



Save Your Sights (SYS)

Vikram Venugopal, Statistical Analyst II

N Jayapandian, Lead Statistical Analyst

Hyderabad, 23FEB2013



The views and opinions expressed in the following slides are those of the individual presenter and should not be attributed to any organization with which the presenter is employed or affiliated



Agenda

Background

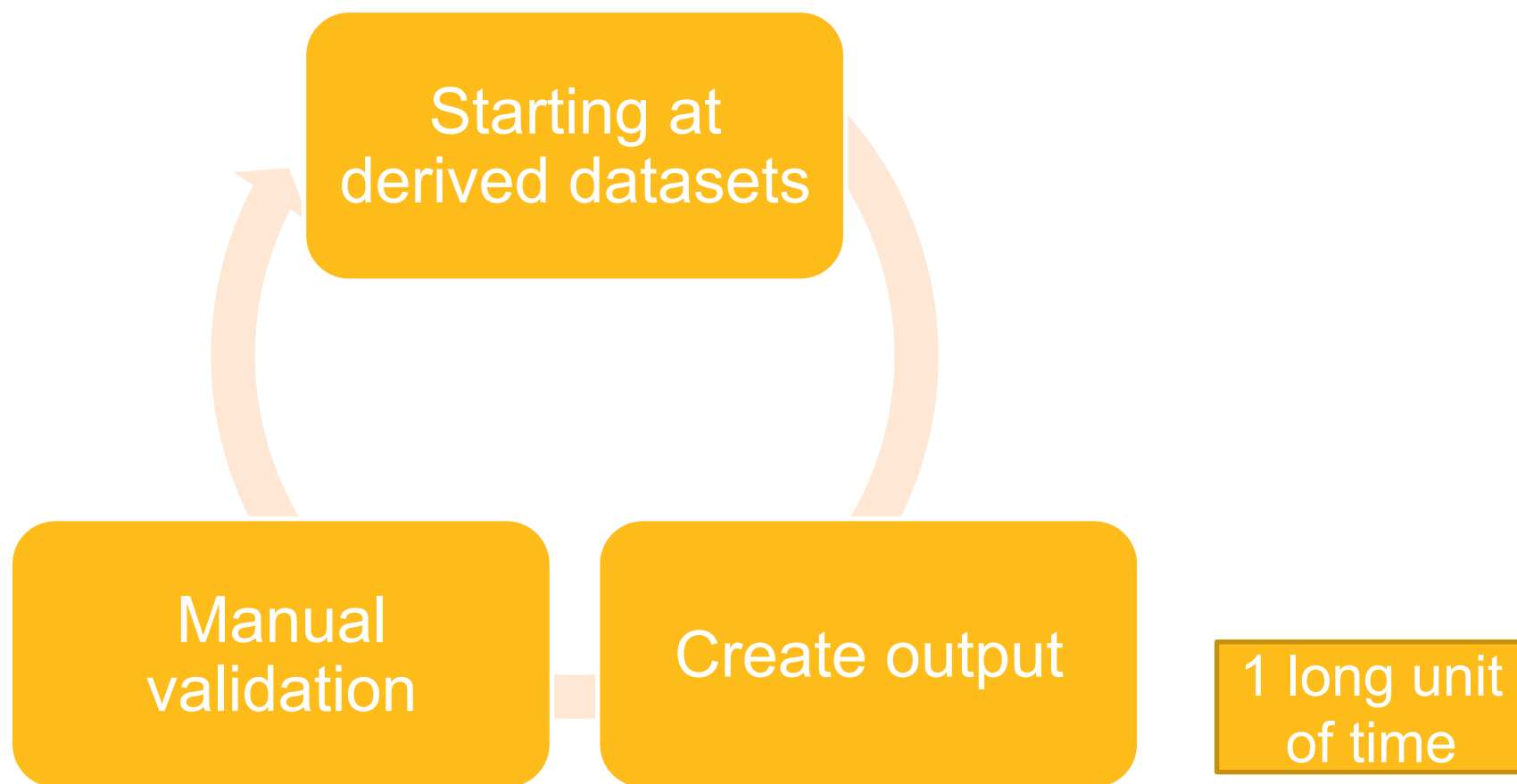
SYS validation process

What is SYS validation tool

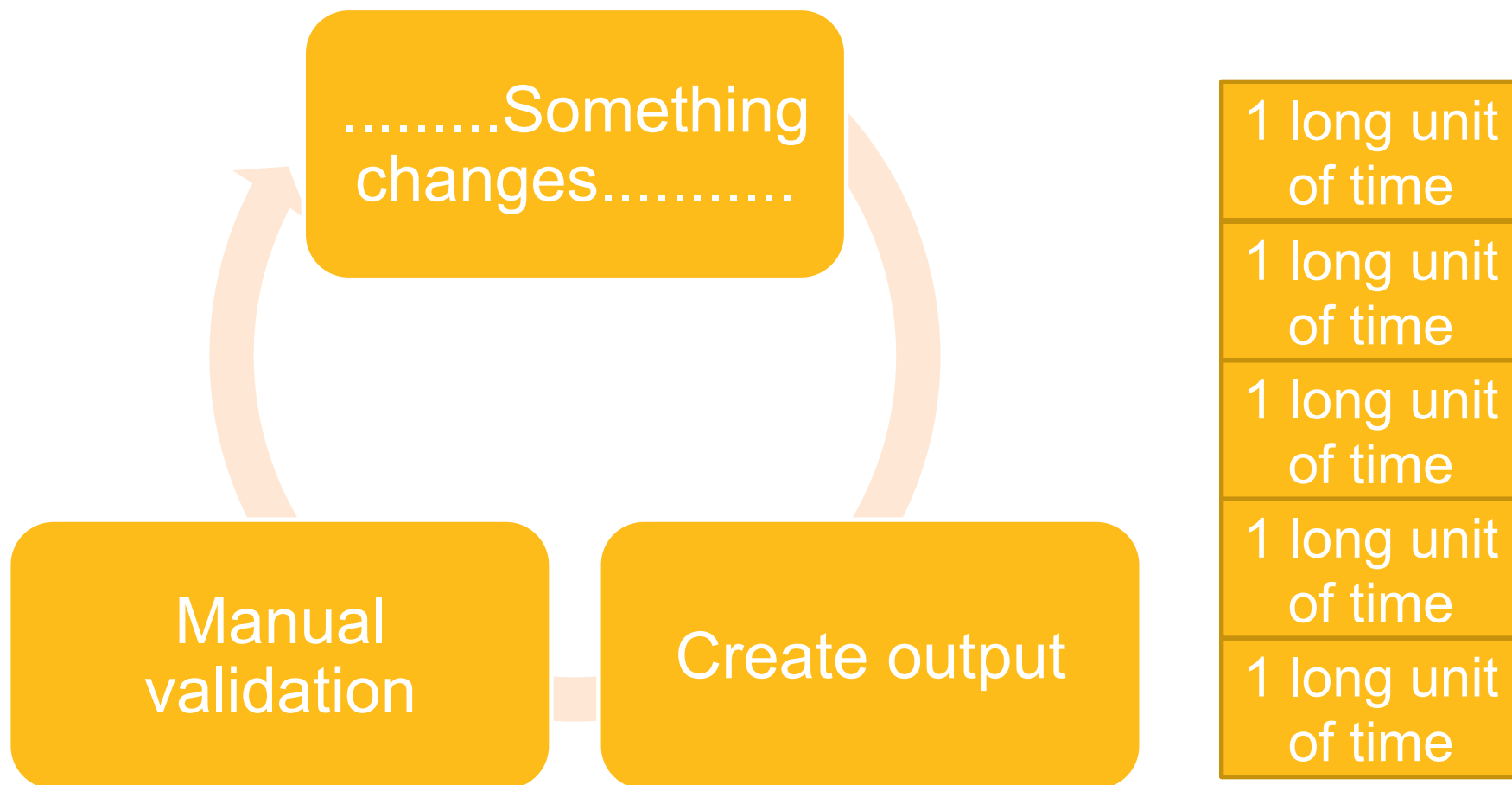
%intext_parse macro

Challenges and benefits

Background – Normal validation process



Background – Normal validation process



Background

From this



Source: Google images

Background

To this...



Source: Google images

Background

What is SYS validation tool

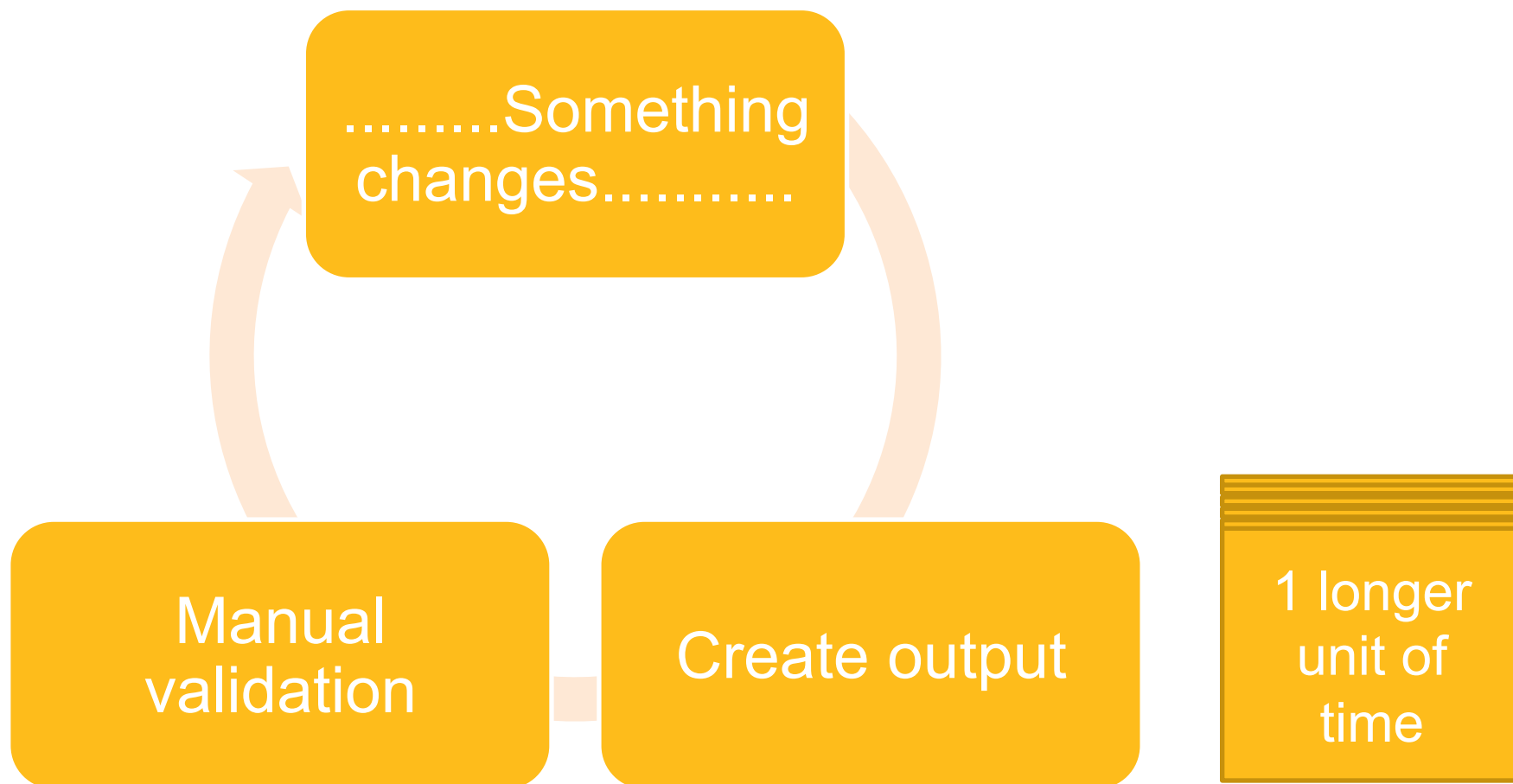
- Save Your Sight validation tool.
- Automates the validation process.
- One-time initial time investment.
 - After that.. Sit back, relax, get a drink, and **push that button**.



Source: Google images

Disclaimer: Picture may not exactly resemble your keyboard button.

New validation process



Type of Outputs

Landscape

Demographics by treatment (Full analysis set)

Demographic Variable	Group1 N=xx	Group2 N=xx	All Patients N=xx
Age (years)			
Q	xx	xx	xx
Mean	xx.x	xx.x	xx.x
SD	x.xx	x.xx	x.xx
Median	xx.x	xx.x	xx.x
Minimum	xx	xx	xx
Maximum	xx	xx	xx
Age category (years) -n (%)			
<65	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
>=65	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Missing	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Sex -n (%)			
Female	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Male	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Missing	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)

Source: Imaginary data representing a Landscape output.

Type of Outputs

Portrait

Demographics by treatment group (Full analysis set)

	Group1 N=xx	Group2 N=xx	All patients N=xx
Age (years)			
n	xx	xx	xx
Mean	xx.x	xx.x	xx.x
SD	xx.xx	xx.xx	xx.xx
Median	xx.x	xx.x	xx.x
Min	xx	xx	xx
Max	xx	xx	xx
Age category (years)			
< 65	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
>= 65	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Missing	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Sex			
Male	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Female	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)
Missing	xx (xx.x%)	xx (xx.x%)	xx (xx.x%)

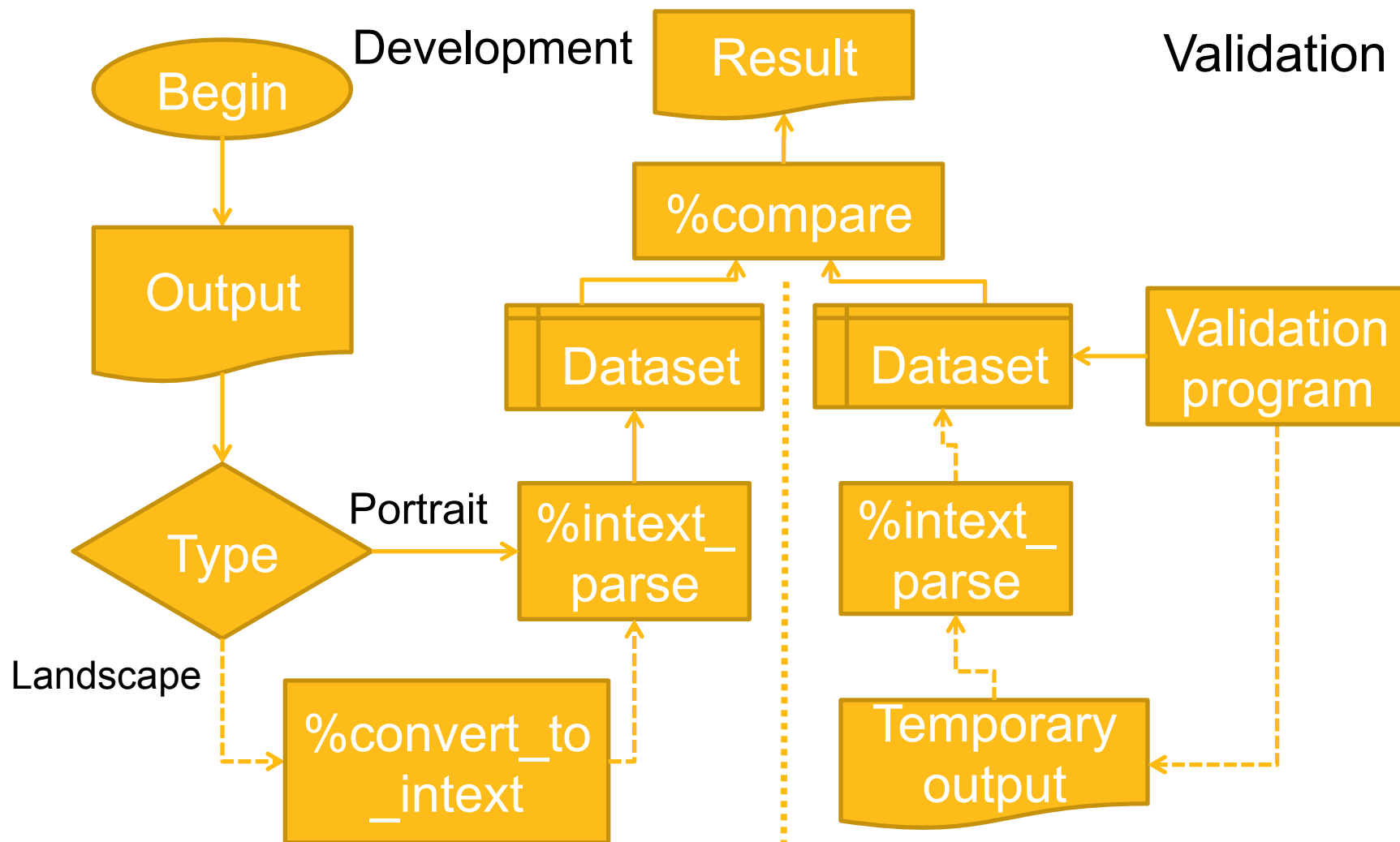
Source: Imaginary data representing a portrait output.

Overview of tool

Components

- %intext_parse
 - Convert Portrait rtf to sas7bdat
- %convert_to_intext
 - Convert Landscape rtf to Portrait rtf
- %compare
 - Compare datasets and generate output

SYS validation Process



%intext_parse macro

Example output

Table 1-1							
Analysis populations by initial treatment (Full Analysis Set)							
Analysis population	<= 320 mg N=12	500 mg N=4	900 mg N=9	1800 mg N=24	2100 mg N=2	3000 mg N=2	All Subjects N=53
Full Analysis Set	12(100.0%)	4(100.0%)	9(100.0%)	24(100.0%)	2(100.0%)	2(100.0%)	53(100.0%)
Dose Determining Set	12(100.0%)	3(75.0%)	3(33.3%)	21(87.5%)	2(100.0%)	1(50.0%)	42(79.2%)
Safety Set	12(100.0%)	4(100.0%)	9(100.0%)	24(100.0%)	2(100.0%)	2(100.0%)	53(100.0%)

Source: Table X.Y

- Full analysis set includes all patients who received at least one dose of study medication.
- Safety set includes all patients who received at least one dose of study medication, and had at least one post-baseline safety assessment.
- Dose-determining set includes all subjects from the safety set of dose escalation phase who received all three doses of weekly study drug and were observed for at least 7 days following their 3rd dose and/or who experienced a dose-limiting toxicity.

Path of program

Production

2

3

4

5

6

Source: Imaginary data generated in a RTF file.

%intext_parse macro

Example Call and Arguments passed

■ Example Call

```
%intext_parse(fpath      = Path of the file
               ,dsetOut   = RTFTable
               ,ignoreCase = N
               ,ignoreWS   = N
               ,compBL     = Y
               ,combSects  = Y
               ,debug      = N);
```

■ Arguments

- FPATH = Full path of intext rtf file
- DSETOUT = Name of output dataset
- IGNORECASE = Y or N flag to convert to lower case for a case insensitive match
- IGNOREWS = Y or N flag to compress all spaces
- COMPBL = Y or N flag to compress out double spaces and remove leading and lagging whitespace
- COMBSECTS = Y or N flag to specify whether sections should be combined. If set to N, a dataset will be created for each section.
- DEBUG = Y or N flag for basic debugging

%intext_parse macro

Flow of the macro

1) Parameter checking:

%superq

Fileexist

%bquote

%sysevalf

Macro will check whether
all the parameters are passed correctly.

Verify

%sysfunc

%qupcase

%intext_parse macro

Flow of the macro

2) Reading the rtf data:

Macro will Collect all the information from rtf file and put them into a dataset.

Each cell becomes an observation.

Page number, Row number, Column number and Header flag are generated.

Prxparse

CatX

Prxsubstr

Missing

Prxchange

Tranwrdr

Substr

EO

%intext_parse macro

Flow of the macro

3) Separating Table content from header and footer:

Macro searches for multiple cells per row.
If it finds multiple cells then it is table content.

First.variable

Lag

Last.variable

EO

%intext_parse macro

Flow of the macro

4) Defining column header:

Column position metadata is stored in the RTF for the first cell in a row.
Extract this information and match it to the corresponding cell.
First row will become the header information
for the following rows in the tables.

Prxnext

Prxsubstr

EO

%intext_parse macro

Flow of the macro

5) By Variable grouping:

This section extracts category values from the header sections. The byval category is assumed to be the longest common substring.

Array

Hash

EO

%intext_parse macro

Flow of the macro

6) Transposing data and removing blank rows:

Transposing the data followed by removing the blank rows if exist, to get the final dataset for comparison purpose.

Proc Transpose

Missing

Cats

EO

%intext_parse macro



Output dataset

ROW	Analysis population	<= 320 mg N=12	500 mg N=4	900 mg N=9	1800 mg N=24	2100 mg N=2	3000 mg N=2	All Subjects N=
1	Full Analysis Set	12(100.0%)	4(100.0%)	9(100.0%)	24(100.0%)	2(100.0%)	2(100.0%)	53(100.0%)
2	Dose Determining Set	12(100.0%)	3(75.0%)	3(33.3%)	21(87.5%)	2(100.0%)	1(50.0%)	42(79.2%)
3	Safety Set	12(100.0%)	4(100.0%)	9(100.0%)	24(100.0%)	2(100.0%)	2(100.0%)	53(100.0%)

Source: Imaginary data generated in a SAS7BDAT file.

%intext_parse macro

Properties of output dataset

- The output dataset includes the following properties to assist in validation:
 - *Row variable* – A unique numeric variable that represents the row number
 - *Byval variable* – Character variable used to store by variable values. Will be empty if no by-variables are used in the analysis.
 - *Col1-coln variables* – Variables on the original table will be read as col1 through to coln.
 - *Headers as labels* – Headers on the report are stored as variable labels

Challenges and Benefits



■ Challenges:

- Initial investment of time required
- Nested headers

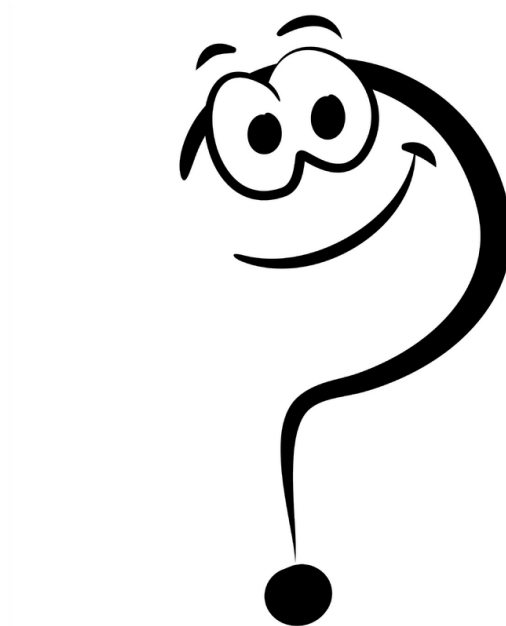
■ Benefits:

- Improved quality – 100% validation performed
- Extremely fast validation – no monotonous checking
- Reproducibility – re-running program highlights issues observed

Acknowledgements



- Karma Tarap (Author of the Macro)
- N Jayapandian (Co-Author for Presentation)
- Suresh Gunturu (Support)
- Programming team : Novartis (Oncology)
- PhUSE Committee



Source: Google images