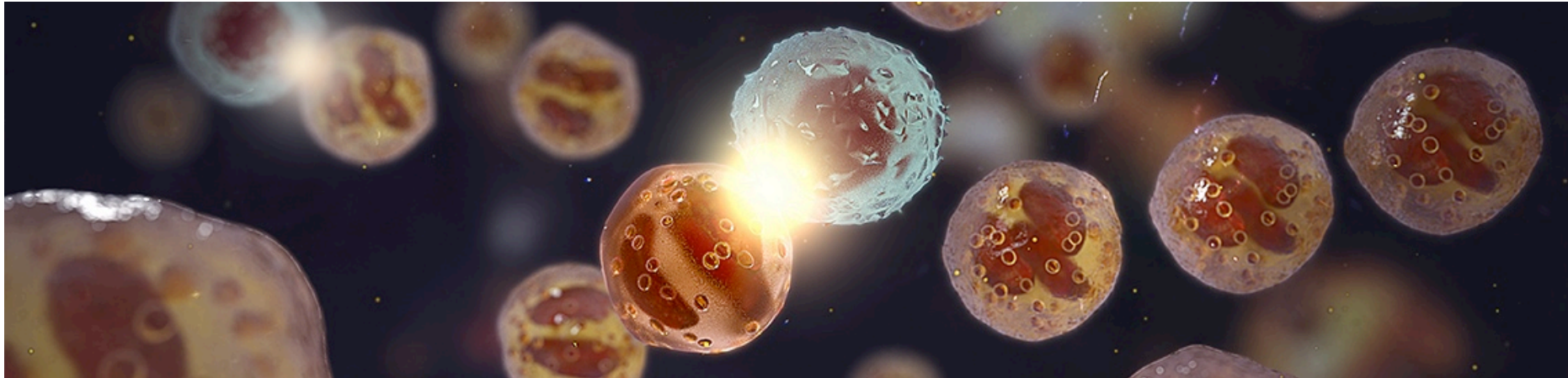


SDTM/ADaM Datasets using R

Prasanna Murugesan

April 2019



SDTM/ADaM Datasets using R

The rise of R

- Open source statistical computing software
- Widely used in exploratory analysis and visualizations
- Most preferred among academic institutions
- Popularity increased with big data and data analytics
- Used in exploratory analysis by statisticians
- Used for analyzing non clinical data (Ex – Post marketing data / HEOR)



SDTM/ADaM Datasets using R

R for data transformation

- How can R be used by Clinical trial data analysis?
- Can R be used for data transformation/manipulation?
- Can R be used for creating CDSIC - SDTM/ADaM data sets?
- Can R perform the functions done by SAS?



SDTM/ADaM Datasets using R

R for data transformation

- R packages for Data Transformation:
 - Base : Base R functions
 - Dplyr : designed for data transformation
 - Lubridate : Date functions
 - Haven : Loads and exports SAS files
 - Foreign : for csv file creation and SAS import file creation



SDTM/ADaM Datasets using R

R- Dplyr Syntax

R syntax	SAS equivalent
select	keep/drop
mutate	data step derivations
group_by	by statement
filter	if/where statement
arrange	proc sort
summarise	statistical procedures (means/tabulate)
spread/gather	transpose
left/right/inner/full joins	merge
bind_cols/bind_rows	set by rows/columns



SDTM/ADaM Datasets using R

R- Dplyr Syntax

- Libname setup
- Subject level derivation
- Merging parent and supplemental dataset
- Baseline flag/change from baseline derivation



SDTM/ADaM Datasets using R

R- Dplyr Syntax

Libname setup equivalent in R- Sample code:

```
sdtm <- "//product/study/analysis/sp/data/final"
```

```
out  <- "C:/"
```

Read files:

```
dm    <- read_sas(file.path(sdtm,"dm.sas7bdat"))
```

```
ds    <- read_sas(file.path(sdtm,"ds.sas7bdat"))
```

```
sv    <- read_sas(file.path(sdtm,"sv.sas7bdat"))
```

```
suppsv <- read_sas(file.path(sdtm,"suppsv.sas7bdat"))
```



SDTM/ADaM Datasets using R

R- Dplyr Syntax

Subject level derivation -Sample code:

```
adsl <- dm %>%      # read %>% as “and then”  
  select(studyid, subjid, age, sex, height, weight, race, scrfl) %>%  
  mutate(bmi = (weight*703)/height^2 ) %>%  
  filter(scrfl == “Y”) %>%  
  select(-scrfl) %>%  
  arrange(studyid, subjid)
```



SDTM/ADaM Datasets using R

R- Dplyr Syntax

Merging parent and supplemental dataset in R- Sample code:

```
suppds_ <- suppds %>%  
  mutate(idvarval_ = as.numeric(idvarval)) %>%  
  select(usubjid,idvarval_,qnam,qval) %>%  
  spread(.,qnam,qval)  
ds_all <- left_join(ds, suppds_, by = c("usubjid"="usubjid",  
  "dsseq"="idvarval_"))
```



SDTM/ADaM Datasets using R

R- Dplyr Syntax

Baseline flag/change from baseline derivation in R- Sample code:

```
advs <- advs_ %>%  
  group_by(subjid,paramcd) %>%  
  arrange(subjid,paramcd) %>%  
  mutate ( base = aval[visitnum==1],  
           ablfl = ifelse(visitnum == 1,"y",na),  
           chg = ifelse ( !is.na(aval) & !is.na(base),aval-base,na),  
           pchg= ifelse ( !is.na(aval) & !is.na(base),((aval-  
           base)/base)*100,na)  
  )
```

subjid	paramcd	visitnum	aval	base	chg	pchg
x1	HR	-1	72	70	-2	-2.9
x1	HR	1	70	70	0	0
x1	HR	2	75	70	-5	-7.1
x1	HR	3	78	70	-8	-11
x1	HR	4	76	70	-6	-8.6



SDTM/ADaM Datasets using R

R for data transformation

Benefits:

- An independent open source tool to validate SAS datasets.
- More statistical functionalities available in R.
- Datasets viewed in R-Shiny make it more user friendly.



SDTM/ADaM Datasets using R

R for data transformation

Challenges:

- Packages are written by users and constantly changes.
- Packages may not have gone through SDLC.
- SAS datasets names/vars read in R are case sensitive and R syntaxes are case sensitive.
- Logs are not detail oriented as SAS.



SDTM/ADaM Datasets using R

Questions

