PharmaSUG 2018 - Paper BB-13

Report Compatibility: System to evaluate your code against volatile sources.

Randy Santiano, Eli Lilly; Joe Cooney, ICON; Craig Brelage, Eli Lilly

ABSTRACT

Developing programmed reports and queries via SQL, SAS, or R Code may take several cycles of development and resources. Although there is value with the final product, additional return on investment is obtained when all or portions of the report's code can be reused. However, variation in how data is collected across the clinical trial industry can reduce your organization's ability to reuse programmed reports and queries. This issue can be addressed by applying five basic principles to ensure that your organization has the ability to evaluate programmed report/query compatibility. This will allow for a proactive evaluation for the report code's reusability, and it will also support an environment where the users and developers are given transparency towards new specification development, allow for better workload management, and foster the ability for more accurate report/query catalog planning.

INTRODUCTION

Reviewing data collected from clinical trials is important to maintaining data integrity, improving the accuracy of statistical analysis, and ensuring compliance from all parties involved. The task of creating reports based on Data Verification Plans is normally given to the Data Management Team in partnership with other functions involved in data review (e.g. Medical review teams or third party vendor). Data Verification Plans may contain a combination of review requirements such as exception events, data reconciliation, summary listings, and vendor oversight. Typically, each of the review items will require a report or data listing to be created for each trial.

Typically, study teams generating the Data Verification Plan submit report development requests for each individual study. This creates an environment where unique reports are created for each new trial and there is little to no reuse of reports across studies. This practice leads to multiple inefficiencies and variations in how reports are constructed. Examples include estimations for hours of report development will vary between studies, variations in report specifications for common Data Verification Plan items with impact on the quality and usability for each report, and potential interpretation by each Study Team and supporting Report Programmers.

As the programmers responsible for creating these reports grow in experience, the most

common observations for the code used between studies are as follows:

- There are similarities between Data Verification Plans from the same Therapeutic Area
- There are common sets of data verification specifications, based on defined Clinical Domains
- Many Studies reference the same data sources
- The types of data between Studies are similar

The most logical approach in reducing the turn-around-time for reports needed to execute Data Verification Plan data reviews is to have sets of reusable code. However, the issues with developing and maintaining a library of reusable code can be attributed to any of the following study design circumstances:

- Each study setup may be developed and implemented independently, leading to an increased chance of study variability
- Each study team may develop unique Data Verification Plans to address a Therapeutic Area specific property
- Newer studies may introduce different data sources
- The Medical Review Team may define report specifications outside of the Data Verification Plan
- Post production changes to the study design occur often

A PRINCIPLED APPROACH TO REPORT COMPATIBILITY

Each of these unpredictable situations characterized the environmental volatility that reduces the ability to have a library of reusable code. Although these issues may be seen as a natural part of the clinical trial environment, there are a few basic principles that can be adopted to lessen the effects of the environmental volatility.

Principle #1: Identify Your Common Report Themes

Establishing a standard set of common Data Verification Plan Targets to support the FDA submission process is not a new concept, but often a difficult one to achieve and maintain. What is often overlooked is developing sets of branching reporting themes to accommodate different case report form designs. If the reporting themes can be based on Therapeutic Area Domain targets, then it may be easier to implement common reporting themes across studies.

For example, the branching structure for Adverse Event information Case Report Form designed is often influenced by what is relevant to the study drug or Therapeutic Area. The Adverse Event data is also often chronologically associated with the subject's Medical History and correlated with their concomitant medications. Adverse Event related Data Verification Plan Targets

are based on identifying conditional exceptions, chronological issues, and logical discrepancies.

By organizing each of these reporting themes into common groups, the code that generates the reports for these groups can be catalogued into library themes.

Below are example for reports themes that can be created:

- If there is a report that is designed to identify discrepancies with Adverse Event Statuses, then designate that report at AE101.
- If there is a report that is designed to identify overlaps between Medical History and Adverse Events, then designate that report as AE102
- If there is a report that is designed to have information for both Adverse Event Status discrepancies and Medical History overlaps, then designate that report as AE101_AE102.
- If there are study design difference between Therapeutic Area 1 and 2 for the same AE101_AE102 report, then distinguish between the two reports by differentiating the report name to TA1_AE101_AE102 and TA2_AE101_AE102.
- If there are two report design variations for the AE101_AE102 report for Therapeutic Area 1, then distinguish between the reports by differentiating the report names to TA1_AE101_AE102 and TA1_AE101_AE102_VAR1.
- If STUDY_X for Therapeutic Area 1 has a unique design configuration for the AE101_AE102 report, then segregate the report to a STUDY_X folder.

Below is an example for independent, but similar reports themes that can be combined:

- If there is a report that is designed to identify gender-based discrepancies in according to their Adverse Event and Medical History event terms, then designate that report as AE103_MH103.

The structure of the library will start to take shape with every report defined by the Study Team. Below is an example of the report hierarchy structure based on the report theme examples.

AE → AE101
AE102

AE101_AE102
AE103_MH103

AE → TA1 → TA1_AE101_AE102
AE → TA1 → TA1_AE101_AE102_VAR1
AE → TA1 → STUDY_X → TA1_AE101_AE102

AE → TA2 → TA2_AE101_AE102

The benefits for establishing a common reporting theme not only creates a library structure that is easy to understand, but can also accommodate different report combinations. The report hierarchy can also serve as a way to define commonly reusable reports, Therapeutic Area reports, and Study Specific reports.

Principle #2: Support a Standards Library

As the study design often dictates the study's database schema, it should be expected that each study would bring new challenges to report designs. Most of the report design challenges stem from the development of different Case Report Form configurations and identification of new sources of data, such as different electronic Clinical Outcomes Assessment (eCOA) or Patient Diaries.

The development and use of study build templates is instrumental to driving standardization, where the responsibility for decisions made for managing such a library is best given to a governing body. The purpose of the governing body would be to ensure consistency regarding decisions made for the study build templates and Case Report Form library components and help ensure more standardization at the study level.

Although it sounds simple to implement a Standard Case Report Form Library and Study Build Templates, many factors complicate the governance and maintenance of such a library. Below are just a few examples of when Case Report Form Library or Study Build Template components may require updates or additions to accommodate the following:

- Forms Design may be modified to reduce Investigator Site Confusion.
- Current Form Designs may be split or merged to accommodate different collection schemas.
- New Form Designs may be developed to accommodate different disease states or stages.
- Changes to one primary Form Design may cause cascading changes to subsequent forms.
- Internal and/or external standards may evolve

Even though any of these changes will cause redevelopment to reports supporting the Data Verification Plan and corresponding Oversight Reports, there are a few considerations to help mitigate the downstream effects. Below are a few examples, with comments towards the benefits and disadvantages:

- For a Governing Body to define details published about what are acceptable changes for every Case Report Form and Study Build Templates.
  Benefit: to establish a process of control, which would limit variations.
  Disadvantage: the possibility of longer processing timelines for newer studies with a unique study setup.

- Maintain alignment between common Study Build Templates to Data Verification Targets.
  Benefit: the establishment of a set of Core Study Build Templates.
  Disadvantage: the possibility of increasing the number of Forms per Study.

- Reduce the removal of variables when modifying the Study Build Template, when possible.
  Benefit: to accommodate reports downstream from modifications.
  Disadvantage: increasing amount of variables that do not contain data, which may affect the intent of several reports downstream and increase study notation.

- Develop new Data Verification Targets when designing new Study Build Templates.
  Benefit: to manage expectations towards new report development.
  Disadvantage: the possibility of increasing the number of reports per study, with possible consolidation needed after a certain point in time.

A library of Standard Study Templates would allow for a direct Data Verification Plan association. The example below showcases how a Study Build template would be associated with various references and version updates in addition to which study where the Data Verification Plan is used.

TABLE 1: Study Build Template Alignment

Therapeutic Area	Template Name	Template Version	DVP Reference	Study A	Study B	Study C	Study D
TA1	AE_1000	1.0	AE101	1	1	0	0
TA1	AE_1000	1.0	AE102	1	1	0	0
TA1	AE_1000	1.0	AE103	1	1	0	0
TA1	AE_1000	2.0	AE101	1	1	0	0
TA1	AE_1000	2.0	AE102	1	1	0	0
TA1	AE_1000	2.0	AE103	1	0	1	0
TA2	AE_1000	1.0	AE101	0	0	1	0
TA2	AE_1000	1.0	AE102	0	0	1	0
TA2	AE_1000	1.0	AE103	0	0	0	0
TA3	AE_1000	1.0	AE201	0	0	0	1

Principle #3: Convert Report Code Specifications into a Database

When designing reports, a developer often references some type of formal documentation that outlines the report specifications. Adding notation and report header information into the code is a common practice for many code developers. Per most validation standard operating procedures, the completed validation package would contain the report code specifications, the report code with header information, and a document to acknowledge the completion of the report validation process.

What is not common is a process to convert the code specifications into a standard table format that can be entered into a central database. Even if the programming software can generate the table schema from the code, there are additional information that about the code that can only be identified by the developer as it related to the design of the report.

For example, a report for Therapeutic Area 1 for Study A with an identity of AE104 had been requested to look for discrepancies between Panel X, Panel Y, and Panel Z. The report specifications gives the developer general statements and instructions on how to combine the tables and identify the discrepancies.

Below is an example of the SQL code that the developer would use to generate the report titled TA1_AE104:

```
SELECT x.SITE, x.SUBJID, x.TERM, x.CODELIST, x.STATUS, y.START_DATE,
z.END_DATE, z.NOTES

        CASE WHEN x.TERM = y.TERM THEN 'OK'
                WHEN x.TERM = z.TERM THEN 'DUPLICATE'
                WHEN x.TERM is Null THEN 'ERROR'
                ELSE 'PENDING'
                END as ERROR_MESSAGE

        FROM PANEL_X x, PANEL_Y y, PANEL_Z z
        Where x.SUBJID=y.SUBJID(+) and x.SUBJID=z.SUBJID(+) and
x.STATUS is not null
```

The SQL code would be translated to the following Code Specification Table:

TABLE 2: Report Code Specification Table

Therapeutic Area	Report Name	Report Description	Version	Completion Date	Completed by	Relevance	Panel	Variable
TA1	AE104	AE104: Discrepancies between Panel X, Panel Y, and Panel Z	1.0	08MAR2018	John Smith	CORE	PANEL_X	SITE
TA1	AE104	AE104: Discrepancies	1.0	08MAR2018	John Smith	KEY	PANEL_X	SUBJID

Therapeutic Area	Report Name	Report Description	Version	Completion Date	Completed by	Relevance	Panel	Variable
		between Panel X, Panel Y, and Panel Z						
TA1	AE104	AE104: Discrepancies between Panel X, Panel Y, and Panel Z	1.0	08MAR2018	John Smith	CASE	PANEL_X	TERM
TA1	AE104	AE104: Discrepancies between Panel X, Panel Y, and Panel Z	1.0	08MAR2018	John Smith		PANEL_X	CODELIST
TA1	AE104	AE104: Discrepancies between Panel X, Panel Y, and Panel Z	1.0	08MAR2018	John Smith	LOGIC	PANEL_X	STATUS
TA1	AE104	AE104: Discrepancies between Panel X, Panel Y, and Panel Z	1.0	08MAR2018	John Smith		PANEL_Y	START_DATE
TA1	AE104	AE104: Discrepancies between Panel X, Panel Y, and Panel Z	1.0	08MAR2018	John Smith	LOGIC	PANEL Y	TERM
TA1	AE104	AE104: Discrepancies between Panel X, Panel Y, and Panel Z	1.0	08MAR2018	John Smith		PANEL_Z	END_DATE
TA1	AE104	AE104: Discrepancies between Panel X, Panel Y, and Panel Z	1.0	08MAR2018	John Smith	LOGIC	PANEL_Z	TERM
TA1	AE104	AE104: Discrepancies between Panel X, Panel Y, and Panel Z	1.0	08MAR2018	John Smith		PANEL_Z	NOTES

The table also provides an easier way for a non-programmer to understand the report with respect to the information collected from the Case Report Form.

In summary, the intent of capturing details about a developed report outside of the code specifications is to create a referential database of reports and information elements. The utilization of this table is best described in Principles #4 and #5.

Principle #4: Use the Report Code Specifications in multiple Environments

When the Report Code Specifications are uploaded into a central database of reports, we can now evaluate the report details across different table schemas for multiple studies of interest.

TABLE 3: Report Compatibility Review

Therapeutic Area	Report Name	Relevance	Panel	Variable	Study A	Study B	Study C	Study D
TA1	AE104	CORE	PANEL_X	SITE	1	1	1	1
TA1	AE104	KEY	PANEL_X	SUBJID	1	1	1	1
TA1	AE104	LOGIC	PANEL_X	TERM	1	1	1	1
TA1	AE104		PANEL_X	CODELIST	1	1	1	1

Therapeutic Area	Report Name	Relevance	Panel	Variable	Study A	Study B	Study C	Study D
TA1	AE104	LOGIC	PANEL_X	STATUS	1	1	1	1
TA1	AE104		PANEL_Y	START_DATE	1	1	0	0
TA1	AE104	LOGIC	PANEL Y	TERM	1	1	1	0
TA1	AE104		PANEL_Z	END_DATE	1	1	0	0
TA1	AE104	LOGIC	PANEL_Z	TERM	1	1	1	0
TA1	AE104		PANEL_Z	NOTES	1	1	1	0

The table above describes three different scenarios that can be used to evaluate report compatibility and estimate the work needed to develop a report variant, where a “1” indicates that the target variable is found in the study and “0” is when the variable is not present.

Scenario #1: Report AE104 can be used in Studies B, as it has the same Panels and Variables as Panel A.

Scenario #2: Since the report had identified several missing variables for Study C, the Report AE104 cannot be used unless modification are made. The development for this report is minimal, as all of the missing variables were not relevant to the report. Additional investigations conclude that Study C does not collect the Start and Stop Dates and can be omitted or replaced from the report. A variant of the report can be created and named as TA1_AE104_VAR1.

Scenario #3: Report AE104 cannot be used in Study D, as Table 2 identifies that Panel Y and Z are missing. Since some of the relevant Panels and Variables for the report are missing, the report would have to be completely redesigned once the appropriate panels are found and/or the report specifications do not apply for this study.

Trends across different studies can be observed with considerations to library planning and general estimations for workload if similar reports are required for each Study using this report compatibility information. The information for report compatibility can also be used to guide future Study Designs and provide a case for Study Build Consistency.

For example, if Study E is associated with Therapeutic Area 1, then the study teams can be given information on a set of existing reports associated with the same or similar Therapeutic Area.

TABLE 4: Report Compatibility Summary Review

Therapeutic Area	Report Name	Study A	Study B	Study C	Study D	Study E
TA1	AE104	YES	YES	NO	NO	?
TA1	AE104_VAR1	NO	NO	YES	NO	?
TA2	AE104	NO	NO	NO	YES	?

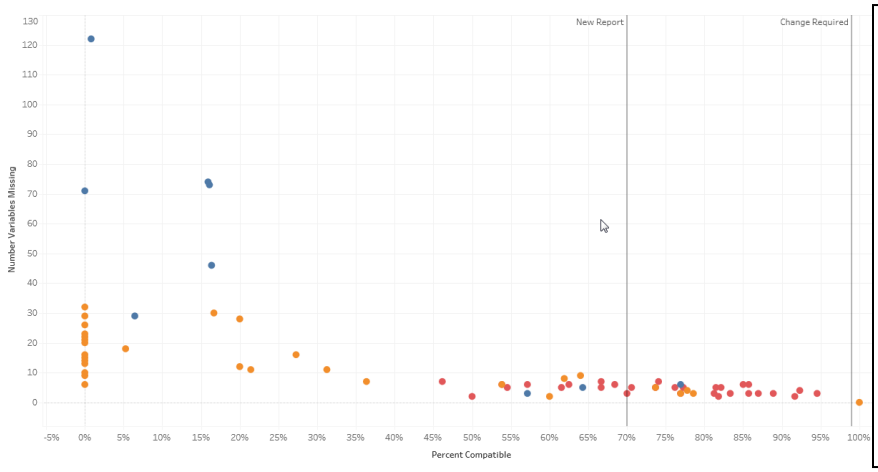
In Table 4, an assumption can be made that Report AE104 may be compatible if Study E is similar to Study A and B.

By having this type of report review process available to the study team, they can quickly evaluate report compatibility across one to multiple reports within the library and assess the workload potential based on evidence. As the library of reports gets larger, having a way to evaluate report compatibility across all reports for any given Study provides insight into implications such as:

- Study design impact on report reusability
- Consistent use of and points of departure from study build templates
- Distribution of report compatibility across different Therapeutic Areas.
- Reusable reports compatibility across different Therapeutic Areas.
- How many existing reports are compatible for a new Study.

For Example, if Study E is associated with Therapeutic Area 1, then a data visualization can be generated to reference the Report Library for reports that are compatible for that Study.

Graph 1: Review of Study E for Therapeutic Area 1



Graph 1 Notes:
X axis represent the variables present for Study E over the appropriate report variables. A percent is calculated based on variables found that match the report.

Y axis represent the number of report variables missing per individual reports.

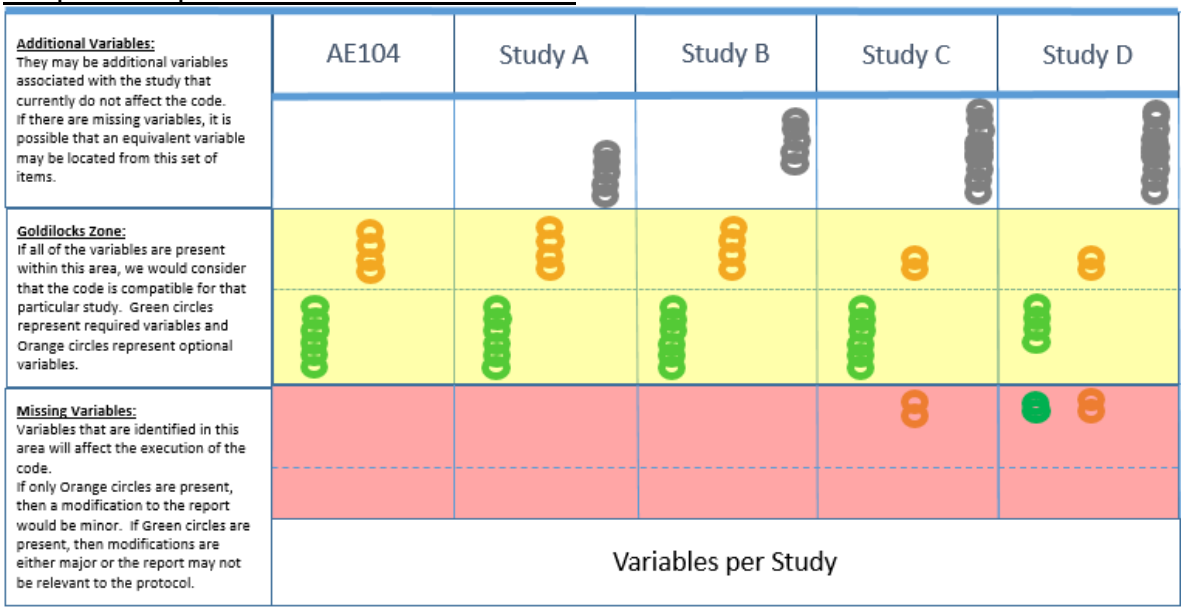
3 dots represent 3 different Therapeutic Area Type reports.

If the red dots represent reports that were specifically created for TA1, then the graph indicates that there will be some development work will be required to have reports compatible for Study E. It also seems that there may be a few other Therapeutic area reports that can be reused, as the blue and orange represent TA2 and TA3 respectively.

The details from the report code specifications can be used to further evaluate modifications to the report. The Graph below is a visual representation of Table 3, where the variables can be grouped by their relevance. The graph introduces the concept of a “Goldilocks Zone”, which is a reference to the children’s story of

Goldilocks and the Three Bears. In this "Goldilocks Zone" graph, a code developer can describe a region to the Users that is "Just Right" for updates to the code that require minimal effort. Any variables that fall above or below this zone may be "too Hot" (i.e. complex redevelopment) or "too Cold" (i.e. insignificant changes to the code and logic, but should be avoided to maintain consistency).

Graph 2: Report AE104 Goldilocks Zone



Considerations to the modifications of the code for report AE104 can be presented to the study teams in addition to workload expectations for development hours if the same report is required for groups of studies.

Principle #5: Develop Proactive Feedback Cycle

Change management is not only a difficult process, but can also be tedious to maintain across all affected areas. As systems are becoming more interdependent, it is often a challenge to understand the downstream effects for any one given change. The process is often very tedious, manual, and requires a lot of resources to maintain and manage per change.

The process of accommodating the changes downstream from the event is often a reactive activity, as updates are often identified after the issues propagate throughout the system.

Having a system that can identify areas affected in real-time can only be possible if the overall system has one common property or table. With the use of a central code specification table, the organization can evaluate their report library across multiple environment and use data visualization techniques to communicate

report or template compatibility as well as downstream effect of changes to the code or template.

The example below outlines a situation where a post-production change requires the creation of a new template. Table 5 describes the original setup.

TABLE 5: Study Built Template and Reporting Alignment

Therapeutic Area	Template Name	Template Version	DVP Reference	Report Reference	TA1 Study A	TA1 Study B	TA2 Study C	TA3 Study D
TA1	AE_1000	1.0	AE101	TA1_AE101_AE102	1	1	0	0
TA1	AE_1000	1.0	AE102	TA1_AE101_AE102	1	1	0	0
TA1	AE_1000	1.0	AE103	TA1_AE103_MH103	1	1	0	0
TA2	AE_1000	1.0	AE101	TA2_AE101_AE102	0	0	1	0
TA2	AE_1000	1.0	AE102	TA2_AE101_AE102	0	0	1	0
TA2	AE_1000	1.0	AE103	TA2_AE103	0	0	1	0
TA3	AE_1000	1.0	AE201	TA3_AE201	0	0	0	1
TA1	MH_1000	1.0	MH103	TA1_AE103_MH103	1	0	0	0
TA2	MH_1000	1.0	MH103	TA2 _MH103	0	0	1	0
TA3	MH_1000	1.0	MH103	TA3_AE103_MH103	0	0	0	1
GLOBAL	DM_1000	1.0	DM100	DM100	1	1	1	1

Table 6 shows the addition of new study templates and outlines the affected Report Reference with respect to the Studies.

TABLE 6: Study Build Template and Reporting Alignment with Version Update

Therapeutic Area	Template Name	Template Version	DVP Reference	Report Reference	TA1 Study A	TA1 Study B	TA2 Study C	TA3 Study D
TA1	AE_1000	1.0	AE101	TA1_AE101_AE102	1	1	0	0
TA1	AE_1000	1.0	AE102	TA1_AE101_AE102	1	1	0	0
TA1	AE_1000	1.0	AE103	TA1_AE103_MH103	1	1	0	0
TA2	AE_1000	1.0	AE101	TA2_AE101_AE102	0	0	1	0
TA2	AE_1000	1.0	AE102	TA2_AE101_AE102	0	0	1	0
TA2	AE_1000	1.0	AE103	TA2_AE103	0	0	1	0
TA2	AE_2000	2.0	AE103	<missing>	0	0	1	0
TA3	AE_1000	1.0	AE201	TA3_AE201	0	0	0	1
TA1	MH_1000	1.0	MH103	TA1_AE103_MH103	1	0	0	0
TA2	MH_1000	1.0	MH103	TA2 _MH103	0	0	1	0
TA2	MH_2000	2.0	MH103	<missing>	0	0	1	0
TA3	MH_1000	1.0	MH103	TA3_AE103_MH103	0	0	0	1
GLOBAL	DM_1000	1.0	DM100	DM100	1	1	1	0
TA3	DM_2000	1.0	DM101	<missing>	0	0	0	1

An automated communication to the affected reporting groups can be initiated by the system, prompting for a response and evaluation of the new study templates.

Table 7 represents the completion of a change management cycle, where the new report references have been added.

TABLE 7: Study Built Template and Reporting Alignment Full Update

Therapeutic Area	Template Name	Template Version	DVP Reference	Report Reference	TA1 Study A	TA1 Study B	TA2 Study C	TA3 Study D
TA1	AE_1000	1.0	AE101	TA1_AE101_AE102	1	1	0	0
TA1	AE_1000	1.0	AE102	TA1_AE101_AE102	1	1	0	0
TA1	AE_1000	1.0	AE103	TA1_AE103_MH103	1	1	0	0

Therapeutic Area	Template Name	Template Version	DVP Reference	Report Reference	TA1 Study A	TA1 Study B	TA2 Study C	TA3 Study D
TA2	AE_1000	1.0	AE101	TA2_AE101_AE102	0	0	1	0
TA2	AE_1000	1.0	AE102	TA2_AE101_AE102	0	0	1	0
TA2	AE_1000	1.0	AE103	TA2_AE103	0	0	1	0
TA2	AE_2000	2.0	AE103	TA2_AE103_MH103	0	0	1	0
TA3	AE_1000	1.0	AE201	TA3_AE201	0	0	0	1
TA1	MH_1000	1.0	MH103	TA1_AE103_MH103	1	0	0	0
TA2	MH_1000	1.0	MH103	TA2 _MH103	0	0	1	0
TA2	MH_2000	2.0	MH103	TA2_AE103_MH103	0	0	1	0
TA3	MH_1000	1.0	MH103	TA3_AE103_MH103	0	0	0	1
GLOBAL	DM_1000	1.0	DM100	DM100	1	1	1	0
TA3	DM_2000	1.0	DM101	TA3_DM101	0	0	0	1

This system can also provide the governance over the Study Build Template the ability to simulate different change scenarios, leading to a data driven forecasting and guidance model.

CONCLUSION

Volatility in data standardization in the highly dynamic clinical trial environment and competitive pharmaceutical industry should always be expected. Fortunately, we have the tools necessary to accommodate these situations. Each principle described in this document can be applied independently, with gains in efficiency achieved as each is implemented and applied with the appropriate intent.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please contact the co-authors at:

Randy Santiano
Associate Consultant-Process/Technology
Eli Lilly and Company
Lilly Corporate Center, Indianapolis IN 46285 U.S.A.
+1 317-433-9767 (office)
randy.santiano@lilly.com | www.lilly.com

Joseph Cooney
Senior Clinical Data Programming Lead
ICON Clinical Research
Internal Tel: 01 6774
External Tel: +1 215 616 6774
Joseph.Cooney@iconplc.com | www.iconplc.com

Craig Brelage
Associate Consultant - Global Data Delivery, Clinical Data Flow and Technology
Eli Lilly and Company
317-277-1240 (Office)
brelage_craig_r@lilly.com | www.lilly.com