phuse·share
connect·share·advance

Single Day Event

Evolving around the Impacts of Automating Statistical Programming Tasks

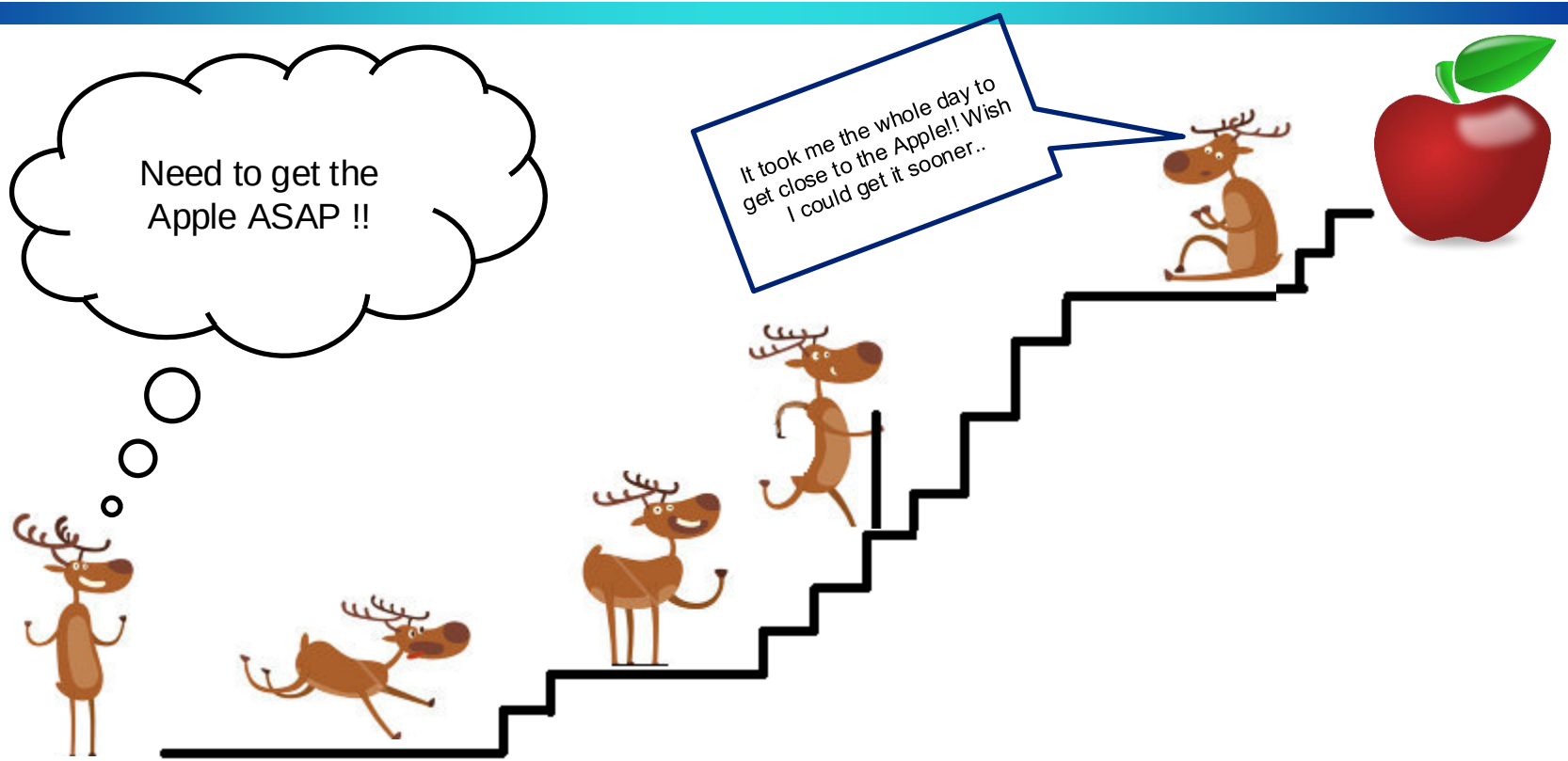
Diganta Bose,
Associate Director, AOD Operations, FSP
Cytel Inc.

phuse.global

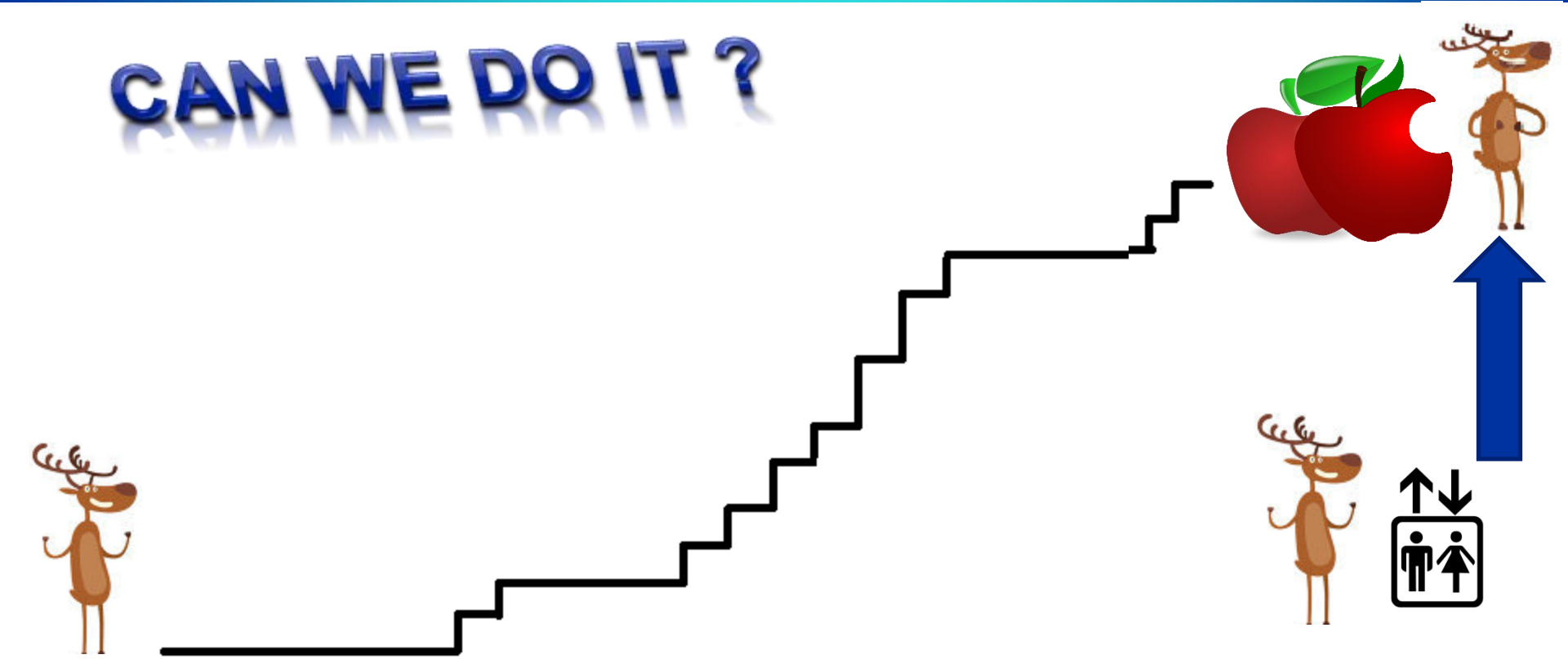
Disclaimer

The purpose of this presentation is to share ideas specific to the field of Statistical Programming for Clinical Trials. The examples and tools discussed are not part of any commercial product but originate from personal, self-driven initiatives aimed at contributing to the community and fostering knowledge exchange. The views and perspectives expressed in this presentation are solely those of the presenter and do not necessarily represent the official stance or opinions of Cytel as an organization.

“Automation” – a Naïve Depiction



“Automation” – a Naïve Depiction



SAS Utility Macros

```
proc sql noprint;
   select count(distinct usubjid) into:bigN from lb;
quit;

%macro labshift(test=,dsout=);
 proc sql;
    create table _tmp_ as
       select a.usubjid,a.lbtest,a.base,b.lpostbl
       from (select usubjid,lbtest,lbnrind as base from lb
                where week=0 and lbtest="&test") as a
          left join
          (select usubjid,lbnrind as lpostbl from lb
           where week=2 and lbtest="&test") as b
       on a.usubjid=b.usubjid
       order by usubjid;
 quit;

 %xtab_nsubj(inds=_tmp_,roweqcol=yes,iscolchar=yes,
                cols=L N H,rvar=base,cvar=lpostbl,
                bign=&bigN,outds=&dsout);
%mend;

%labshift(test=HEMATOCRIT,dsout=hct_shift);
%labshift(test=HEMOGLOBIN,dsout=hgb_shift);
```

Analysis Ready Input Data

tmp Dataset

	usubjid	lbtest	base	lpostbl
1	101	HEMATOCRIT	L	N
2	102	HEMATOCRIT	N	N
3	103	HEMATOCRIT	H	H
4	104	HEMATOCRIT	H	N
5	105	HEMATOCRIT	N	N
6	106	HEMATOCRIT	H	H
7	107	HEMATOCRIT	H	H
8	108	HEMATOCRIT	N	H
9	109	HEMATOCRIT	N	H
10	110	HEMATOCRIT	N	H

Utility Macros

Hematocrit

	ord	c0	c1	c2	Total
1	r0	0 (0.0)	1 (10.0)	0 (0.0)	1 (10.0)
2	r1	0 (0.0)	2 (20.0)	3 (30.0)	5 (50.0)
3	r2	0 (0.0)	1 (10.0)	3 (30.0)	4 (40.0)
4	Total	0 (0.0)	4 (40.0)	6 (60.0)	10 (100.0)

Hemoglobin

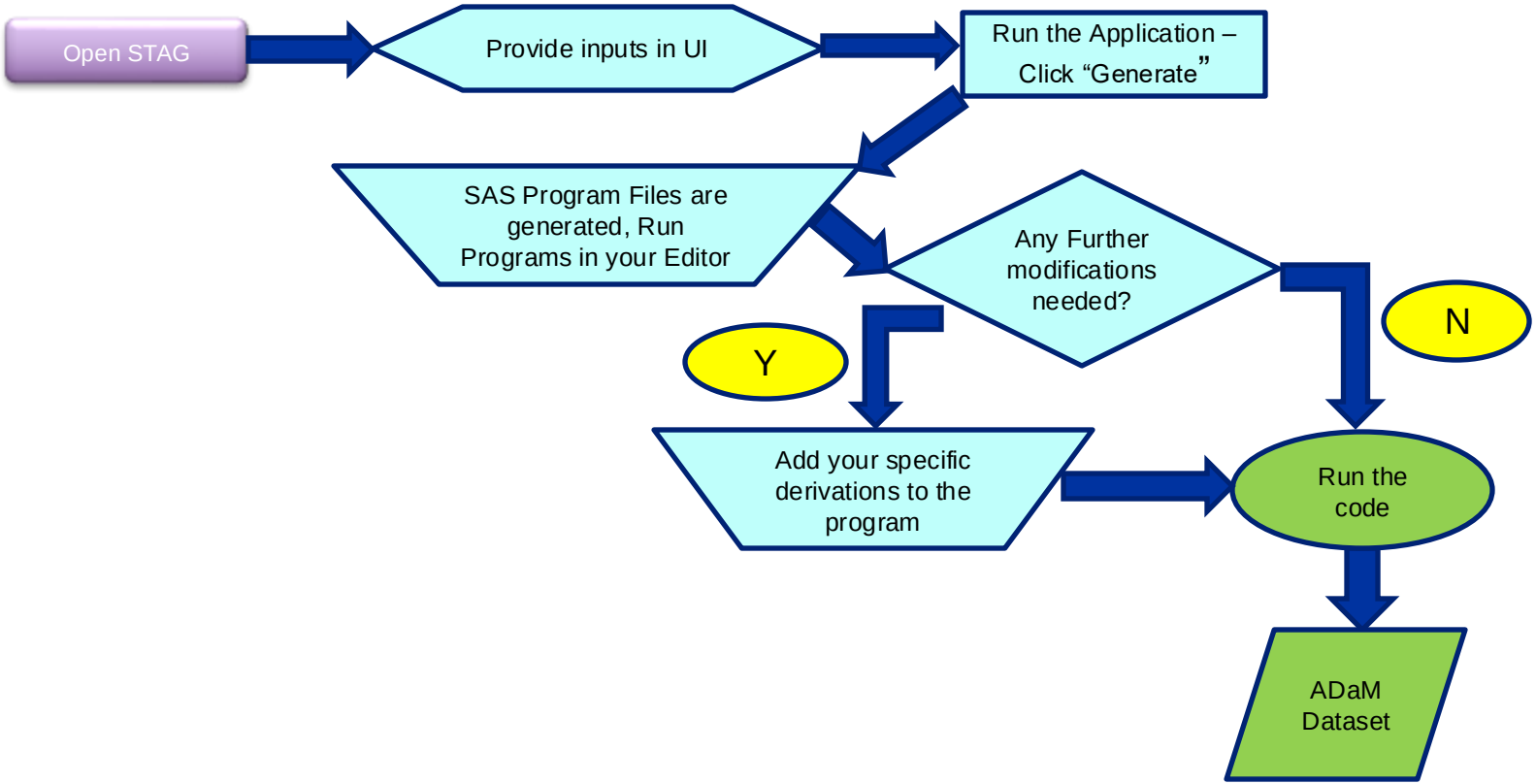
	ord	c0	c1	c2	Total
1	r0	0 (0.0)	1 (10.0)	1 (10.0)	2 (20.0)
2	r1	0 (0.0)	4 (40.0)	4 (40.0)	8 (80.0)
3	r2	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
4	Total	0 (0.0)	5 (50.0)	5 (50.0)	10 (100.0)

Automate CDISC ADaM ?

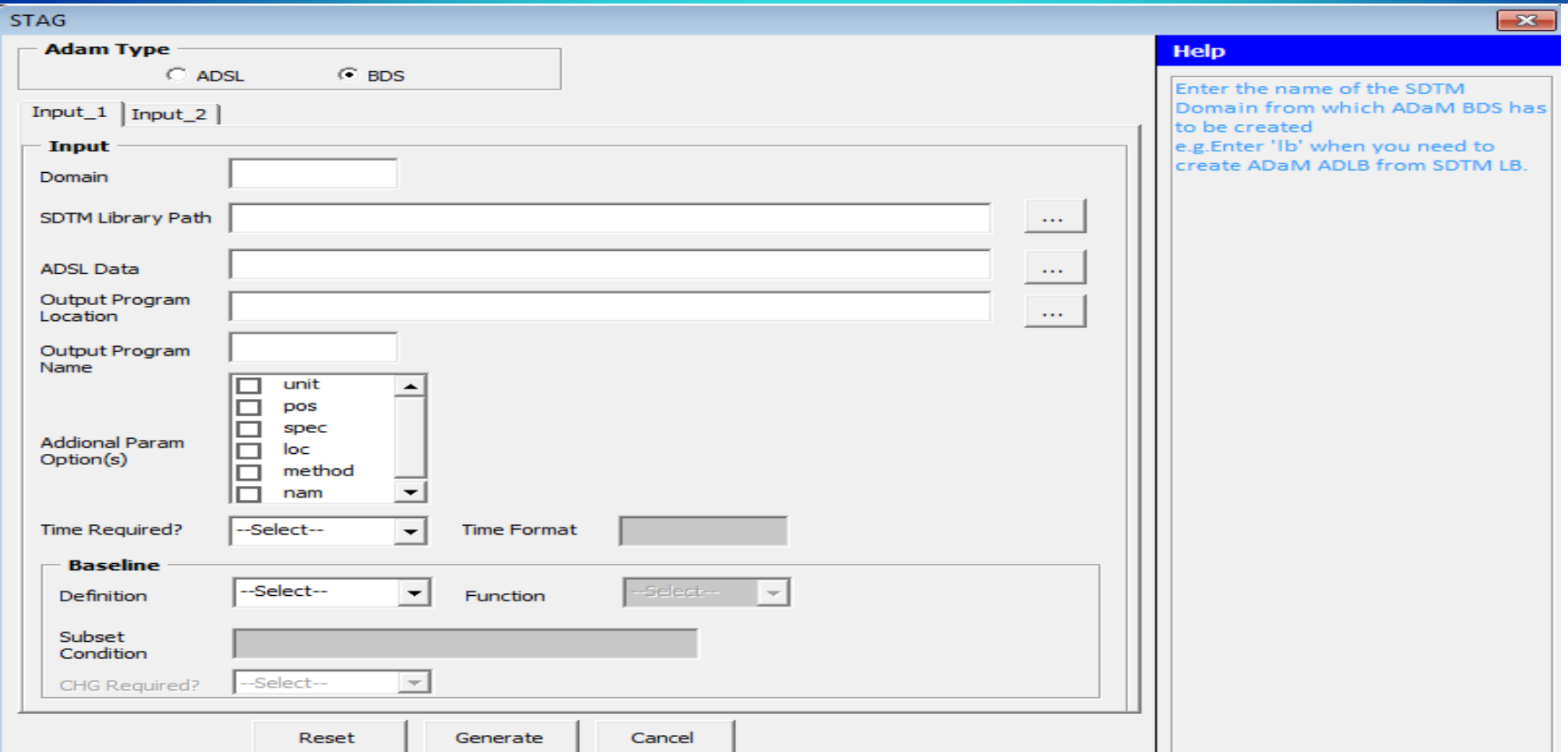
Highlights of the Utility – STAG (SDTM To ADaM Generator)

- Tool based ADaM Generation
- The commonly used key ADSL and BDS variables
- Front end UI
- SAS Macros/programs at the Backend

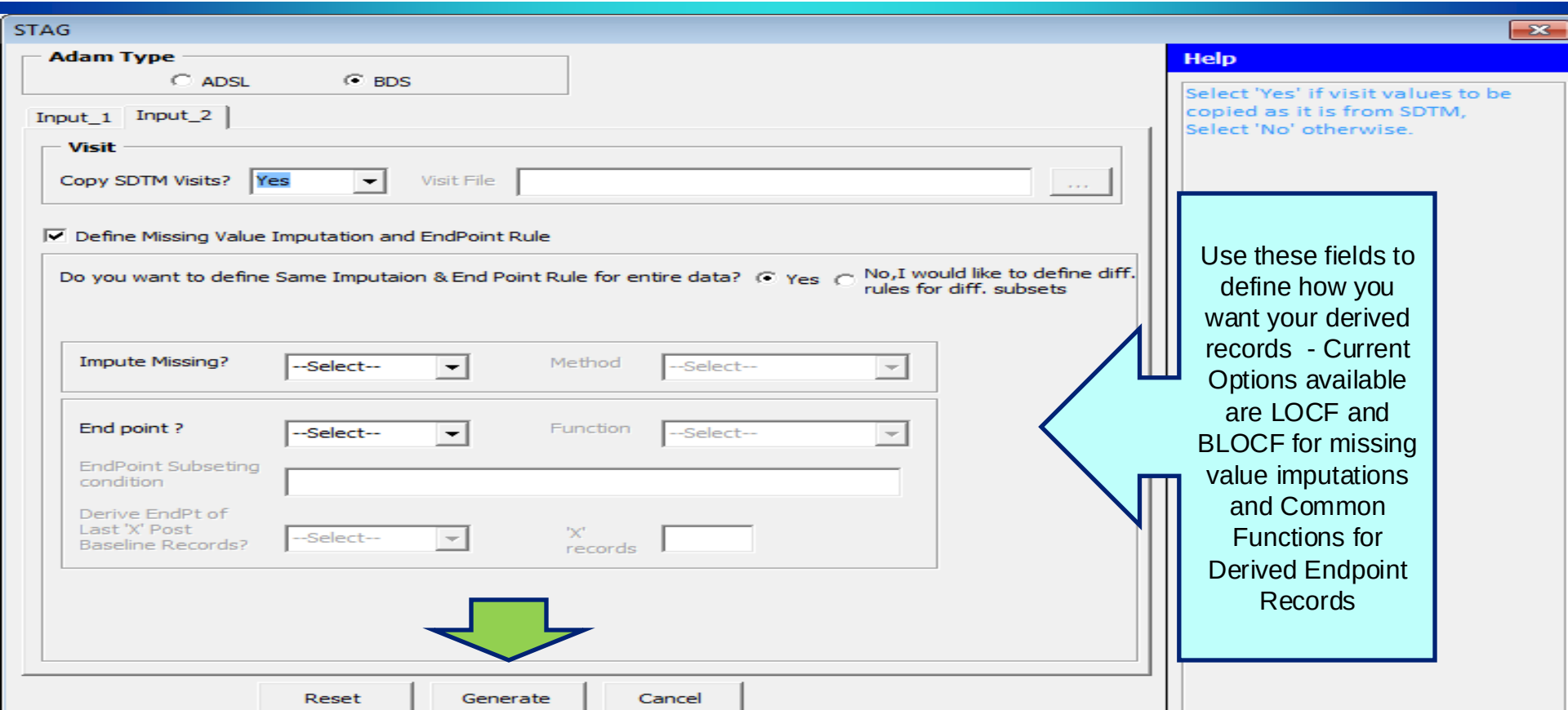
STAG Overview



STAG UI Screen BDS Screen 1

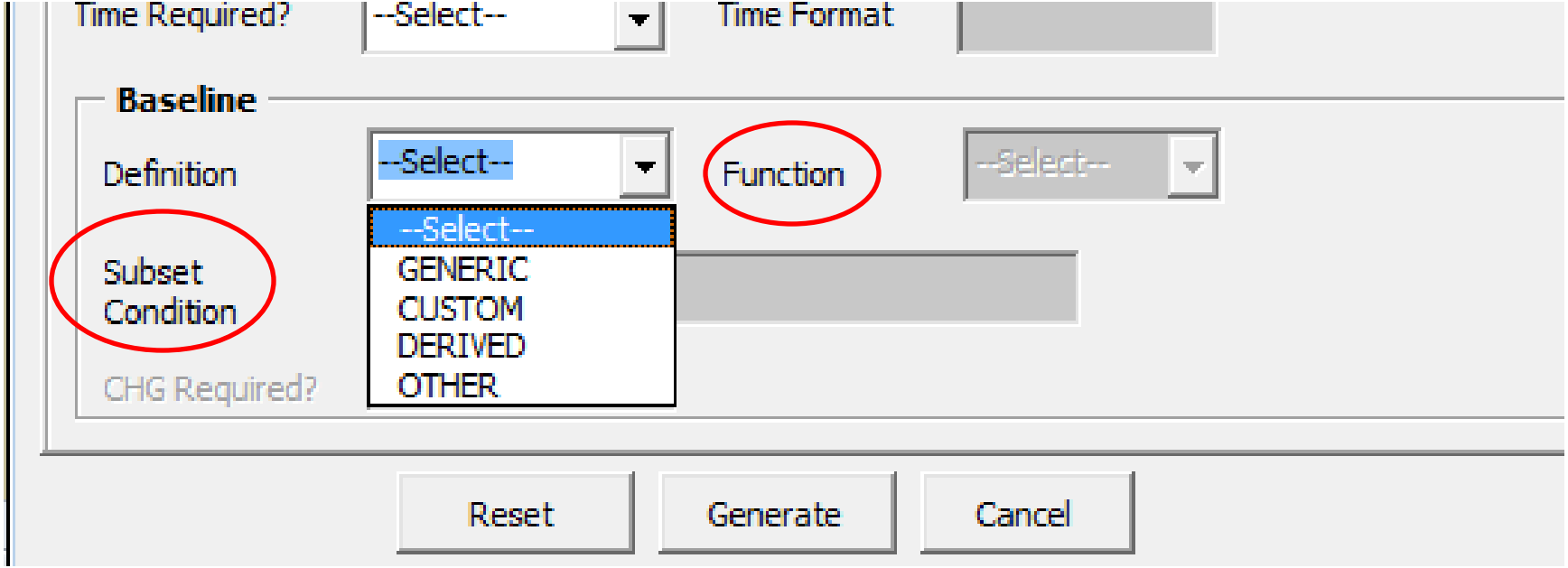


STAG UI Screen BDS Screen 2



BDS Demo

Derivation of Baseline related variables may need you to focus on Certain Terminologies used in STAG to understand your study's Baseline Definition better

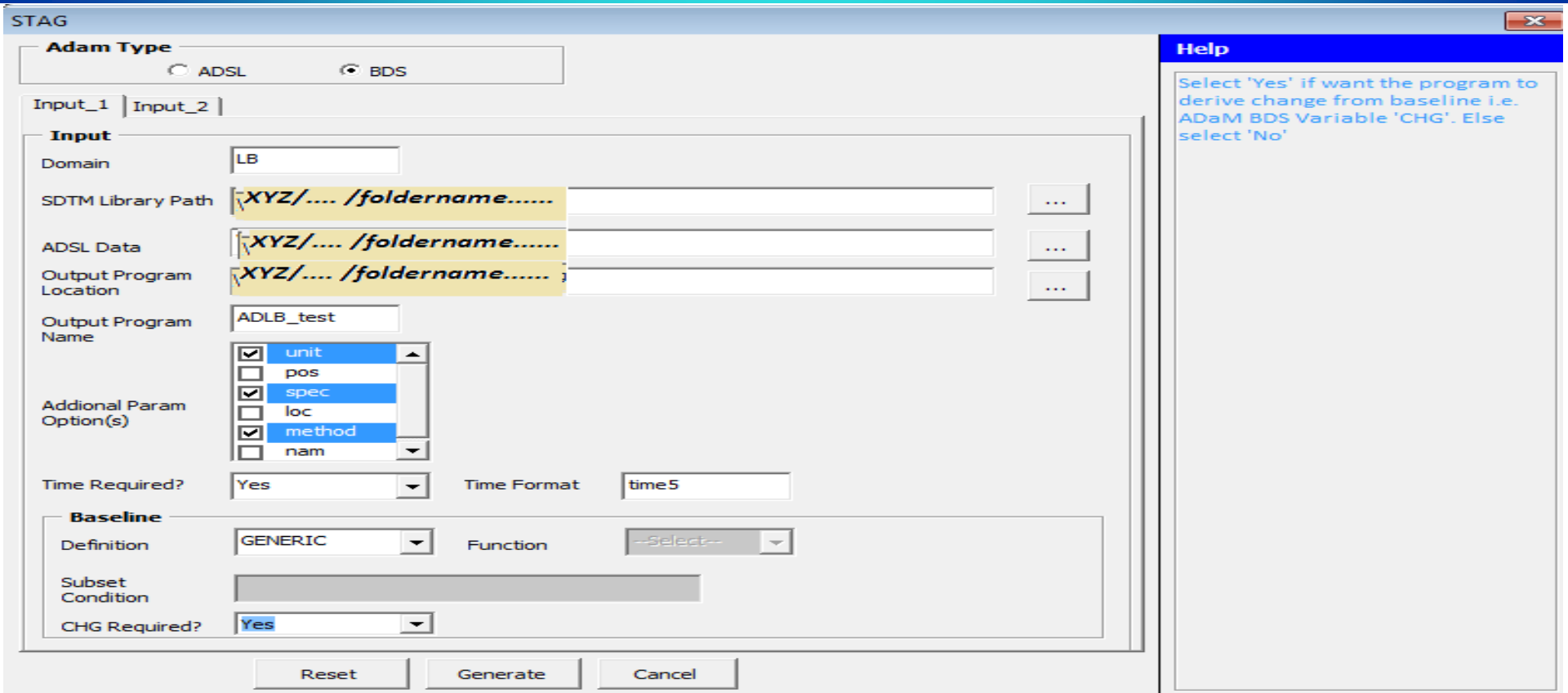


BDS Demo ADLB

STUDYID	USUBJID	LBSEQ	LBTESTCD	LBTEST	LBCAT	LBORRES	LBSTRESC	LBSTRESN	LBSTRESU	LBSPEC	VISIT
CDISC01	CDISC01.100008	1	BILI	Bilirubin	CHEMISTRY	0.4	6.8	6.8	umol/L	BLOOD	SCREEN
CDISC01	CDISC01.100008	2	BILI	Bilirubin	CHEMISTRY	0.3	5.1	5.1	umol/L	BLOOD	WEEK 24
CDISC01	CDISC01.100008	3	BUN	Blood Urea Nitro...	CHEMISTRY	26	9.28	9.28	mmol/L	BLOOD	SCREEN
CDISC01	CDISC01.100008	4	BUN	Blood Urea Nitro...	CHEMISTRY	18	6.43	6.43	mmol/L	BLOOD	WEEK 24
CDISC01	CDISC01.100008	5	GLUC	Glucose	CHEMISTRY	100	5.5	5.5	mmol/L	BLOOD	SCREEN
CDISC01	CDISC01.100008	6	GLUC	Glucose	CHEMISTRY	87	4.8	4.8	mmol/L	BLOOD	WEEK 24
CDISC01	CDISC01.100008	7	VITB12	Vitamin B12	CHEMISTRY	366	270	270	pmol/L	SERUM	SCREEN
CDISC01	CDISC01.100008	8	VITB9	Vitamin B9	CHEMISTRY	55.8	126.4	126.4	nmol/L	BLOOD	SCREEN
CDISC01	CDISC01.100008	9	HCT	Hematocrit	HEMATOLOGY	36.2	0.36	0.36		BLOOD	SCREEN
CDISC01	CDISC01.100008	10	HCT	Hematocrit	HEMATOLOGY	32.9	0.33	0.33		BLOOD	WEEK 24
CDISC01	CDISC01.100008	11	HGB	Hemoglobin	HEMATOLOGY	12.0	120	120	g/L	BLOOD	SCREEN
CDISC01	CDISC01.100008	12	HGB	Hemoglobin	HEMATOLOGY	11.3	113	113	g/L	BLOOD	WEEK 24
CDISC01	CDISC01.100008	13	LYM	Lymphocytes	HEMATOLOGY	1.68	1.68	1.68	10**9/L	BLOOD	SCREEN
CDISC01	CDISC01.100008	14	LYM	Lymphocytes	HEMATOLOGY	1.34	1.34	1.34	10**9/L	BLOOD	WEEK 24
CDISC01	CDISC01.100008	15	OCCBLD	Occult Blood	URINALYSIS	1+	1+	.		URINE	SCREEN
CDISC01	CDISC01.100008	16	OCCBLD	Occult Blood	URINALYSIS	NEGATIVE	NEGATIVE	.		URINE	WEEK 24
CDISC01	CDISC01.100008	17	PH	pH	URINALYSIS	8.0	8	8		URINE	SCREEN
CDISC01	CDISC01.100008	18	PH	pH	URINALYSIS	8.0	8	8		URINE	WEEK 24
CDISC01	CDISC01.100008	19	GLUC	Glucose	URINALYSIS	NEGATIVE	NEGATIVE	.		URINE	SCREEN
CDISC01	CDISC01.100008	20	GLUC	Glucose	URINALYSIS	TRACE	TRACE	.		URINE	WEEK 24
CDISC01	CDISC01.100008	21	GLUC	Glucose	URINALYSIS	10	0.6	0.6	mmol/L	URINE	SCREEN
CDISC01	CDISC01.100008	22	GLUC	Glucose	URINALYSIS	30	1.7	1.7	mmol/L	URINE	WEEK 24

0.6

BDS Demo ADLB



BDS Demo ADLB

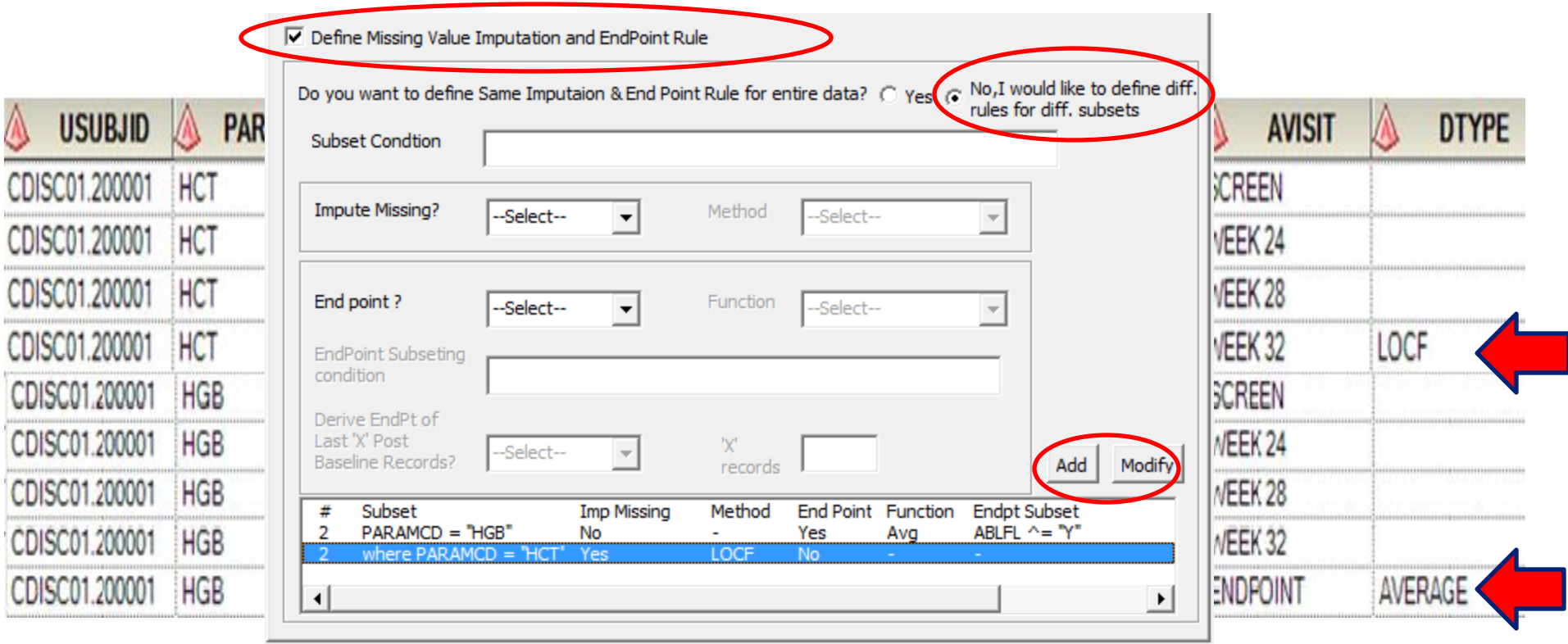
```
libname sdtm_dt 'XYZ/.... /foldername.....
libname adam_dt 'XYZ/.... /foldername.....
%include 'XYZ/.... /foldername.....
%include 'XYZ/.... /foldername.....
%include 'XYZ/.... /foldername.....
%include 'XYZ/.... /foldername.....

/*%generic_bds0x Macro Call*/
%generic_bds0x(sdtmlib=sdtm_dt, domain=LB, add= unit spec method,adsldata=adam_dt.adsl, basedef=GENE

/*Avisit Macro call*/
%avisit(indata=work.bds_LB, copy_sdtmvis=Yes);

data bds_out;
  set bds_vis_out;
run;

/*%generic_bds0y Macro Call*/
%generic_bds0y(inds=bds_out, sdtmlib=s

data pre_bds_LB;
  set bds_Out_der_recs;
run;
```

Run the piece of code generated by STAG to assess to what extent ADaM ADLB variables are derived

BDS Demo ADLB

USUBJID	PARAM	PARAMN	PARAMCD	AVAL	AVALC	ADT	ATM	ADTM	ABLFL	BASE	BASEC	CHG
CDISC01.100008	Bilirubin (umol/L) (BLOOD)	1	BILI	6.8	6.8	15APR2003	11:20	15APR03:11:20:...	Y	6.8	6.8	0
CDISC01.100008	Bilirubin (umol/L) (BLOOD)	1	BILI	5.1	5.1	13OCT2003	11:55	13OCT03:11:55:...		6.8	6.8	-1.7
CDISC01.100008	Blood Urea Nitrogen (mmol/L) (BLOOD)	2	BUN	9.28	9.28	15APR2003	11:20	15APR03:11:20:...	Y	9.28	9.28	0
CDISC01.100008	Blood Urea Nitrogen (mmol/L) (BLOOD)	2	BUN	6.43	6.43	13OCT2003	11:55	13OCT03:11:55:...		9.28	9.28	-2.85
CDISC01.100008	Glucose (URINE) (DIPSTICK)	3	GLUC1	.	NEGATIVE	15APR2003	11:20	15APR03:11:20:...	Y	.	NEGATIVE	.
CDISC01.100008	Glucose (URINE) (DIPSTICK)	3	GLUC1	.	TRACE	13OCT2003	11:55	13OCT03:11:55:...		.	NEGATIVE	.
CDISC01.100008	Glucose (mmol/L) (BLOOD)	4	GLUC2	5.5	5.5	15APR2003	11:20	15APR03:11:20:...	Y	5.5	5.5	0
CDISC01.100008	Glucose (mmol/L) (BLOOD)	4	GLUC2	4.8	4.8	13OCT2003	11:55	13OCT03:11:55:...		5.5	5.5	-0.7
CDISC01.100008	Glucose (mmol/L) (URINE) (QUANT)	5	GLUC3	0.6	0.6	15APR2003	11:20	15APR03:11:20:...	Y	0.6	0.6	0
CDISC01.100008	Glucose (mmol/L) (URINE) (QUANT)	5	GLUC3	1.7	1.7	13OCT2003	11:55	13OCT03:11:55:...		0.6	0.6	1.1
CDISC01.100008	Hematocrit (BLOOD)	6	HCT	0.36	0.36	15APR2003	11:20	15APR03:11:20:...	Y	0.36	0.36	0
CDISC01.100008	Hematocrit (BLOOD)	6	HCT	0.33	0.33	13OCT2003	11:55	13OCT03:11:55:...		0.36	0.36	-0.03
CDISC01.100008	Hemoglobin (g/L) (BLOOD)	7	HGB	120	120	15APR2003	11:20	15APR03:11:20:...	Y	120	120	0
CDISC01.100008	Hemoglobin (g/L) (BLOOD)	7	HGB	113	113	13OCT2003	11:55	13OCT03:11:55:...		120	120	-7
CDISC01.100008	Lymphocytes (10**9/L) (BLOOD)	8	LYM	1.68	1.68	15APR2003	11:20	15APR03:11:20:...	Y	1.68	1.68	0
CDISC01.100008	Lymphocytes (10**9/L) (BLOOD)	8	LYM	1.34	1.34	13OCT2003	11:55	13OCT03:11:55:...		1.68	1.68	-0.34
CDISC01.100008	Occult Blood (URINE)	9	OCCBLD	.	1+	15APR2003	11:20	15APR03:11:20:...	Y	.	1+	.
CDISC01.100008	Occult Blood (URINE)	9	OCCBLD	.	NEGATIVE	13OCT2003	11:55	13OCT03:11:55:...		.	1+	.
CDISC01.100008	Vitamin B12 (pmol/L) (SERUM)	10	VITB12	270	270	15APR2003	11:20	15APR03:11:20:...	Y	270	270	0
CDISC01.100008	Vitamin B9 (nmol/L) (BLOOD)	11	VITB9	126.4	126.4	15APR2003	11:20	15APR03:11:20:...	Y	126.4	126.4	0
CDISC01.100008	pH (URINE)	12	PH	8	8	15APR2003	11:20	15APR03:11:20:...	Y	8	8	0
CDISC01.100008	pH (URINE)	12	PH	8	8	13OCT2003	11:55	13OCT03:11:55:...		8	8	0

BDS Demo ADLB



Challenges in Automation

- Programming Complexity
- Preserving core competencies
- Effort estimations and Implementation Costs
- Addressing new Job Creation

Programming Complexity

Use of the SAS Macro Language

Use of Regular Expressions in SAS

Nested Queries in proc SQL

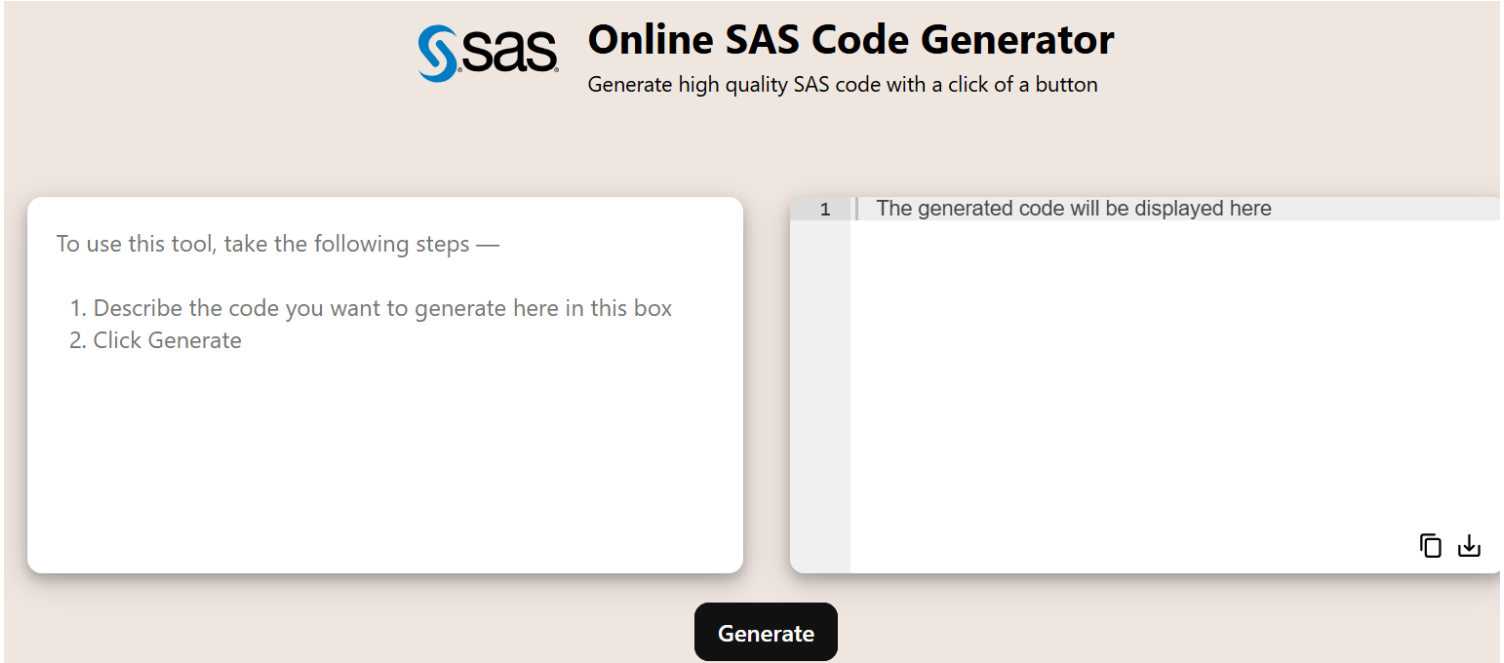
System Programming Techniques

Preserving core competencies

- Core competencies / Understanding underlying concepts
- Cognitive Abilities of Problem Solving/Debug
- Need for thorough Validation
- Sometimes Automated Processes are less flexible than the Human touch

Preserving core competencies

https://www.codeconvert.ai/sas-code-generator



Preserving core competencies

https://www.codeconvert.ai/sas-code-generator -Asked the AI tool to generate a Macro to check log files for errors, warnings

```
%macro check_log(log_file);
    filename mylog "&log_file";

    data _null_;
        infile mylog;
        input;
        if _infile_ contains 'ERROR' then do;
            put 'ERROR found: ' _infile_;
            call symputx('error_found', '1');
        end;
        else if _infile_ contains 'WARNING' then do;
            put 'WARNING found: ' _infile_;
            call symputx('warning_found', '1');
        end;
    run;

    %if &error_found = 1 %then %do;
        %put ERROR: Errors were found in the log file.;
    %end;
    %if &warning_found = 1 %then %do;
        %put WARNING: Warnings were found in the log
    file.;
    %end;

    filename mylog clear;%mend
    check_log;%check_log('path_to_your_log_file.log');
```

How a Handcrafted Code for checking log file looks like ?

Effort estimations and Implementation Costs

- Validators/QC Programmers often need to write code from scratch
- Often at Higher Level, effort is underestimated.
- Initial Implementation costs daunting
- Evaluate implementation costs will be balanced out by the future savings.

Addressing new Job Creation

- Additional Learning curve – for developers as well as users.
- Some Jobs may be automated – Job Cut
- Human workers to focus on more complex, creative, and strategic tasks
- New Job Creation →design, maintain, and oversee automated systems

Concluding Remarks

“World is changing – adapt, implement and improvise”

Let’s value solutions that increase productivity

But at the Same time

Let’s appreciate those solutions in a balanced and responsible way without diminishing our own cognitive abilities

The Global Healthcare Data Science Community

Contact Channels

UK +44 1843 609600
office@phuse.global
phuse.global

Social Media

@phusetwitta
/phusebook
/phusetube
/company/phuse