cdisc

2022 CHINA INTERCHANGE

29-30 JULY | VIRTUAL EVENT

An Efficient Approach to Automatically Trace Back Clinical Data from Analysis to Source

Presented by Yudong Ren, Clinical Programmer, *Jiangsu Hengrui Pharmaceuticals Co., Ltd.*

Meet the Speaker

Yudong Ren

Title: Clinical Programmer

Organization: Jiangsu Hengrui Pharmaceuticals Co., Ltd.

Yudong Ren, master degree, a statistical programmer with more than 2 years working experience in pharmaceutical industry, interested in using variety of tools to improve work efficacy.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*
- *The author(s) have no real or apparent conflicts of interest to report.*

Agenda

1. What’s traceability in clinical trial data
2. Why do we need traceability
3. How to trace back clinical data from analysis to source automatically and efficiently
4. Reference

What’s traceability in clinical trial data

What's traceability in clinical trial data

CDSIC ADaMIG

Traceability – The property that enables the understanding of the data's lineage and/or the relationship between an element and its predecessor(s). Traceability facilitates transparency, which is an essential component in building confidence in a result or conclusion. Ultimately, traceability permits the understanding of the relationship between the analysis results, the ADaM datasets, the SDTM datasets, and the data collection instrument. Traceability is built by clearly establishing the path between an element and its immediate predecessor. The full path is traced by going from one element to its predecessors, then on to their predecessors, and so on, back to the SDTM datasets, and ultimately to the data collection instrument.

There are two levels of traceability:

1. *Metadata traceability* facilitates the understanding of the relationship of the analysis variable to its source dataset(s) and variable(s) and is required for ADaM compliance. This traceability is established by describing (via metadata) the algorithm used or steps taken to derive or populate an analysis variable from its immediate predecessor. Metadata traceability is also used to establish the relationship between an analysis result and ADaM dataset(s).
2. *Datapoint traceability* points directly to the specific predecessor record(s) and should be implemented if practical and feasible. This level of traceability can be very helpful when trying to trace a complex data manipulation path. This traceability is established by providing clear links in the data (e.g., by use of --a --SEQ variable) to the specific data values used as input for an analysis value. The BDS and OCCDS structures were designed to enable datapoint traceability back to predecessor data.

Why do we need traceability[3]

- Clinical studies are conducted to demonstrate new drugs and therapies are safe and effective
- Study data and analysis results are submitted to agencies
- Traceability is the documentation of steps taken between collected data and the analysis results
- Traceability ensures the analysis results are verifiable

Why do we need traceability[3]

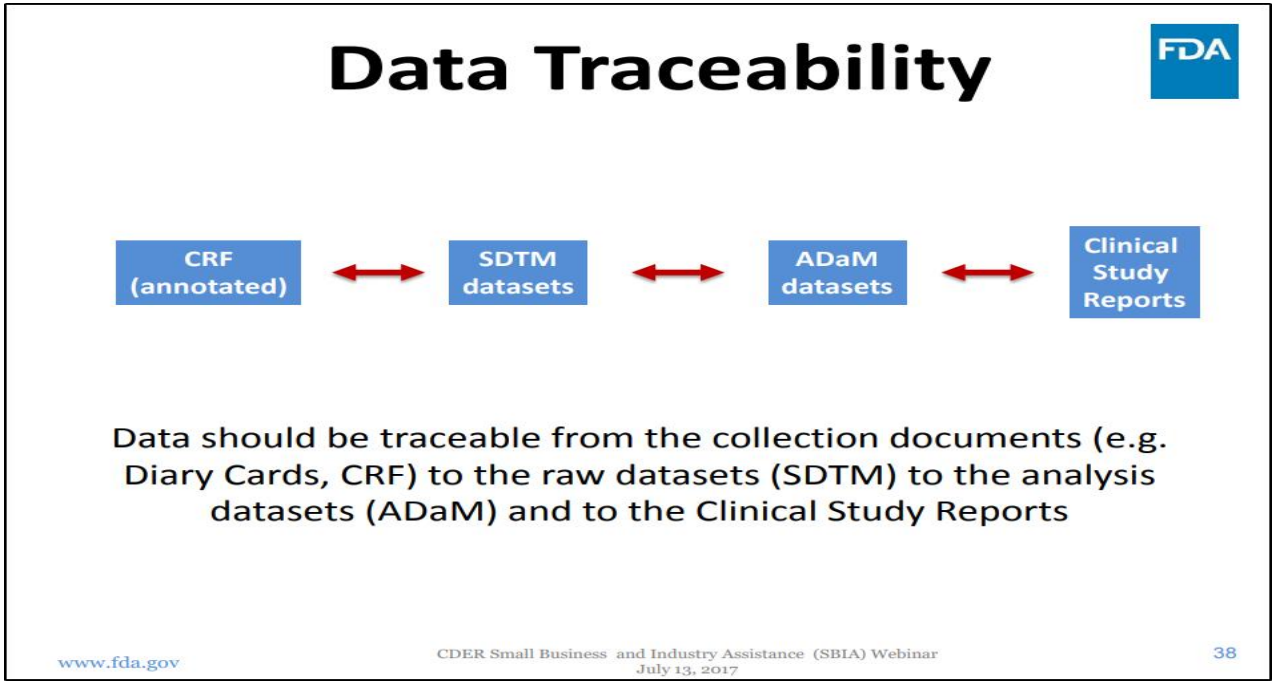
Traceability can answer questions:

- How is a result computed?
- Where is the data supporting a result?
- How is the data derived?
- Which data points in ADaM and SDTM support each subject?

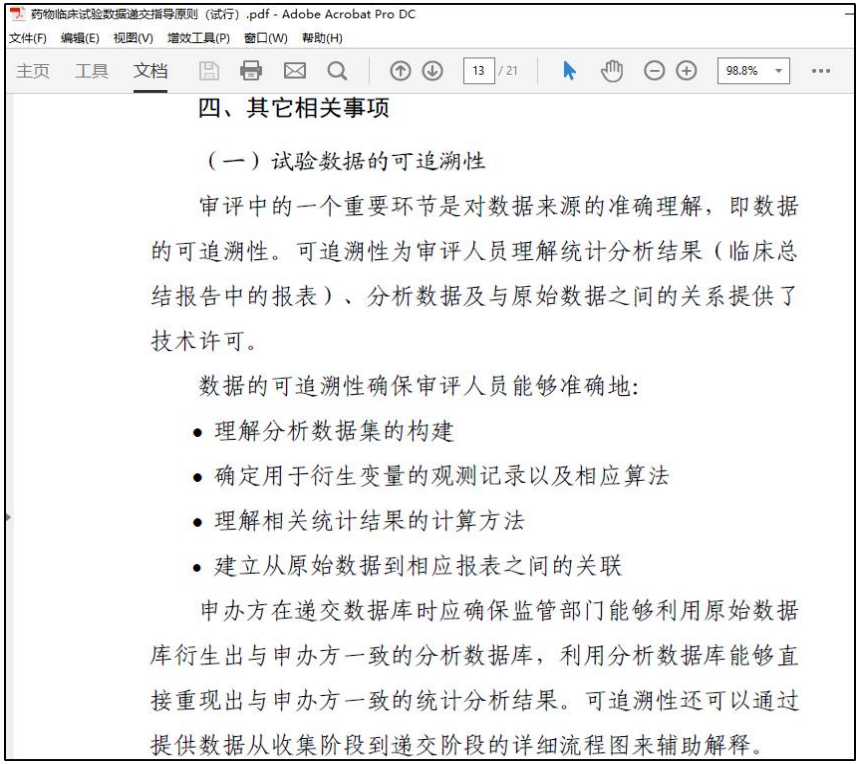
Without traceability, a reviewer must:

- Example submitted program code for answers
- Request meetings with the sponsor for clarification

Why do we need traceability



Why do we need traceability


药物临床试验数据递交指导原则（试行）.pdf - Adobe Acrobat Pro DC

文件(F) 编辑(E) 视图(V) 增效工具(P) 窗口(W) 帮助(H)

主页 工具 文档 13 / 21 98.8%

四、其它相关事项

（一）试验数据的可追溯性

审评中的一个重要环节是对数据来源的准确理解，即数据的可追溯性。可追溯性为审评人员理解统计分析结果（临床总结报告中的报表）、分析数据及与原始数据之间的关系提供了技术许可。

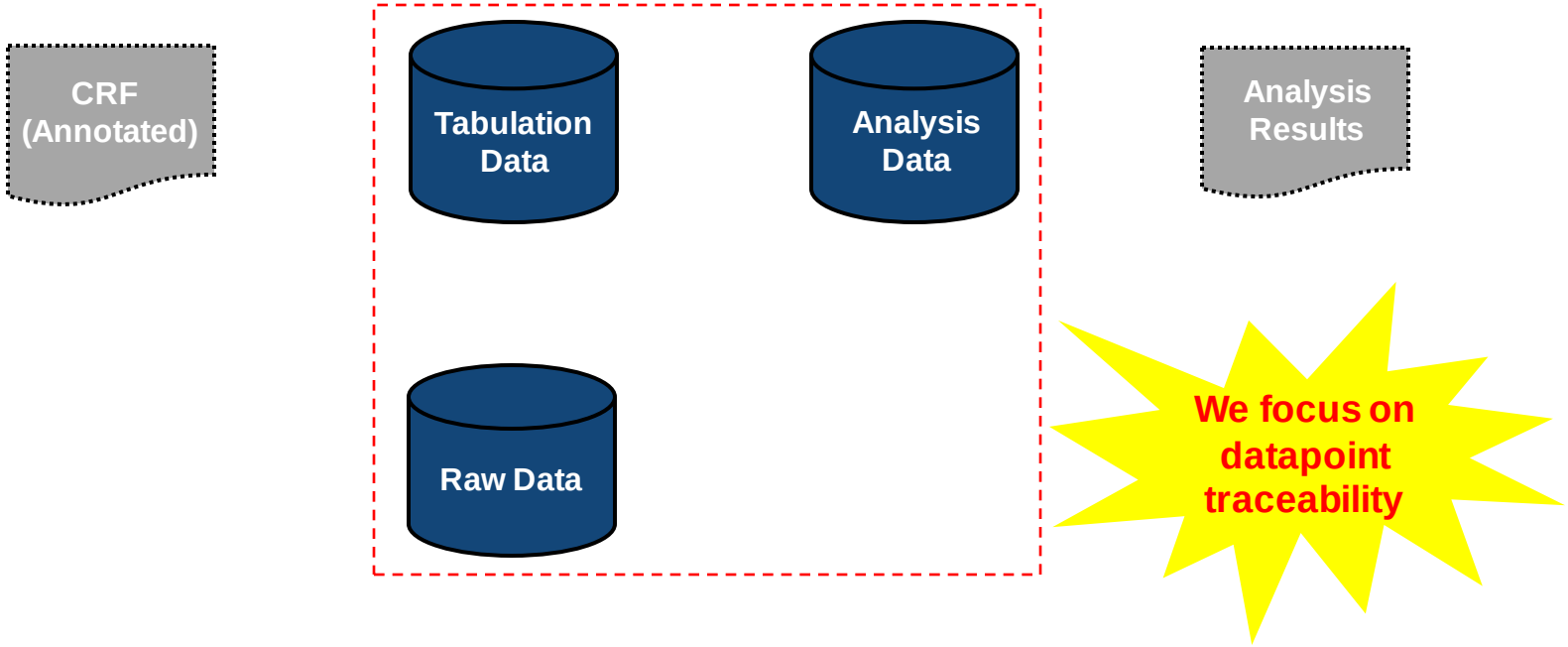
数据的可追溯性确保审评人员能够准确地:

- 理解分析数据集的构建
- 确定用于衍生变量的观测记录以及相应算法
- 理解相关统计结果的计算方法
- 建立从原始数据到相应报表之间的关联

申办方在递交数据库时应确保监管部门能够利用原始数据库衍生出与申办方一致的分析数据库，利用分析数据库能够直接重现出与申办方一致的统计分析结果。可追溯性还可以通过提供数据从收集阶段到递交阶段的详细流程图来辅助解释。

How to trace back clinical data from analysis to source automatically and efficiently

Traceability automation in this presentation:



How to trace back clinical data from analysis to source automatically and efficiently

Traceability Idea + Supportive Automation Tool:

- **Idea**: Build clever traceability supportive data in datasets(e.g. SDTM, ADaM)
- **Tool:** SAS Macro+ VBA

How to trace back clinical data from analysis to source automatically and efficiently

Requirement Analysis

- **Requirement：** Trace clinical data automatically and efficiently

  ADaM (-> ADaMs) -> SDTM(s) -> Raw(s)

- **Difficulties and Solutions:**
  - **Difficulty 1:** Trace clinical data efficiently?

    Solution: Assign Traceability ID for each source record and keep it in variable --SPID.

    --SPID: Sponsor-Defined Identifier

  - **Difficulty 2:** Trace clinical data automatically? (It is impossible to trace data via SAS window)

    Solution: Use Excel datasets(by adding VBA) instead.

    It will open predecessor datasets automatically and filter by value of Traceability ID(s) when click(or find the Traceability ID value on) the cell(s) of --SPID.

How to trace back clinical data from analysis to source automatically and efficiently

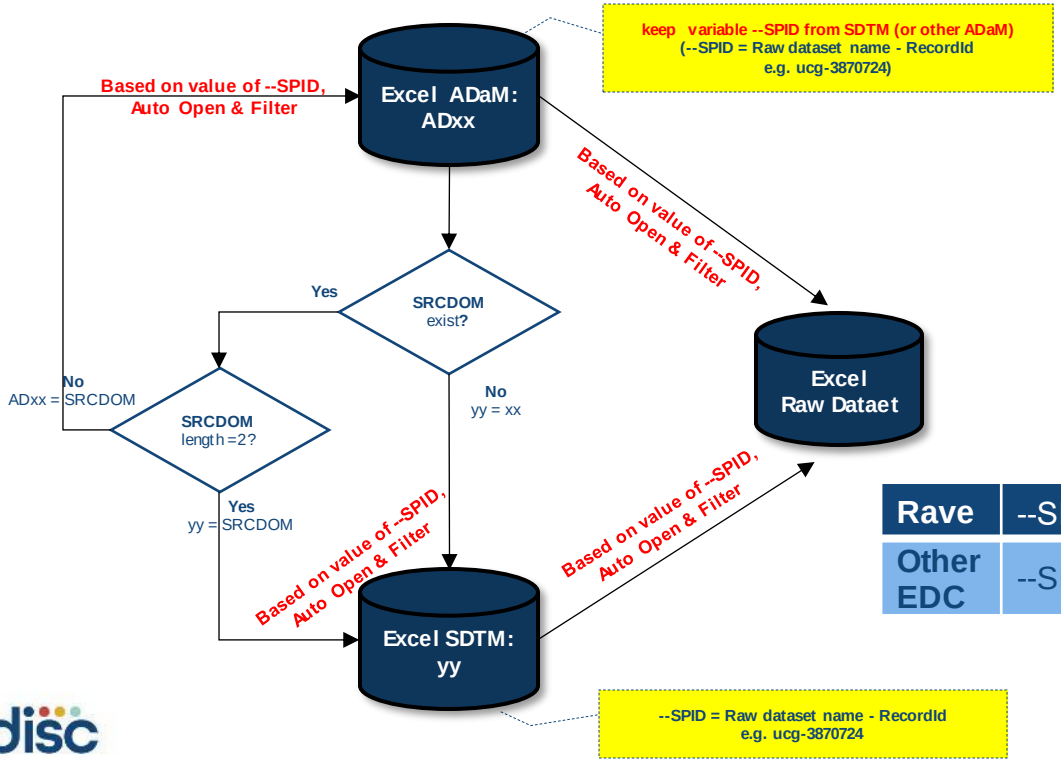
Preparation:

- **SAS Macro Part:** Convert SAS datasets(Raw, SDTM and ADaM) to xlsx.
    - For SDTM, merge main domain and SUPP-- dataset into one single dataset.

- **VBA Part:** Convert .xlsx data files (SDTM and ADaM) to .xlsm and add supportive VBA Codes (batch processing).

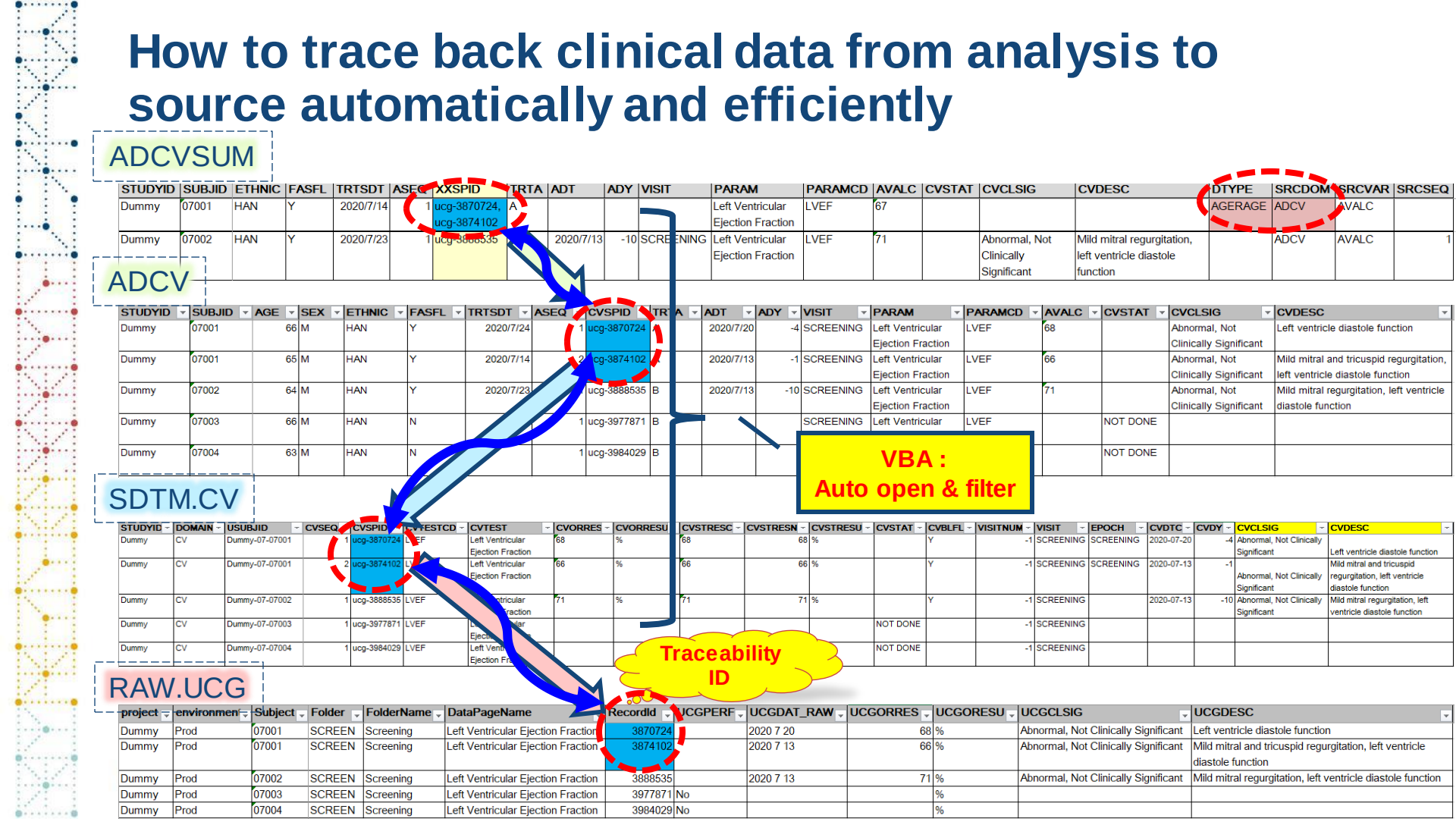
How to trace back clinical data from analysis to source automatically and efficiently

Automation Judgment Rule:



Rave	--SPID value = “Form Name - RecordId”
Other EDC	--SPID value = “Form Name - Its record ID var”

How to trace back clinical data from analysis to source automatically and efficiently



ADCVSUM

STUDYID	SUBJID	ETHNIC	FASFL	TRTSDT	ASEQ	XXSPID	TRTA	ADT	ADY	VISIT	PARAM	PARAMCD	AVALC	CVSTAT	CVCLSIG	CVDESC	DTYPE	SRCDOM	SRCVAR	SRCSEQ
Dummy	07001	HAN	Y	2020/7/14	1	ucg-3870724, ucg-3874102	A				Left Ventricular Ejection Fraction	LVEF	67				AGERAGE	ADCV	AVALC	
Dummy	07002	HAN	Y	2020/7/23	1	ucg-3888535		2020/7/13	-10	SCREENING	Left Ventricular Ejection Fraction	LVEF	71		Abnormal, Not Clinically Significant	Mild mitral regurgitation, left ventricle diastole function		ADCV	AVALC	1

ADCV

STUDYID	SUBJID	AGE	SEX	ETHNIC	FASFL	TRTSDT	ASEQ	CVSPID	TRTA	ADT	ADY	VISIT	PARAM	PARAMCD	AVALC	CVSTAT	CVCLSIG	CVDESC
Dummy	07001	66	M	HAN	Y	2020/7/24	1	ucg-3870724	A	2020/7/20	-4	SCREENING	Left Ventricular Ejection Fraction	LVEF	68		Abnormal, Not Clinically Significant	Left ventricle diastole function
Dummy	07001	65	M	HAN	Y	2020/7/14	2	ucg-3874102	A	2020/7/13	-1	SCREENING	Left Ventricular Ejection Fraction	LVEF	66		Abnormal, Not Clinically Significant	Mild mitral and tricuspid regurgitation, left ventricle diastole function
Dummy	07002	64	M	HAN	Y	2020/7/23	1	ucg-3888535	B	2020/7/13	-10	SCREENING	Left Ventricular Ejection Fraction	LVEF	71		Abnormal, Not Clinically Significant	Mild mitral regurgitation, left ventricle diastole function
Dummy	07003	66	M	HAN	N		1	ucg-3977871	B			SCREENING	Left Ventricular	LVEF		NOT DONE		
Dummy	07004	63	M	HAN	N		1	ucg-3984029	B							NOT DONE		

SDTM.CV

STUDYID	DOMAIN	USUBJID	CVSEQ	CVSPID	CVTESTCD	CVTEST	CVORRES	CVORRESU	CVSTRESC	CVSTRESN	CVSTRESU	CVSTAT	CVBLFL	VISITNUM	VISIT	EPOCH	CVDTC	CVDY	CVCLSIG	CVDESC
Dummy	CV	Dummy-07-07001	1	ucg-3870724	LVEF	Left Ventricular Ejection Fraction	68	%	68	68	%		Y	-1	SCREENING	SCREENING	2020-07-20	-4	Abnormal, Not Clinically Significant	Left ventricle diastole function
Dummy	CV	Dummy-07-07001	2	ucg-3874102	LVEF	Left Ventricular Ejection Fraction	66	%	66	66	%		Y	-1	SCREENING	SCREENING	2020-07-13	-1	Abnormal, Not Clinically Significant	Mild mitral and tricuspid regurgitation, left ventricle diastole function
Dummy	CV	Dummy-07-07002	1	ucg-3888535	LVEF	Left Ventricular Ejection Fraction	71	%	71	71	%		Y	-1	SCREENING		2020-07-13	-10	Abnormal, Not Clinically Significant	Mild mitral regurgitation, left ventricle diastole function
Dummy	CV	Dummy-07-07003	1	ucg-3977871	LVEF	Left Ventricular Ejection Fraction						NOT DONE		-1	SCREENING					
Dummy	CV	Dummy-07-07004	1	ucg-3984029	LVEF	Left Ventricular Ejection Fraction						NOT DONE		-1	SCREENING					

RAW.UCG

project	environmen	Subject	Folder	FolderName	DataPageName	RecordId	UCGPERF	UCGDAT_RAW	UCGORRES	UCGORESU	UCGCLSIG	UCGDESC
Dummy	Prod	07001	SCREEN	Screening	Left Ventricular Ejection Fraction	3870724		2020 7 20	68	%	Abnormal, Not Clinically Significant	Left ventricle diastole function
Dummy	Prod	07001	SCREEN	Screening	Left Ventricular Ejection Fraction	3874102		2020 7 13	66	%	Abnormal, Not Clinically Significant	Mild mitral and tricuspid regurgitation, left ventricle diastole function
Dummy	Prod	07002	SCREEN	Screening	Left Ventricular Ejection Fraction	3888535		2020 7 13	71	%	Abnormal, Not Clinically Significant	Mild mitral regurgitation, left ventricle diastole function
Dummy	Prod	07003	SCREEN	Screening	Left Ventricular Ejection Fraction	3977871	No			%		
Dummy	Prod	07004	SCREEN	Screening	Left Ventricular Ejection Fraction	3984029	No			%		

How to trace back clinical data from analysis to source automatically and efficiently

How to trace back clinical data from analysis to source automatically and efficiently

Some Note:

- This idea can be applied to trace data from source to analysis.
- Need to consider the value of STUDYID when trace data in ISS/ISE study.
- More information can be added to Traceability ID(--SPID), such as SUBJID, VISIT(FolderName), DataPageName
- --SPID can be dropped from datasets when submit, data flow and integrity won't be affected.
- This idea can be used to trace Pinnacle 21 issue to related records of SDTM/ADaM, and then check if they are data issue or mapping issue more efficiently.

Reference

[1]. CDSIC: ADaMIG v1.3

[2]. FDA: Final - Study Data Technical Conformance Guide v4.8.1_Oct2021

[3]. PharmaSUG SDE RTP 2020: More Traceability: Clarity in ADaM Metadata and Beyond. Richann Watson, Wayne Zhong, Daphne Ewing, Jasmine Zhang

[4]. PhUSE SDE Shanghai 2020: Tracecart - Trace Back Clinical Data from Analysis to Source Automatically and Efficiently. Haiqiang Luo

Acknowledgement

罗海强 (Haiqiang Luo) : Perfect leader and teacher

Thank You!

Contact：

Email: yudong.ren@hengrui.com

WeChat: r15114608662