PhUSE US Connect 2018

Paper SI11

Using the TFL Workbench for Standards Governance

Hansjörg Frenzel, PRA Health Sciences, Germany
Iain Humphreys, PRA Health Sciences, UK

ABSTRACT

The TFL Workbench has been developed by PRA Health Sciences to support programming of standard tables, figures, and listings (TFLs), and is currently used to produce tables based on standard table shells. A compliance report is available and used by PRA's Global Data Standards team to gather metrics about standards usage. These metrics currently report the percentage of tables in a study produced by the TFL Workbench, the percentage of TFL Workbench templates that were altered in order to comply to the TFL shells, and whether any pre- or post-processing is necessary to create tables. The paper discusses how the TFL Workbench can support governance of a TFL standard, and the future potential of the compliance report, in helping Governance Committees to decide which standard data displays need updates.

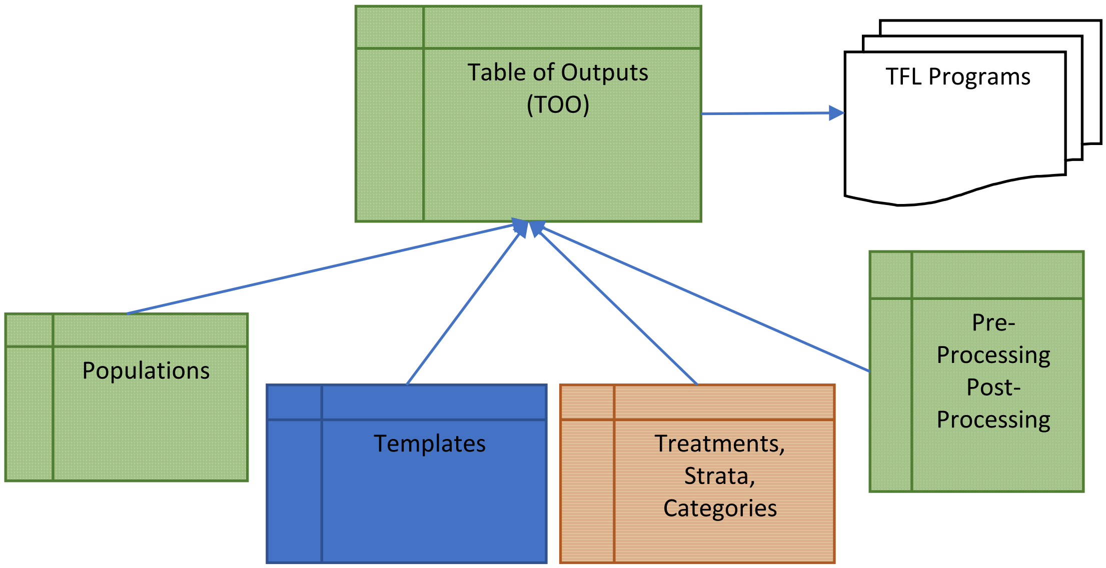
INTRODUCTION

PRA Health Sciences has developed a TFL standard which is available to clients who do not have a dedicated standard. This is done with the intention of raising the quality and consistency of data displays used for their clinical study reports, and to support our Biostatisticians in the development of TFL shells. Quality gains and consistency are accomplished by using standard programs for data displays (tables, figures, and listings). For this to be successful, the resulting TFL Shells have to be sufficiently consistent with the standard as well as amongst themselves. But of course, the standard has to be flexible enough to allow for changes based on study design, scientific needs, and specific data. Finally, standard usage has an impact on costs – the more closely a study follows a standard, the less manual intervention is required, which results in lower effort for creating and reviewing data displays. The converse is also true.

PRA has also developed the TFL Workbench, a tool that is intended to write SAS® programs for tables, figures, and listings based on machine-readable metadata. The TFL Workbench provides programming teams with libraries of configurations based on PRA's own or sponsor-specific TFL standards, which they can download for their study. These can be modified to accommodate changes that are dictated by study data and methodology. Changes to these configurations can be made comparatively easily, but the tool is resilient enough to prevent unwanted dissimilarities. In studies where standards are not available, inconsistency can easily creep into data displays because of inhomogeneous shells or bespoke programs. The TFL Workbench seeks to overcome this by building TFL programs from a combination of reusable sets of configurations, which remain invariant from one output to another. In its current state, the TFL Workbench is capable of writing programs for summary tables only.

THE TFL WORKBENCH

The TFL Workbench is a database that stores machine-readable metadata about the structure of summary tables and writes table programs utilizing these configurations. The configurations can be changed by study teams to adapt the layout and contents of tables to the needs of an individual study. It requires building tables from distinct sets of configurations, describing their vertical structure (called "templates"), their columns (usually representing treatment groups), as well as different sets of strata, should they be required. Each template available in the TFL Workbench describes a unique type of table (e.g. disposition, adverse events, demography, laboratory results). The configuration of treatment columns is stored independently from templates in treatment sets. Stratification can be applied using strata sets. Populations are defined in their own set. Pre- and post-processing was added so that programmers gain more flexibility in merging data or updating the formatting of the results. The central database table, called "Table of Outputs" or "TOO" for short, is the place where distinct sets of configurations are combined to create tables.



Consistency is achieved when similar tables use the same template, e.g. all summaries of adverse events by system organ class and preferred term use the template available for this type of table. Idiosyncratic differences between tables of the same type would require multiple templates, each implementing miniscule changes. Similarly, all tables requiring a certain configuration of treatment columns will share the same treatment set. This way, a set of metadata needs to be configured once and can be re-used multiple times. In contrast, SAS® programs often diverge from one another by using different methods to summarize data or having different structures or applying formatting differently. In this way inconsistency is easily introduced, and it requires much effort to detect and resolve those inconsistencies during QC and review. When using the TFL Workbench, a higher level of consistency is achieved from reusing configurations where appropriate and required.

THE COMPLIANCE REPORT

TFL Workbench comes with a compliance report, which compares study-specific configurations back to the standard they originate from. The report is used by study teams to see where changes to the standard settings have been made, and this guides their review. The report is created by validated SAS® macros which produce a Microsoft Excel® workbook, with tabs for each distinct type of configurations, plus one tab per template. Each tab except the TOO presents the original standard configurations in the upper half of the worksheet and the study-specific configurations below that. Changes at the study level are presented by highlighting the cell that contains the modification: data-related changes are regarded as potentially more critical than formatting changes, and deserving of particularly close attention, since they are likely to have a direct bearing on the numerical results being displayed in a summary.

	A	B	C	D
1	Sour	Popul-ation	Where Clause	Data Source
2	1	All Randomized	randfl eq 'Y'	ADSL
3	1	Full Analysis Set	fasfl = 'Y'	ADSL
4	1	Per Protocol Analysis Set	pprotfl = 'Y'	ADSL
5	1	Safety Analysis Set	saffl = 'Y'	ADSL
6	2	All Randomized	randfl eq 'Y'	ADSL
7	2	Full Analysis Set	fasfl = 'Y'	ADSL
8	2	Per Protocol Analysis Set	pprotfl = 'Y'	ADSL
9	2	Safety Population	saffl in ('Y', ' ')	ADSL

The example shows modifications to a population configuration. In the upper half is the standard configuration for these populations, in the lower half the study configuration. The changes to the standard are made visible by highlighting the modified cells: the label of the safety population was changed from "Safety Analysis Set" to "Safety Population" (this is a cosmetic change), while the where clause to filter appropriate records in the datasets used is changed from "saffl = 'Y'" to "saffl in ('Y', ' ')" (this is a data change). Similar displays are available for other elements: treatment and strata sets, pre- and post-processing code, and each individual template used in the study.

The TOO tab of the compliance report is different, as it details which instances of the above elements were used (in columns Template ID, Population, Treatment Set, Strata Set, Category set, Pre- and Post Proc) to create a specific table. The highlighting indicates which tables were built on unmodified or modified configurations, and whether format or data changes were driving the change to the configuration of the elements.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Output ID	Template ID	Popul- ation	Trt Se	Trt Cc	Strata Se	Strata Cc	Categ Se	Categ Cc	Row Label	Data Sourc	Tot.Exp. Var	Table Filter	Treat Var	Pre- Proc	Post- Proc
2	14-01-01-01	PRA-STD-T-001	All Randomized	STD1	1		0		0		ADSL			TRT01P	_PreProc1_	_PostProc1_
3	14-01-01-03	PRA-STD-T-003	Full Analysis Set	STD1_N	1		0		0		ADSL			TRT01P		
4	14-01-02-01	PRA-STD-T-006	Full Analysis Set	STD1_N	1		0		0		ADSL			TRT01P		
5	14-01-02-02	PRA-STD-T-006	Per Protocol Analysis Set	STD1_N	1		0		0		ADSL			TRT01P		
6	14-01-02-03	PRA-STD-T-006	All Randomized	STD1_N	1		0		0		ADSL			TRT01A	_PreProc1_	
7	14-01-02-04	PRA-STD-T-006	Safety Population	SAF1	1	SEX_N	1		0		ADSL			TRT01A		
8	14-01-02-05	PRA-STD-T-006	Safety Population	SAF1_N	1	SEX	0		0		ADSL			TRT01A		
9	14-03-02-01	PRA-STD-T-013	Safety Population	SAF1_N	1		0		0		ADAE		TRTEMFL = 'Y'	TRTA		
10	14-03-02-02	PRA-STD-T-014	Safety Population	SAF1_N	1		0		0	Any Adverse Events	ADAE		TRTEMFL = 'Y'	TRTA		
11	14-03-02-06	PRA-STD-T-014	Safety Population	SAF1_N	1		0		0	Any Treatment-Related Adverse Events	ADAE		TRTEMFL = 'Y' and RELGR1N = 1	TRTA		
12	14-03-02-12	PRA-STD-T-014	Safety Population	SAF1_N	1		0		0	Any Serious Adverse Events	ADAE		TRTEMFL = 'Y' and AESER = 'Y'	TRTA		
13	14-03-02-14	PRA-STD-T-014	Safety Population	SAF1_N	1		0		0	Any Treatment-Related Serious Adverse Events	ADAE		TRTEMFL = 'Y' and AESER = 'Y' and RELGR1N = 1	TRTA		
14	14-03-03-01-07-opt1	PRA-STD-T-025	Safety Population	SAF2	1		0	SUB1	1		ADLB		PARCAT1N = 1 and ANL01FL = 'Y'	TRTA		
15	15-01-01	PRA-STD-T-006	Per Protocol Analysis Set	STD1_N	1	AGE	1		0		ADLB		missing(filter)	TRT01P	_PreProc1A_	_PostProc2_

The example of a TOO shows that some of the configurations used to write a table program were modified. In column "Template ID" cell highlighting is used to indicate use of modified templates (darker cells indicate data changes). Column "Population" (Column C) reports which tables use an unmodified or modified population configuration, with the modified population sets being presented using cell highlighting. Similarly, "Treatment Sets" (Column D) and "Strata Sets" uses highlighting to report which treatment sets were modified (analogous to changes in populations). From the example above one can also see that the TOO contains more metadata than just the configuration sets described above: most importantly which dataset feeds into a specific table, how the dataset is filtered (in addition to the filter defined in the population set and the treatment/strata sets), and the name of the treatment variable in the dataset. But because this information is highly study dependent, it is not compared to configurations in the standard database.

The compliance report is an essential tool that allows study teams to see which changes they made to standard configurations. Because it is built to compare study configurations to standard configurations, it is also the source for metrics about the usage of the standard.

THE LIST OF METRICS WE COLLECT

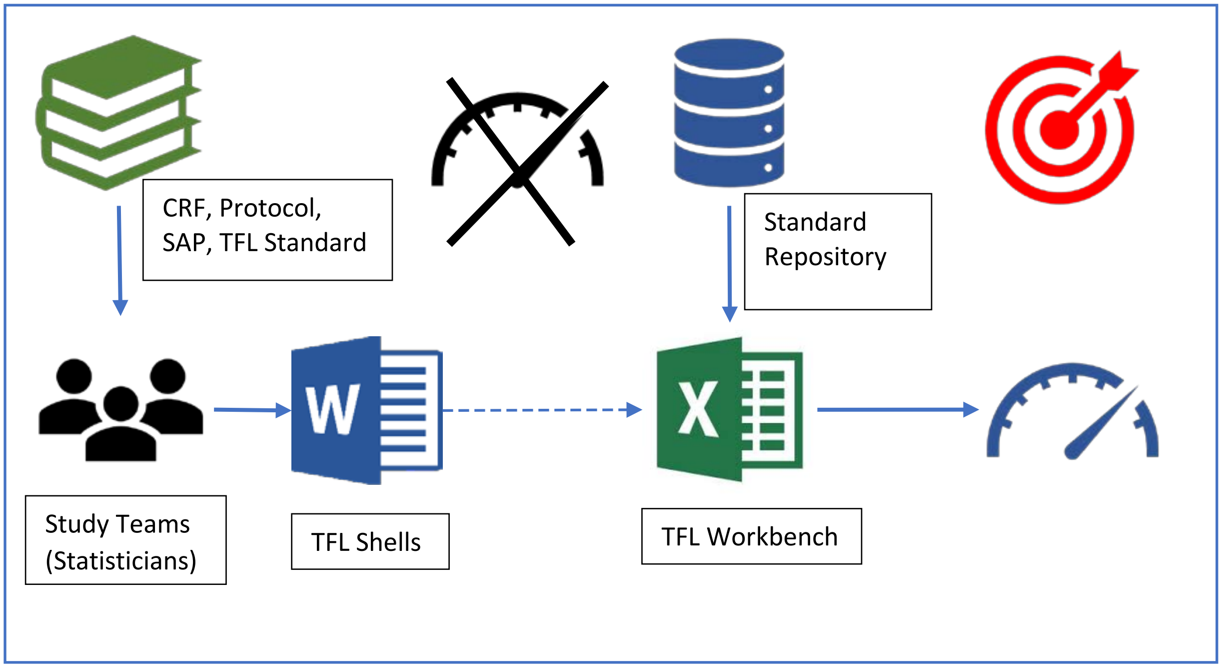
Beside being nicely formatted documents with content every user of the standard agrees upon, it is important to ensure that standards are used in daily work, to enhance quality and reduce effort in creating data displays. Knowing to which extent standards are used, and how often standard elements must be modified by study teams to be useful for a study, is important in understanding how good a standard is and how it impacts quality and consistency of the deliverable and its costs. Both, improving quality and consistency and lowering costs are achieved by using standard programs. To find out how useful a standard is, data must be available to answer the following questions:

- How often is a standard used? – What is the percentage of clinical studies in which the standard was used compared to all clinical studies eligible for the standard
- What is the proportion of outputs that are based on standard shells?
- What is the proportion of TFLs created by the TFL Workbench based on standard shells vs custom shells?
- How many TFLs required modification of standard shells?
- Which are the standard shells used the most, and how often are they changed?
- What did change in individual shells?

We cannot answer most of these questions directly, but we can utilize the TFL Workbench to help reach an answer.

HOW WE MEASURE

PRA Statisticians broadly follow the following flow diagram when creating TFL Shells; a process that is almost ubiquitous:

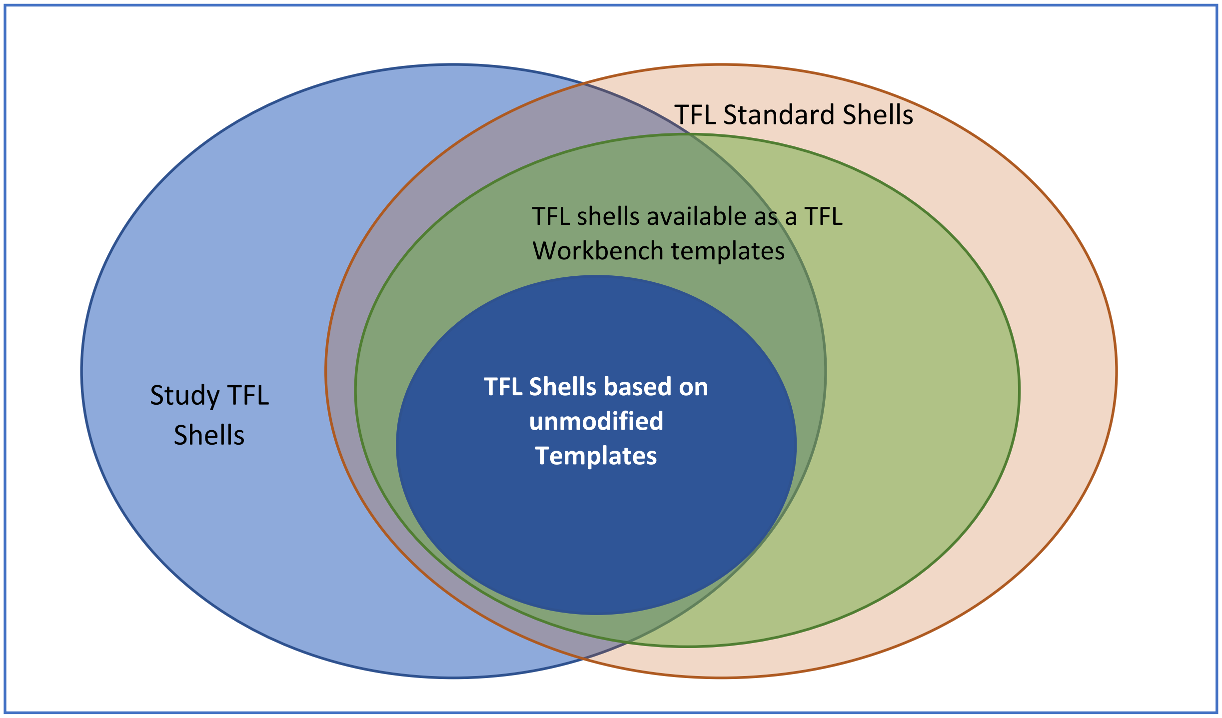


Sponsor representatives and PRA Statisticians agree upon the TFL shells, based on CRF, study protocol, and statistical analysis plan (SAP). Although the TFL shells are based on a template resembling the standard, teams can deviate if the analysis of the study data requires it. They should not, however, deviate because of their personal preference. The final product – TFL shells for a study – might only loosely resemble the standard. But because traditionally TFL Shells are usually created in Microsoft Word®, manual review and comparison would be the only way of obtaining useful metrics data at this stage: a labor-intensive manual task with unreliable results. We do not collect our metrics by manually comparing TFL shells for a study to standard shells.

Because it provides data about the usage of standards as a by-product of producing data displays, we can use the TFL Workbench to gauge how a standard is used. How does the TFL Workbench help determining how many templates were changed? As mentioned above, TFL Workbench configurations can be – once implemented for a study – compared to the standard TFL Workbench configurations via use of the compliance report. But there are limitations imposed by this method: out of all the shells Statisticians create for a study, only a subset might come from the TFL standard shells. Other data displays might be added to present data for which no standard template is available, or they may require a statistical technique or a graphic that is not represented in a standard. Only rarely can 100% of the TFL shells for a study be built using standard shells, even when they are modified to conform to study requirements.

Furthermore, only a subset of the TFL shells in a standard are implemented as templates in the TFL Workbench. In fact – at the time of writing – only table templates containing descriptive summary statistics (limited to those that can be obtained with PROC MEANS, plus counts and percentages for categorical data) are implemented. Figures and listings templates will be added in the future, but currently there are no plans to add templates for more complex inferential statistics. In addition, of the TFL Workbench templates that can be used in a study, not all can be used without modification.

Metrics generated using the TFL Workbench metadata only cover a sub-set of all shells created for a study, and TFL Workbench metrics only provide the ratio between all outputs created by the TFL Workbench and the subset of tables using unmodified templates. But as the TFL Workbench is the platform for standard SAS® programs, we can measure how many changes to our standard configurations are made in each study, which directly translate to changes in standard programs.

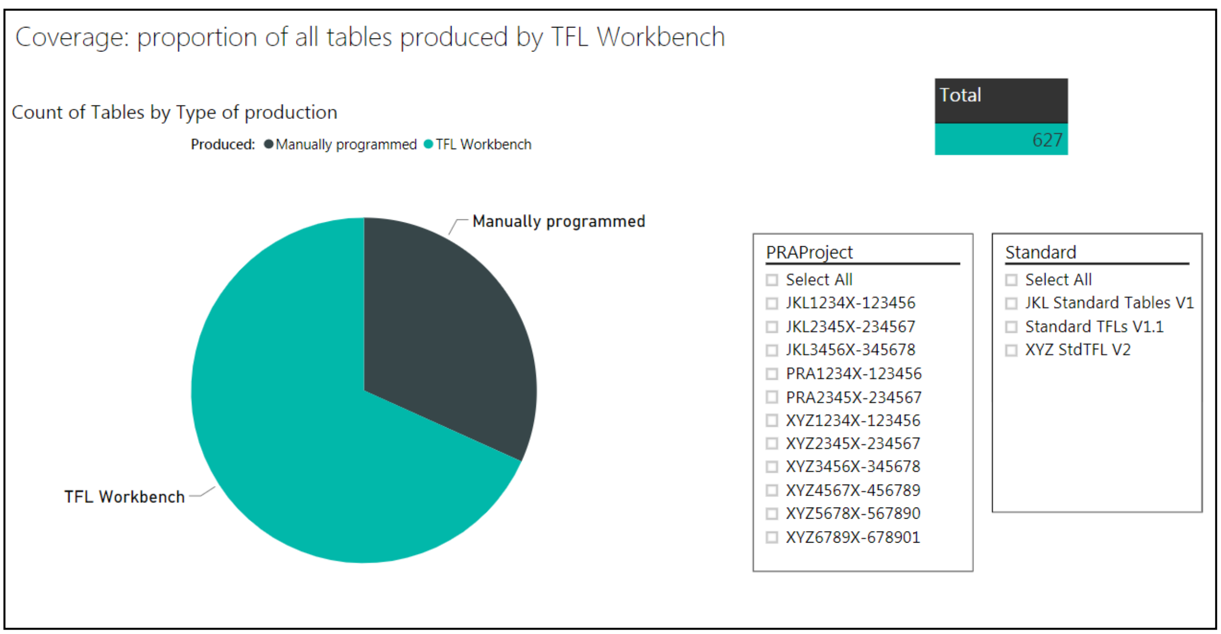


INDIVIDUAL METRICS

HOW OFTEN IS A STANDARD USED?

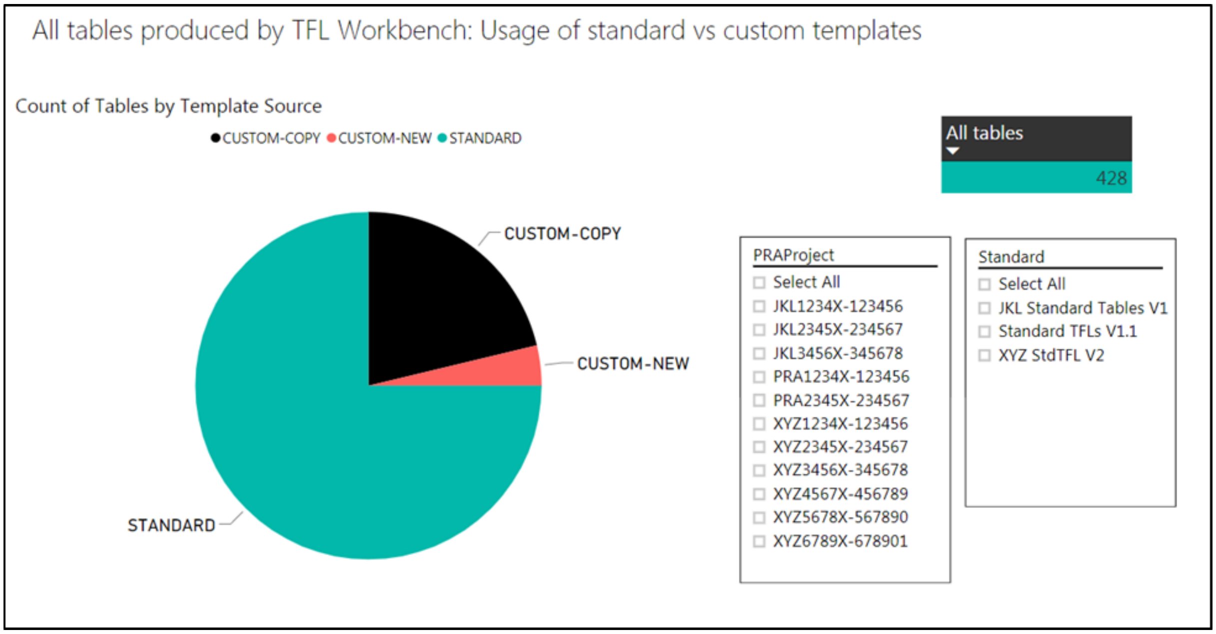
This metric is currently gathered manually by comparing different trackers. Although we know which projects use the TFL Workbench, and which TFL Standard they use, because we automatically track the usage of the tool, information about projects which do not use the TFL Workbench has to be obtained from separate systems. We currently look at studies that commence programming within a certain year and compare it with the number of studies for which a TFL Workbench database is downloaded during this year.

WHAT IS THE PROPORTION OF OUTPUTS THAT ARE BASED ON STANDARD SHELLS?



This metric provides insight in how good a standard is in supporting the goal of reporting clinical study results. As we draw our metrics from TFL Workbench usage, as an approximation for this metric we calculate the number of tables created by the TFL Workbench and based on standard programs, compared to all tables created for a study. These can include tables based on custom configurations, and not on any standard template at all but are created because the programming team decided it is more efficient to use the TFL Workbench. This is compared to all tables, either created via the TFL Workbench or created by bespoke programs. This metric also tells us how much of the study output we can cover using the TFL Workbench, i.e. how useful the tool is. Results can be filtered down by standard, as well as by study (or group of studies).

WHAT IS THE PROPORTION OF TFLS CREATED BY THE TFL WORKBENCH BASED ON STANDARD TEMPLATES VS CUSTOM TEMPLATES?

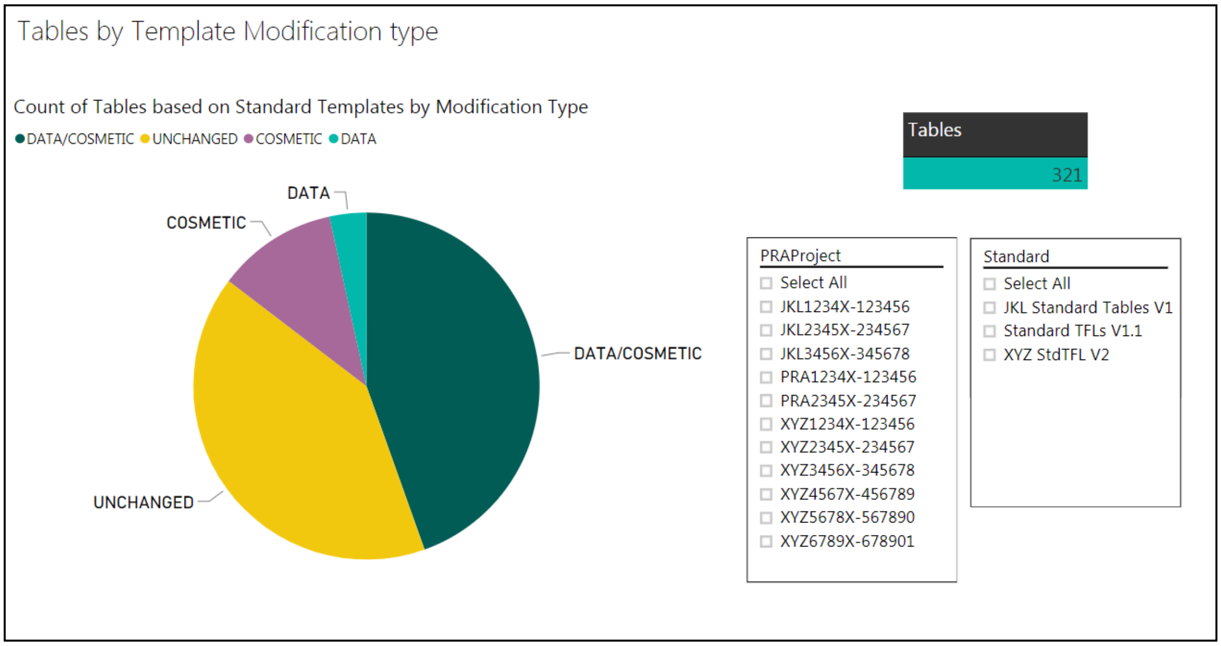


Again, the metric is an approximation for the use of standard shells, as it shows the proportion of tables that were created based on standard templates, and for how many tables custom templates needed to be created. STANDARD refers to a standard template being used, altered or unaltered. CUSTOM-COPY stands for templates that originated as a copy of a standard template. Copies of a standard template are often done because there is a set of tables that deviates slightly from tables created using the standard template. By making a copy of a standard template and modifying it, these tables can be created relatively easy. CUSTOM-NEW are templates that are added without explicitly copying a standard template. Theoretically, CUSTOM-NEW could represent tables that have no shell in the standard, but for the time being it is mostly chosen to create tables similar to standard tables.

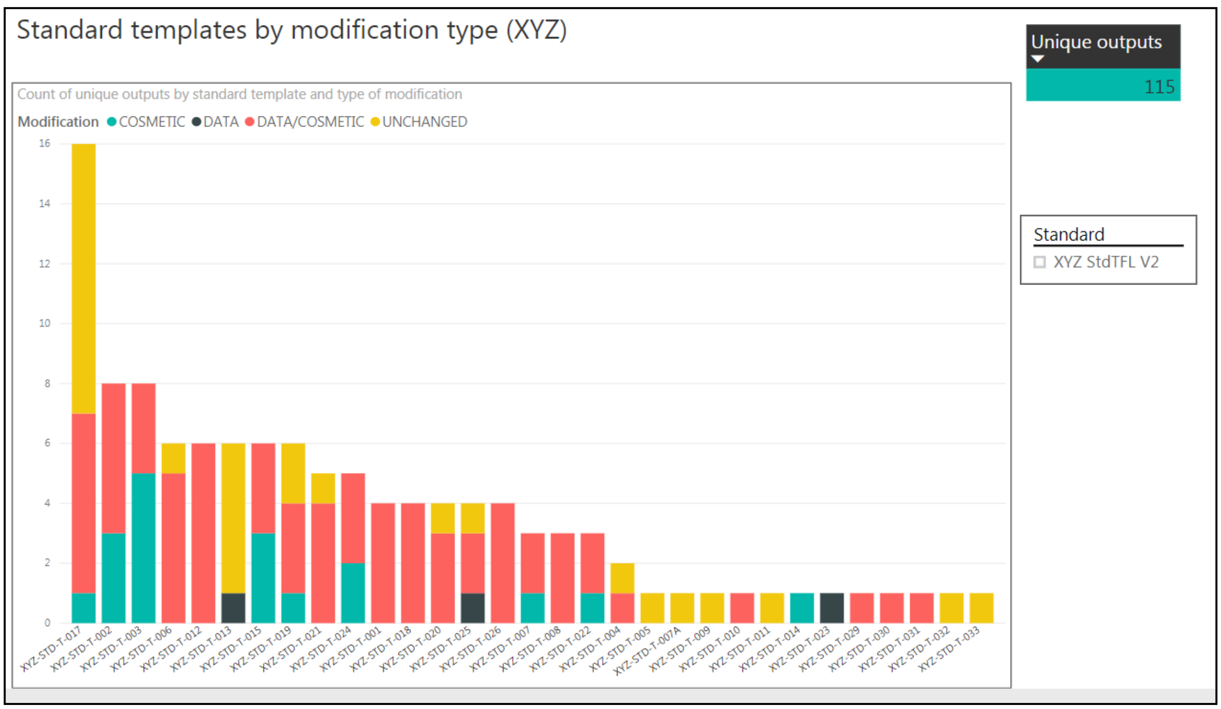
HOW MANY TFLS REQUIRED MODIFICATION OF STANDARD SHELLS?

Standard templates are usually used to create standard tables, thus changes applied to them represent (necessary) deviations from the standard. Templates are the integral set of configurations that represents the standard, as they represent major structural elements of a data display. Although we recognize that some changes might be necessary to adapt a standard to match specific study designs and specific questions, we obviously want the standard to change as little as possible. Especially, we want to avoid idiosyncratic changes, thus a high compliance with available templates represents a low volume of such changes.

The next picture shows the changes made to standard templates. "Cosmetic changes" relate mostly to label changes, which are of a lesser importance, while "DATA" changes represent modifications to analysis variables, class variables, where clauses or statistics, and which are more critical to get right.



WHICH ARE THE STANDARD SHELLS USED THE MOST, AND HOW OFTEN ARE THEY CHANGED?



Presenting the data from the metric above by template provides us information about which templates are used most, and which are the most frequently modified.

As in the slide before, UNCHANGED represents templates that were not modified for use in a study – they were fit for purpose. Other templates might change often, either implementing COSMETIC or DATA changes, or both (DATA/COSMETIC), but this is expected. An example of this would be a demography and baseline characteristic table (baseline characteristics are often related to the product under investigation, or to the study design), or populations (often driven by the investigation), or compliance and exposure tables. Other tables, such as those presenting occurrence data (adverse events, medical history, concomitant medications), or standard laboratory summaries, rarely require a change to the template.

WHAT DID CHANGE IN INDIVIDUAL SHELLS?

	Added	Deleted	Modified
>20%			
>40%			
>60%			
>80%			

Comparing the instances of a template across the multiple studies that used it gives us an insight into which parts of a template were changed most often. Although this is not a metric in a strict sense, its allows for easy review of frequent modifications of a standard table shell. Again, we use data from the configuration workbook as a basis for this report. Deletions and modifications can easily be highlighted, as they affect existing elements of a standard templates. Additions are a little bit trickier as new element with the same meaning can be added to the template in multiple ways. We therefore decided not to show the individual additions, but rather add a note (a blue line) at the top of the template (if new sections were added) or below each section (if new lines were added), the coloring of the line indicating the frequency of additions.

METADATA	SECTION	ROW	SUBROW	PARENT	DATA TYPE	SUPPRESS IF NONE	ROWLABEL1	INDENT1	ROWLABEL	INDENT2	TREATMENT	STATISTIC	PRECISION	CLASS VARS	ANALYSIS VA	WHERE	DENOM
							Additional sections were inserted in this template										
blank						N											
row label hdr						N	Characteristic	0									
label	1	1			CAT	N	Sex [n (%)]	0									
statistic	1	2			CAT	N	Female	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	SEXN eq 1	_BigN_
statistic	1	3			CAT	N	Male	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	SEXN eq 2	_BigN_
statistic	1	4			CAT	Y	Missing	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	SEXN eq .	_BigN_
blank						N											
label	2	1			CAT	N	Race [n (%)]	0									
statistic	2	2			CAT	N	White	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ARACEN eq 5	_BigN_
statistic	2	3			CAT	N	Black or African American	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ARACEN eq 3	_BigN_
statistic	2	4			CAT	N	Asian	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ARACEN eq 2	_BigN_
statistic	2	5			CAT	N	American Indian or Alaska Native	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ARACEN eq 1	_BigN_
statistic	2	6			CAT	N	Native Hawaiian or other Pacific Islander	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ARACEN eq 4	_BigN_
statistic	2	8			CAT	N	Other	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ARACEN in (6 7)	_BigN_
statistic	2	9			CAT	Y	Missing	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ARACEN eq .	_BigN_
							Additional elements were inserted in this section										
blank						N											
label	3	1			CAT	N	Ethnicity [n (%)]	0									
statistic	3	2			CAT	N	Hispanic or Latino	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ETHNICN eq 1	_BigN_
statistic	3	3			CAT	N	Not Hispanic or Latino	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ETHNICN eq 2	_BigN_
statistic	3	4			CAT	Y	Missing	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	ETHNICN eq .	_BigN_
blank						N											
label	4	1			CON	N	Age (years)	0									
statistic	4	2			CON	N	n	2			xx	n	0		AGE		
statistic	4	3			CON	N	Mean (std)	2			xx.x (xx.xx)	mean#std	0		AGE		
statistic	4	4			CON	N	Median	2			xx.x	median	0		AGE		
statistic	4	5			CON	N	Q1, Q3	2			xx.x, xx.x	q1#q3	0		AGE		
statistic	4	6			CON	N	Min, Max	2			xx, xx	min#max	0		AGE		
blank						N											
label	5	1			CAT	N	Age Group [n (%)]	0									
statistic	5	2			CAT	N	<65 years	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	AGEGR1N eq 1	_BigN_
statistic	5	3			CAT	N	&SYM_GE. 65 years - <75 years	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	AGEGR1N eq 2	_BigN_
statistic	5	4			CAT	N	&SYM_GE. 75 years	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	AGEGR1N eq 3	_BigN_
statistic	5	5			CAT	Y	Missing	2			xx (xx.x)	n#pct	1		COUNT(_ALL_)	AGEGR1N eq .	_BigN_
blank						N											
label	6	1			CON	N	Weight (kg)	0					1		BASEWT		
statistic	6	2			CON	N	n	2			xx	n	1		BASEWT		
statistic	6	3			CON	N	Mean (std)	2			xx.x (xx.xx)	mean#std	1		BASEWT		
statistic	6	4			CON	N	Median	2			xx.x	median	1		BASEWT		
statistic	6	5			CON	N	Q1, Q3	2			xx.x, xx.x	q1#q3	1		BASEWT		
statistic	6	6			CON	N	Min, Max	2			xx, xx	min#max	1		BASEWT		

The picture below shows a snippet from a demography table template. The row label header (The text that is presented as a header of the label column) was modified in > 40% - 60% of the study templates, the age groupings and corresponding where clauses were changed in >80% of the study templates. Also, in > 40% - 60% of study templates, new sections were added and at least one new category added to the RACE section. Finally, in > 20% - 40% of studies the quartiles were deleted from the statistics displayed.

CONCLUSION

TFL Standards support quality and consistency of data displays created to report the outcome of clinical studies. Metrics are an essential tool to govern the use of standards. Understanding the proliferation of a standard, and the changes necessary to fulfil the study design, is valuable information to further develop TFL standards. If it is possible to pass the improvements on to standard programs, quality and consistency of data displays can be improved as well. Rather than collecting metrics data manually, it is beneficial to use information that is created as a by-product of the TFL production process. Microsoft Word® does not provide any handle to extract such data. The TFL Workbench was developed to allow the creation of TFL programs based on machine-readable metadata. This feature comes in handy, as it provides us with data that serve as an approximation to the requested metrics. Although often an approximation, the TFL Workbench allows us valuable insight into the use of standard TFL shells. Does this mean we always have to rely on surrogate data to measure the use of standards? Only if we insist in using tools like Microsoft Word® to create TFL shells. The TFL Workbench might in future allow us to streamline the process of writing shells and producing data displays by giving Biostatisticians the opportunity to use the libraries of standard configurations to also create the shells. This would reduce significantly the effort downstream of adapting configurations to meet the TFL shells, as well as copying titles and footnotes. If we can do this, we can collect metrics about the use of TFL standards, which are created as a by-product of the process of writing shells, rather than collecting them manually.

RECOMMENDED READING

Iain Humphreys, Hansjörg Frenzel: “The TFL Workbench: Tools to Standardize and Accelerate TFL Programming”, AD01, PhUSE 2016

Iain Humphreys, Hansjörg Frenzel: “The TFL Workbench: From Pilot to Production”, AD05, PhUSE 2017

Hansjörg Frenzel, James Witt: “Save Time: Program your Output Specs”, PD01, PhUSE 2006

CONTACT INFORMATION

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

Hansjörg Frenzel
PRA Health Sciences (Germany)
Gottlieb Daimler Str. 10
68165 Mannheim
Email: FrenzelHansjoerg@prahs.com
Web: www.prahs.com

Iain Humphreys
PRA Health Sciences (UK)
500 South Oak Way
Reading, RG2 6AD
Email: HumphreysIain@prahs.com