SDTM/ADaM Datasets using R

Prasanna Murugesan

June 2018
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
R in Pharmaceutical Industry

- Used in exploratory analysis by statisticians
- Used for analyzing non-clinical data (Ex – Post marketing data / HEOR)

SDTM/ADaM Datasets using R

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
  – Sas7bdat : Loads and exports SAS files
## SDTM/ADaM Datasets using R

### R- Dplyr Syntax

<table>
<thead>
<tr>
<th>R syntax</th>
<th>SAS equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>select</td>
<td>keep/drop</td>
</tr>
<tr>
<td>mutate</td>
<td>data step derivations</td>
</tr>
<tr>
<td>group_by</td>
<td>by statement</td>
</tr>
<tr>
<td>filter</td>
<td>if/where statement</td>
</tr>
<tr>
<td>arrange</td>
<td>proc sort</td>
</tr>
<tr>
<td>summarise</td>
<td>statistical procedures (means/tabulate)</td>
</tr>
<tr>
<td>spread/gather</td>
<td>transpose</td>
</tr>
<tr>
<td>left/right/inner/full joins</td>
<td>merge</td>
</tr>
<tr>
<td>bind_cols/bind_rows</td>
<td>set by rows/columns</td>
</tr>
</tbody>
</table>
SDTM/ADaM Datasets using R

R- Dplyr Syntax

• Libname setup
• Subject level derivation
• Merging parent and supplemental dataset
• Baseline flag/change from baseline derivation
Libname setup equivalent in R - Sample code:

```r
sdtm <- "/product/study/analysis/sp/data/final"
out <- "C:\"
```

Read files:

```r
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:

```r
adsl <- dm %>>

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
Merging parent and supplemental dataset in R - Sample code:

```r
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_"))
```
R- Dplyr Syntax

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

```r
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)
)
```

### SDTM/ADaM Datasets using R

<table>
<thead>
<tr>
<th>subj</th>
<th>paramcd</th>
<th>visitnum</th>
<th>aval</th>
<th>base</th>
<th>chg</th>
<th>pchg</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>HR</td>
<td>-1</td>
<td>72</td>
<td>70</td>
<td>-2</td>
<td>-2.9</td>
</tr>
<tr>
<td>x1</td>
<td>HR</td>
<td>1</td>
<td>70</td>
<td>70</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>x1</td>
<td>HR</td>
<td>2</td>
<td>75</td>
<td>70</td>
<td>-5</td>
<td>-7.1</td>
</tr>
<tr>
<td>x1</td>
<td>HR</td>
<td>3</td>
<td>78</td>
<td>70</td>
<td>-8</td>
<td>-11</td>
</tr>
<tr>
<td>x1</td>
<td>HR</td>
<td>4</td>
<td>76</td>
<td>70</td>
<td>-6</td>
<td>-8.6</td>
</tr>
</tbody>
</table>
SDTM/ADaM Datasets using R

R for data transformation

Benefits:

• An independent tool to validate SAS datasets.
• R being an open source tool can reduce license cost.
• 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