTM should be displayed in a listing. One case to consider is when handling protocol deviations; the SDTM DV domain is expected to hold only those deviations that cannot be derived from the clinical database data. The programmable deviations are not collected as deviations in explicit form, they are often derived in the same program that produces the listing so they can seem to come out of nowhere.

We think a better process is to create a dataset for deviations, say an ADaM dataset ADDV which will contain both programmable and non-programmable deviations where the derivations can be reviewed. Although ADaM is not really intended for listings there is no restriction from using it for that purpose and if deviation summaries are desired they can then be readily produced from the ADaM dataset.

In this case, to be able to produce the listing below we will need information from several sources. Directly from SDTM.DV we get the non-derivable deviations, from SDTM.DM we get treatment, sex and investigator name. Derivable deviations may come from several SDTM domains such as CM, LB, DS, EX, etc.

In our hypothetical required listing in addition to sex and treatment we must display the following possible protocol deviations: *'Drug non-compliance'*, *'IWRS error'*, *'Use of med x combination not allowed'* which cannot be derived and will be available from SDTM.DV. Also the term *'Prohibited Medication'* needs to be derived from SDTM.CM and a list of medications explicitly forbidden in the protocol, the term *'Subject has discontinued from the study prior to 24 weeks'* is to be derived from SDTM.DS and the term *'Subject has missed two adjacent visits'* from SDTM.SV.

USUBJID	DVSEQ	DVTERM	DVDECOD	DVSTDT
TEST-001-00110	1	Drug non-compliance	TREATMENT DEVIATION	2010-10-25
TEST-007-00700	1	IWRS error	TREATMENT DEVIATION	2010-10-03
TEST-012-01204	1	Use of med x combination not allowed	PROHIBITED MEDICATION	2010-10-18

Table 3 Example of SDTM DV domain. Selected columns shown.

Our ADaM dataset will contain variables suitable to display sex, treatment, investigator, etc. from ADSL. It may also contain variables like ADVTERM and ADVDECOD, analysis counterparts of the corresponding variables from SDTM.DV. The investigator name could be included in the ADaM dataset or displayed in the listing using a format.

Row	STUDYID	USUBJID	DVSEQ	ADVTERM	ADVDECOD	TRT01P	SEX
1	TEST	TEST-001-00110	1	Drug non-compliance	TREATMENT DEVIATION	PLACEBO	F
2	TEST	TEST-007-00700	1	IWRS error	TREATMENT DEVIATION	PLACEBO	F
3	TEST	TEST-012-01204	1	Use of med x combination not allowed	PROHIBITED MEDICATION	PLACEBO	M
4	TEST	TEST-001-00107		Subject has discontinued from the study prior to 24 weeks	SCHEDULE DEVIATION	PLACEBO	M
5	TEST	TEST-004-00401		Prohibited Medication	PROHIBITED MEDICATION	ACTIVE	F
6	TEST	TEST-006-00601		Subject has missed two adjacent visits	SCHEDULE DEVIATION	PLACEBO	F

Table 4 ADaM Deviation Dataset. Selected columns shown.

Row	DVTERM	DVDECOD	AGE	TRTSDT	ASTDT	ASTDY
1	Drug non-compliance	TREATMENT DEVIATION	47	7/6/2010	10/25/2010	112
2	IWRS error	TREATMENT DEVIATION	61	5/3/2010	10/3/2010	154
3	Use of med x combination not used in accordance with product label	PROHIBITED MEDICATION	56	7/20/2010	10/18/2010	91
4			39	9/15/2010	11/9/2010	56
5			58	6/5/2010	8/13/2010	70
6			63	7/20/2010	10/11/2010	84

Table 5 ADaM Deviation Dataset (cont)

Listing 16.2.2
Listing of Protocol Deviations
Safety Population

Investigator	Subject Id	Treatment	Age Sex	Day	Deviation Term	Deviation Coded
Alvarado, D	601	PLACEBO	63F	84	Subject has missed two adjacent visits	SCHEDULE DEVIATION
Murray, J	107	PLACEBO	39M	56	Subject has discontinued from the study prior to 24 weeks	SCHEDULE DEVIATION
Collins, C	700	PLACEBO	61F	154	IWRS error	TREATMENT DEVIATION
Murray, J	110	PLACEBO	47F	112	Drug non-compliance	TREATMENT DEVIATION
Stevens, M	401	ACTIVE	58F	70	Prohibited Medication	PROHIBITED MEDICATION
Irving, A	1204	PLACEBO	56M	91	Use of med x combination not allowed	PROHIBITED MEDICATION

X:\XXXXXXXX\XXXXXXXX-000\Biostats\programs\TLFs\1-protdev.sas 02JUL2013 19:09

Table 6 Listing of protocol deviations

CONCLUSION

We hope that we have provided a little history behind one of the more obscure entities in clinical research, the lowly data listing, and given you some reasons why they should be taken a little more seriously than they are. Data listings can be the basis of some very valuable tools in your arsenal, but only if you consider them important contributors to the data review process and properly design/develop them. They do not exist by themselves; they should be populated with SDTM/ADaM data to maintain traceability, but most of all they should be easily read and make the clinical data more accessible by others.

REFERENCES

CDISC Study Data Tabulation Model (SDTM) v1.4 and Study Data Tabulation Model Implementation Guide (SDTMIG) v3.2

<http://www.cdisc.org/sdtm>

Analysis Data Model (ADaM) Implementation Guide (ADAMIG v 1.1)

<http://www.cdisc.org/adam>

CFR - Code of Federal Regulations Title 21 Part 314

<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?fr=314.50>

STRUCTURE AND CONTENT OF CLINICAL STUDY REPORTS E3

http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf

Dave Christiansen, *Regulatory Submission Datasets in the World of Evolving Standards*. 2015 ASA Biopharmaceutical Section Statistics Workshop

http://www.amstat.org/meetings/fdaworkshop/presentations/2005/p07_christiansen_cdisc.ppt

FDA Guidance for Industry. E3 Structure and Content of Clinical Study Reports Questions and Answers (R1)

<http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm336889.pdf>

CONTACT INFORMATION

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

Mario Widel:

Address: Lilly Corporate Center

City, State ZIP: Indianapolis, IN 46285

Work Phone: (317)459-5520

Fax:

E-mail: mwidel@lilly.com

Henry B Winsor:

Address: 100 Cardinal Way

City, State ZIP: Redwood City, CA 94063

E-mail: hwinsor@relypsa.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.