PharmaSUG China 2025 - Paper AA-180

Outline of Shell and TFL Automation Tools

Huafeng Gao, Zhiyuan Guo, Xianghua Tang, Pei Liu and Kaifan Li, Keymed Biosciences Inc.

ABSTRACT

In clinical trials, statistical analysis Tables, Figures and Listings (TFLs) are the primary deliverables. Usually, statisticians manually create Shell for TFLs, and statistical programmers generate TFLs according to the Shell.

This paper shares a set of tools which implemented in Keymed Biosciences Inc. (Keymed), including Shell generation tools, TFLs automation, and quality control within the Analysis Task Tracker system (ATTS).

We compiled general conventions and Shell Library. And accordingly, we developed this set of tools. In Within ATTS, statisticians can select shells from the Shell Library and effortlessly create new ones. Programmers, meanwhile, do not need to write code for TFLs that are already in the library.

INTRODUCTION

This set of tools are built around the Shell Library and integrating SDTM and ADaM automation (See PharmaSUG China 2025 - Paper AA - 179). Through the development of flexible SAS macros tailored for diverse scenarios, it establishes a comprehensive and efficient automated framework for statistical analysis.

The following outlines the overall workflow, which structured into four sections:

- the implementation of the Tables and Figures Shell tool (Section 2).
- the auto-generation of SAS programs to produce Tables and Figures Output (Section 3).
- the implementation of the Listing Shell and Output tools (Section 4).
- quality control for TFLs automation with the Analysis Task Tracker (Section 5).

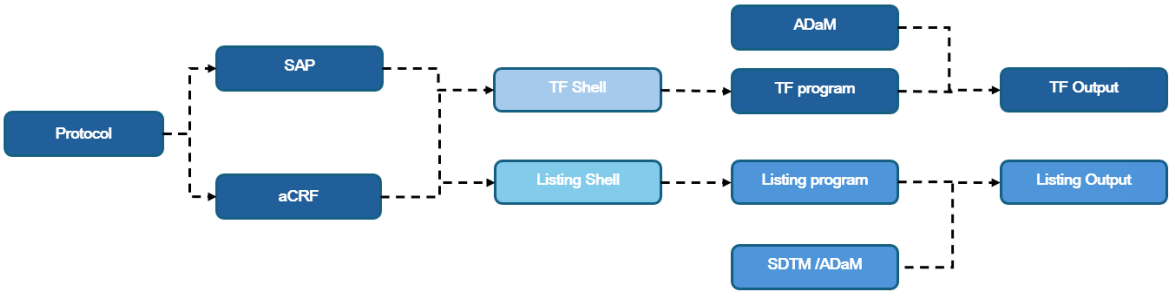


Figure 1. The Overall Workflow

1. OVERVIEW

1.1 INTRODUCTION TO SHELL TOOLS AND TFLS AUTOMATION

CRF with an annotated version will display the information about the data, including variable name, label, CT, etc.. SAP illustrate the statistical considerations, including definition and derivation rule of estimands, statistical methods, handling of missing data, the analysis specification, etc.. Shell is an important bridge for communication between statisticians and statistical programmers. With the guidance of CRF and SAP, it will be designed to display the TFLs' profile, including the content, format and layout. For the table in the Shell, generally the title, footnote, header, content format and program description should be drawn. For

the figure in the Shell, generally the title, footnote, graphic example and elements of figure should be described in details. For the listing in the Shell, generally the title is enough because the requirement of ICH E3 is clear.

Based on the above information, we mainly adopt a configuration file-driven mode to integrate relevant metadata information. For the listing, as it mainly focuses on presenting raw data, it has been implemented with synchronization processing with the SDTM (Study Data Tabulation Model) generation process, combining the mapping file during SDTM generation with ICH requirements to output the Shell; for table and figure, output is achieved by using a unified Shell Library and CT values as well as the titles, footnotes, and program descriptions provided by statisticians.

With the Shell, SDTM and ADaM already in place, we have achieved the automated output of TFLs by using the Shell process file configuration and matching with macros.

2. TABLES AND FIGURES SHELL

2.1 METHODS OF SHELL GENERATION

The methods of shell generation are mainly divided into the following three categories:

1. Fixed format: Generate tables using preset standard Shell Library.
2. CT: Align output with both raw and analysis CT, for example, summaries of sex and race in demographic characteristics tables.
3. Graphical Data Simulation: Generate outputs using simulated data that conforms to specific trends, such as PK concentration.

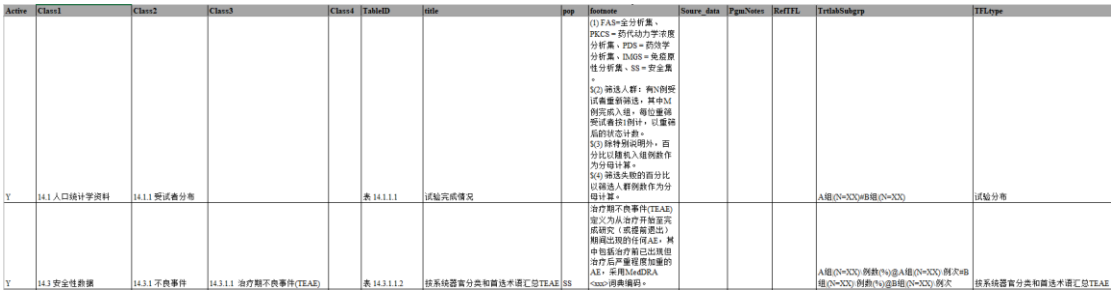
	A组 (N=XX)	B组 (N=XX)	C组 (N=XX)	合计 (N=XX)
性别				
女性	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
男性	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
Total(Missing)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
年龄 (岁)				
n(Missing)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
Median	xx.x	xx.x	xx.x	xx.x
Q1, Q3	xx, xx	xx, xx	xx, xx	xx, xx
Min, Max	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x
民族				
其他	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
汉族	xx(xx.x)	xx(xx.x)	xx(xx.x)	xx(xx.x)
Total(Missing)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
身高 (cm)				
n(Missing)	xx(xx)	xx(xx)	xx(xx)	xx(xx)
Mean(SD)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)	xx.x(xx.x)
Median	xx.x	xx.x	xx.x	xx.x
Q1, Q3	xx, xx	xx, xx	xx, xx	xx, xx
Min, Max	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x	xx.x, xx.x

Figure 2 TF Shell Example

2.2 CONFIGURATION OVERVIEW

During the Shell generation, based on the above three methods, we have designed differentiated key component modules to achieve a high-efficiency, standardized, and reusable output process. These core components mainly include:

1. Tracker: Records the metadata provided by the statistician, including Outline, Title, Footnote, Programming Note and Header information, etc., to ensure the uniformity and traceability of the output document format. It also records key parameters essential for downstream TFL automation, such as data sources and filter criteria.



Active	Class1	Class2	Class3	Class4	TableID	title	pop	footnote	Soure_data	PgmNotes	RefTFL	TrtlabSubgrp	TFLtype
Y	14.1 人口统计学资料	14.1.1 受试者分布			表 14.1.1.1	试验完成情况		(1) FAS=全分析集，PKCS = 药代动力学浓度分析集，PDS = 药效学分析集，IMGS = 免疫原性分析集，SS = 安全集。 5(2) 筛选入群：例N例受试者重新筛选，其中M例完成入组，每位重筛受试者按1例计，以重筛后的状态计数。 5(3) 除特别说明外，百分比以随机入组例数作为分母计算。 5(4) 筛选失败的百分比以筛选入群例数作为分母计算。				A组(N=XX)vsB组(N=XX)	试验分布
Y	14.3 安全性数据	14.3.1 不良事件	14.3.1.1 治疗期不良事件(TEAE)		表 14.3.1.1.2	按系统器官分类和首选术语汇总TEAE	SS	治疗期不良事件(TEAE)定义为从治疗开始至完成研究（或提前退出）期间出现的任何AE，其中包括治疗前已出现但治疗后严重程度加重的AE。采用MedDRA <xxx>词典编码。				A组(N=XX) 例数(%)@A组(N=XX) 例次vsB组(N=XX) 例数(%)@B组(N=XX) 例次	按系统器官分类和首选术语汇总TEAE

Figure 3 Tracker Example

2. Shell Library: As a standardized template library, it stores a unified Shell structure that has been discussed and confirmed, ensure the consistency of the shell structure between different projects.

fmtname	Bold	Text	Avalc
交叉表	1	<检查指标x (单位)>	
交叉表	1	基线后最差结果	
交叉表	1	xxxx (N=xx)	
交叉表		Normal	xx (xx.x)
交叉表		NCS	xx (xx.x)
交叉表		CS	xx (xx.x)
交叉表		Total	xx (xx.x)
交叉表		Missing	xx

Figure 4 Shell Library Example

3. Codelist: Based on the Controlled Terminology (CT) of SDTM and ADaM, thereby precluding data discrepancies caused by terminology entry errors at the source.

Dataset	ID	SUPPID	CodeListName	CodeList	Order	Decode_Value
	ENRTPT	TONY		曾经吸烟	1	
	ENRTPT	ALNY		曾经饮酒	2	
	ENRTPT	TONY		目前吸烟	3	
	ENRTPT	ALNY		目前饮酒	4	
	ETHNIC			其他	2	
	ETHNIC			汉族	1	
	ETHNICC			其他	2	
	ETHNICC			汉族	1	

Figure 5 Codelist Example

4. TFLShell Simds: Record the parameterized configuration generated according to SAP analysis requirements, and add codelist mapping to ensure that the results accurately correspond to statistical requirements.

catgory	Encatgory	vname	Conditions	line_break	vtype	vlabel	codelistid
试验分布	ADSL		subjid ne ''		Char	筛选人群	
试验分布	ADSL		SCRFFL='是'		Char	筛选失败	
试验分布	ADSL		RANDFL='是'		Char	随机入组	
试验分布	ADSL		SUBPAHNY='是' or index(EOTSTT,'完成') or SECDONY='是' or "<strip(SVSTDTC)<strip(EOTDTC)	1	Char	主研究完成第一轮治疗	
试验分布	ADSL		index(EOTSTT,'完成') or SUBPAHNY='是'		Char	主研究完成治疗	
试验分布	ADSL	DCTREAS	index(EOTSTT,'提前') and SUBPAHNY='否'		Char	主研究提前终止治疗	NCOMPLT/PROTMLST
试验分布	ADSL		index(EOSSTT,'完成') or SUBPAHNY='是'	1	Char	主研究完成研究	
试验分布	ADSL		EOSSTT='' and SUBPAHNY='否'		Char	主研究进行中	
试验分布	ADSL	DCSREAS	index(EOSSTT,'提前') and SUBPAHNY='否'		Char	主研究提前退出研究	NCOMPLT/PROTMLST
试验分布	ADSL		SUBPAHNY='是'	1	Char	进入子研究	
试验分布	ADSL		SUBPAHNY='是' and TR02SDT=.	1	Char	子研究未接受试验用药品~{super #}	
试验分布	ADSL		index(EOTSTT,'完成') and SUBPAHNY='是'		Char	子研究完成治疗~{super #}	
试验分布	ADSL	DCTREAS	index(EOTSTT,'提前') and SUBPAHNY='是'		Char	子研究提前终止治疗~{super #}	NCOMPLT/PROTMLST
试验分布	ADSL		index(EOSSTT,'完成') and SUBPAHNY='是'	1	Char	子研究完成研究~{super #}	
试验分布	ADSL	DCSREAS	index(EOSSTT,'提前') and SUBPAHNY='是'		Char	子研究提前退出研究~{super #}	NCOMPLT/PROTMLST

Figure 6 TFLShell Simds Example

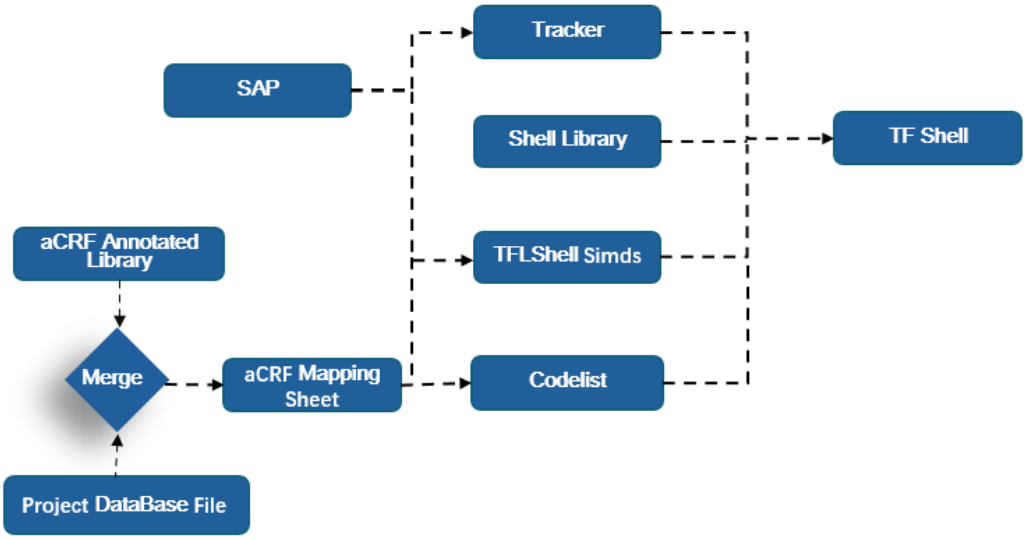


Figure 7 TF Shell Workflow

2.3 SHELL VISUALIZATION (COMPARISON BEFORE AND AFTER OPTIMIZATION)

To optimize the Shell workflow and enhance the user experience, a web-based Shell visualization system was designed and developed. Transfer the Shell configuration document filling to the Kangshu platform and interact with the background program through the web page to achieve Shell output. The process is as follows:

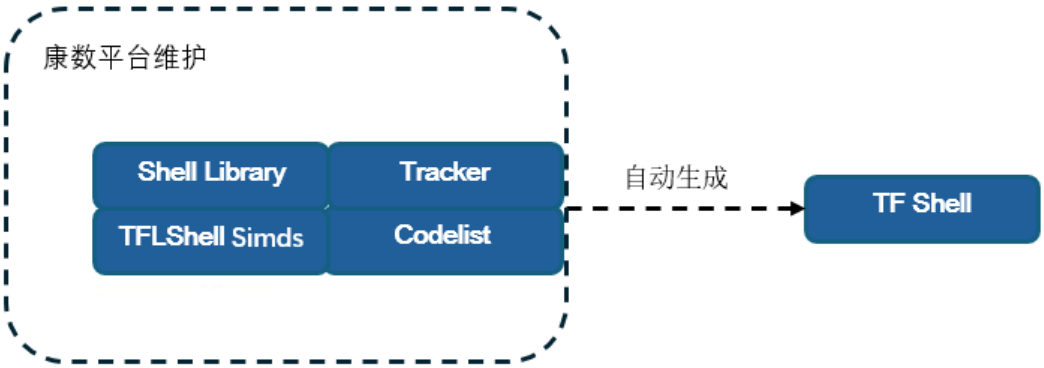


Figure 8 WebWorkflow

Compared with the Excel version, this innovative solution has achieved significant improvements in functionality and usability. Its core advantages are reflected in the following aspects:

Comparison	Shell First Edition (Excel)	Shell Visualization (Webpage)
Visual Interaction	The Header style cannot be visually demonstrated.	Real-time rendering and display of Header styles can be achieved
Automated Processing	1. Manually enter the TableID 2. When inserting or deleting tables, the associated TableID needs to be manually modified	1. Realize the intelligent generation and automatic maintenance of TableID 2. Support dynamic management of table association relationships
Operational Efficiency	1. The header construction allows you to concatenate content by yourself 2. Newline characters should be used to handle the format of footnotes, program descriptions and other contents	1. Introduce a structured input mode to simplify the table header construction process 2. Adopt a itemized management mechanism to optimize metadata maintenance
Maintenance Mode	Manual maintenance	System maintenance

Table 1. Comparison between Excel and the web version

Figure 9 Web Shell Example

3. TABLES AND FIGURES OUTPUT

3.1 HOW TO INTEGRATE INFORMATION FOR OPTIMAL PROGRAMMING EFFICIENCY

1. Macro Design: Generate SAS macro templates based on different shell styles. Leverage configuration files to pass macro parameters.
2. Automatic parameter matching: Automatically associate corresponding macro name, analysis set variables, and analysis dataset, etc., based on title information.

3.2 SELECT MACROS BASED ON INTEGRATED INFORMATION AND GENERATE SAS PROGRAM

1. Structure the integrated configuration information and output it as SAS dataset.

2. Based on the configuration dataset, match the corresponding SAS macro program and call it.
3. Integrate program headers, configure datasets, macro calls, and other core components to generate SAS programs that can run independently.

```
data dummy;
    length text cond $200;
    infile datalines dlm='$' missover;
    input text cond $;
    ord=_n_;
    datalines;
TEAE $ TRTEMFL='是'
1级 $ TRTEMFL='是' and AESEV="1级"
2级 $ TRTEMFL='是' and AESEV="2级"
3级 $ TRTEMFL='是' and AESEV="3级"
4级 $ TRTEMFL='是' and AESEV="4级"
5级 $ TRTEMFL='是' and AESEV="5级"
与试验用药品相关 $ RELGR1N=1
1级 $ RELGR1N=1 and AESEV="1级"
2级 $ RELGR1N=1 and AESEV="2级"
3级 $ RELGR1N=1 and AESEV="3级"
4级 $ RELGR1N=1 and AESEV="4级"
5级 $ RELGR1N=1 and AESEV="5级"
≥3级 $ SEVGR1N=1
与试验用药品相关 $ RELGR1N=1 and SEVGR1N=1
SAE $ AESER='是'
与试验用药品相关 $ RELGR1N=1 and AESER='是'
导致提前退出研究 $ AEDIS='是'
与试验用药品相关 $ RELGR1N=1 and AEDIS='是'
导致死亡 $ AESDTH='是'
与试验用药品相关 $ RELGR1N=1 and AESDTH='是'
;
run;

data dummy;
    set dummy;
    if (index(cond,'RELGR1N') and index(cond,'and')) or index(cond,'AESEV') then blank=1;
run;

%t_aesum2table(order=1 ,popcond=SAFFL='是' ,cond=and TRTEMFL='是' ,ManualWidthx=text1*21);
```

Figure 10 Automated Program Display

3.3 PROGRAM TEMPLATE OPTIMIZATION AND ITERATION

At present, the TFLs automation has completed the basic development based on the existing therapeutic areas (TA) of our company. However, when dealing with the analysis requirements of new therapeutic areas, in-depth iteration and functional optimization are still needed. New tasks typically involve various statistical analysis methods, including multiple comparison correction, dynamic adjustment of covariates, refined subgroup analysis, and modeling of complex time-event data, all of which pose new challenges to the intelligence level and computational accuracy of automation. The future automation upgrade will focus on the following key directions:

1. Enhance the dynamic parameter configuration capability, identify the parameter configurations with the characteristics of different therapeutic domains, to support the calculation of efficacy endpoints for different research designs;
2. Optimize the statistical model adaptation mechanism to ensure the precise invocation of complex methods such as survival analysis and mixed-effects models;
3. Establish a standardized verification process from data input to result output, and ensure the repeatability and robustness of the analysis results through cross-validation and other means.

Through continuous iteration, TFLs automation will gradually cover a wider range of analytical scenarios, thereby enhancing the efficiency and reliability of clinical research data processing. Since our team mainly uses the SAS language as the primary language, with other languages (such as Python, R, etc.) as supplementary, we adopt a multi-language collaboration model for development. We modularized the functions, successfully implemented them on SAS, and then combined them with other languages to achieve visualization.

4. LISTING

4.1 CONCEPT OF LISTING AUTOMATION IMPLEMENTATION

By matching the aCRF Annotated Library with the database file, the aCRF Mapping Sheet is obtained and then exported as the Listing Mapping Sheet(including key information such as the display content of each list, the source data set, the variable mapping relationship, and the filtering conditions). Based on the predefined configuration template, choosing different configurations can directly Output the Listing Shell and output.

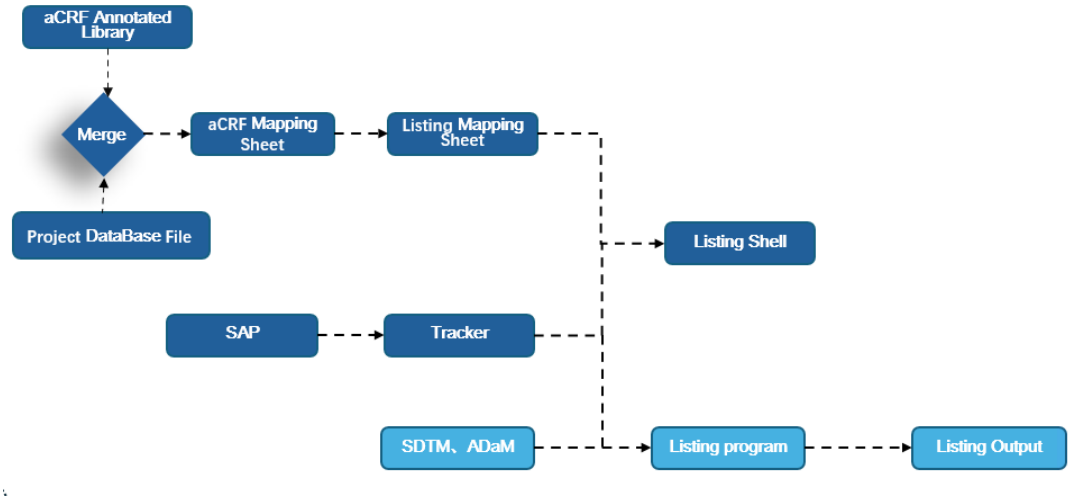


Figure 11 ListingWorkflow

4.2 HOW TO USE THE ACRF MAPPING SHEET TO CREATE THE LISTING MAPPING SHEET

During the process of constructing the aCRF Mapping Sheet, The metadata such as Variable Name and Variable Label in the original Data set were systematically matched with the SDTM (Study Data Tabulation Model) standard. However, when generating the Listing of clinical research reports, some data elements need to be specifically integrated and restructured according to the actual analysis requirements, and these treatments often differ from the standard requirements of SDTM. To this end, the Listing Mapping Sheet, an intermediate mapping file, was created, aiming to build a bridge from standardized SDTM to customized analysis output, ensuring compliance with regulatory requirements as well as meeting the data presentation needs in specific research scenarios. Through clear mapping rules, the traceability association among source data, SDTM standards and the final list output has been achieved.

Description	orde	Dataset	ListingHeader	Source
筛选失败受试者	1	adam.adsl	受试者|筛选号	SUBJID
筛选失败受试者	1	adam.adsl	签署知情|同意版本号	ICPVER
筛选失败受试者	1	adam.adsl	签署知情|同意日期	RFICDTC
筛选失败受试者	1	adam.adsl	筛选失败的原因	catx(": ",SCRFRES,SCRFRESP)
筛选失败受试者	1	adam.adsl	是否重新|进行筛选	RESCRSNY
筛选失败受试者	1	adam.adsl	若重筛，首次筛选号|及筛选失败原因	catx(", ",PRESUBID,FSCRFRES)
受试者入组情况	2	adam.adsl	组别	TRT01P
受试者入组情况	2	adam.adsl	受试者|筛选号	SUBJID
受试者入组情况	2	adam.adsl	若重筛，首次筛选号|及筛选失败原因	catx(", ",PRESUBID,FSCRFRES)
受试者入组情况	2	adam.adsl	签署知情|同意日期	RFICDTC
受试者入组情况	2	adam.adsl	签署知情|同意版本号	ICPVER

Figure 12 Listing Mapping Sheet Example

列表 16.2.1.1 筛选失败的受试者

研究阶段	受试者编号	知情同意		筛选失败原因	再筛选	
		签署日期	版本号		是否	首次筛选号

Figure 13 Listing Shell Example

4.3 CREATE A LISITNG MACRO TO CONVERT THE LISTING MAPPING SHEET INTO THE LISTING SHELL AND OUTPUT

Based on the configuration information of the Listing Mapping Sheet, operations such as data merging, transposing and format conversion are executed programmatically, and the output opening is set. According to the actual research requirements, flexibly choose to Output the Listing Shell structure template that conforms to the CDISC standard or directly generate the final standardized analysis output based on the existing SDTM dataset.

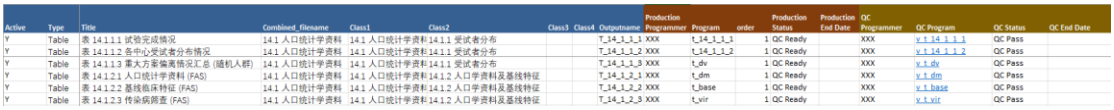
```
*列表 16.2.1.1 筛选失败的受试者*;

%Listing(order=1
        ,inds=adam.adsl
        ,AddPopVar=Y
        ,cond=%str( )
        ,var=
        ,varlabel=
        ,layout=l
        ,FreezeWidth=
        ,ManualWidth=
        ,DisplayOrder=1
        ,Compute=%str( )
        );

*列表 16.2.1.2 受试者入组及完成情况(入组人群)*;

%Listing(order=2
        ,inds=adam.adsl
        ,AddPopVar=Y
        ,cond=%str( )
        ,var=
        ,varlabel=
        ,layout=l
        ,FreezeWidth=
        ,ManualWidth=
        ,DisplayOrder=1
        ,Compute=%str( )
        );
```

Figure 14 Listing program Example

5. QUALITY CONTROL OF TFL AUTOMATION IN THE ANALYSIS TASK TRACKER

5.1 QUALITY CONTROL IN NON-AUTOMATED SCENARIOS

Quality control in the traditional workflow mainly relies on manual labor. Although the working mode of tracking and maintenance through Excel is flexible, there are also problems and challenges in the implementation process. Due to the working characteristics of Excel, the preparation status, update time, comparison status, etc. of the main and QC programs all need to be entered by oneself, which is prone to untimely update and affect the control of progress.



Active	Type	Title	Combined_filename	Class1	Class2	Class3	Class4	Outputname	Production Programmer	Program	order	Production Status	Production End Date	QC Programmer	QC Program	QC Status	QC End Date
Y	Table	表 14.1.1.1 试验完成情况	14.1 人口统计学资料	14.1 人口统计学资料	14.1.1 受试者分布			T_14_1_1_1	XXX	t_14_1_1_1	1	QC Ready		XXX	v t 14 1 1 1	QC Pass	
Y	Table	表 14.1.1.2 各中心受试者分布情况	14.1 人口统计学资料	14.1 人口统计学资料	14.1.1 受试者分布			T_14_1_1_2	XXX	t_14_1_1_2	1	QC Ready		XXX	v t 14 1 1 2	QC Pass	
Y	Table	表 14.1.1.3 重大方案偏离情况汇总 (随机人群)	14.1 人口统计学资料	14.1 人口统计学资料	14.1.1 受试者分布			T_14_1_1_3	XXX	t_dv	1	QC Ready		XXX	v t dv	QC Pass	
Y	Table	表 14.1.2.1 人口统计学资料 (FAS)	14.1 人口统计学资料	14.1 人口统计学资料	14.1.2 人口学资料及基线特征			T_14_1_2_1	XXX	t_dm	1	QC Ready		XXX	v t dm	QC Pass	
Y	Table	表 14.1.2.2 基线临床特征 (FAS)	14.1 人口统计学资料	14.1 人口统计学资料	14.1.2 人口学资料及基线特征			T_14_1_2_2	XXX	t_base	1	QC Ready		XXX	v t base	QC Pass	
Y	Table	表 14.1.2.3 传染病筛查 (FAS)	14.1 人口统计学资料	14.1 人口统计学资料	14.1.2 人口学资料及基线特征			T_14_1_2_3	XXX	t_vir	1	QC Ready		XXX	v t vir	QC Pass	

Figure 15 Traditional Tracker Example

5.2 AUTOMATION AND ANALYSIS TASK TRACKER QUALITY CONTROL

Automated data management based on the Web platform migrates various types of data previously managed in Excel file form to a centralized Web management platform, achieving standardized storage and unified control of data. The system can automatically distribute the processed data to various downstream automated tools. For instance, migrate entries such as TFL that were previously managed in Excel to the Web platform for management; On this basis, Shell automation and Tracker quality tracking can be integrated.

The Tracker quality tracking system automatically monitors the source code files of the main and QC programs and their derivative outputs, tracks comparison results in real time, and synchronously outputs the project progress and project urgency in combination with the milestone plan, achieving an automated closed-loop management of quality control. 6. Practice and Application.

6. PRACTICE AND APPLICATIONQ

6.1 COMPARISON BETWEEN THE AUTOMATED TOOLS AND THE TRADITIONAL WORKFLOW

Comparison dimension	Traditional working mode	Automated tool mode
Work efficiency	Highly dependent on manual operation, it takes a long time and is easily affected by human factors.	Reduce repetitive work and be suitable for batch data processing.
Quality control	Rely on manual review	Built-in real-time quality control, combined with regular manual spot checks, forms an efficient quality control closed loop.
Process standardization	It depends on personal experience and varies greatly	Standardized process output
Traceability	Relying on manual records, traceability is difficult	Complete log records, with full traceability throughout the process
Repeatability	The same analysis for different projects needs to be redone	Controllable code reuse across projects

Table 2. Comparison between the past and current working modes

7. EXPLORATION OF SAP TOOL

Through practical exploration and experience summary in multiple clinical research projects, we have found that the compilation process of SAP documents has significant potential for standardization and can build a unified template system. Moreover, by establishing a specific rule engine, the intelligent capture of Shell configuration information can be achieved. This breakthrough will significantly enhance the work efficiency of statisticians.

ACKNOWLEDGMENTS

We sincerely thank all the team members who participated in, supported and tested during the tool development process. Special thanks go to Dr.Guoqing Zhao and Rongqin Xie for their guidance and

feedback, as well as to statistician Li Zhao for providing the initial shell toolization ideas, and Xiaoyu Ding for his technical support in clinical data visualization platform.

CONTACT INFORMATION

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

Author Name: Huafeng Gao

Company: Keymed Biosciences Inc.

E-mail: huafenggao@keymedbio.com