

# Mr. (R)ight – Why R Should Be Used for Validating TLFs in Clinical Trials?

Monika Ludwinska  
PHUSE EU Connect 2023  
Birmingham, UK  
5–8 November 2023

# Agenda

- ✓ Introduction
- ✓ Quality Control Procedure
- ✓ Principles of Independent Validation for Quality Control Procedure
- ✓ Practical examples
- ✓ R traps
- ✓ Summary



# Introduction

- ✓ Isn't SAS enough?
- ✓ Usage of R in Quality Control Procedure
- ✓ Programmers' future

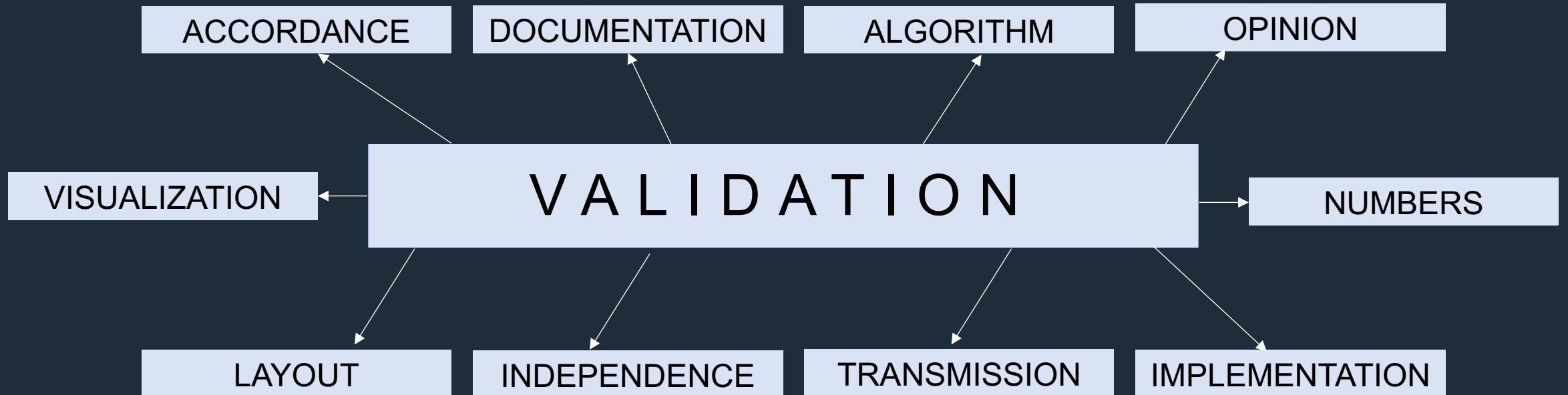


# Quality Control (QC) Procedure

- ✓ Tables, Listings, Figure (TLF) verification
- ✓ Multiple steps process
- ✓ Demands for more effort and accuracy from the Programmer



## Principles of Independent Validation for Quality Control Procedure



# R as a Validation Tool

## R packages

- ✓ Haven – loading SAS data formats (sas7bdat)
- ✓ Dplyr, tidyr – data manipulations
- ✓ Diffdf – comparing results
- ✓ Logrx – generating log file

# Tables

Population: Enrolled

Table 1  
Summary of Baseline Characteristics

| Characteristic                       |         | Treatment X<br>(N=9) | Placebo<br>(N=8) | Total<br>(N=17) |
|--------------------------------------|---------|----------------------|------------------|-----------------|
| BMI (kg/m <sup>2</sup> ) at Baseline | n       | 9                    | 8                | 17              |
|                                      | Mean    | 24.534               | 26.461           | 25.441          |
|                                      | SD      | 5.1625               | 5.6305           | 5.3083          |
|                                      | Median  | 24.570               | 27.020           | 25.070          |
|                                      | Min.    | 16.67                | 19.11            | 16.67           |
|                                      | Max.    | 35.51                | 35.38            | 35.51           |
| Age group (%)                        | n       | 8                    | 7                | 15              |
|                                      | <18     | 2 (25%)              | 1 (14%)          | 3 (20%)         |
|                                      | >=18-64 | 3 (38%)              | 4 (57%)          | 7 (47%)         |
|                                      | >=65-84 | 3 (38%)              | 2 (29%)          | 5 (33%)         |
|                                      | >=85    | 0                    | 0                | 0               |

Note: This is footnote 1.

Note: This is footnote 2.

# Tables

```
# Adding total group, filtering, bigN, label
tab <- adsl %>%
  mutate(TRT01P = "Total",
         TRT01PN = 9) %>%
  bind_rows(adsl) %>%
  filter(ENRLFL == "Y") %>%
  #BigN
  group_by(TRT01P) %>%
  mutate(bign = n(),
         label = paste0(TRT01P, "~(N=", bign,

# First part of the table
qc1 <- tab %>%
  group_by(TRT01PN) %>%
  summarise(n = as.character(n()),
            Mean = as.character(format(round(
            SD = as.character(format(round(
            Median = as.character(format(round(
            Min. = as.character(min(BMIBL)),
            Max. = as.character(max(BMIBL))) %>%
  pivot_longer(cols = -TRT01PN, names_to = "stats") %>%
  pivot_wider(values_from = value, names_from = TRT01PN, names_prefix = "col") %>%
  mutate(char = "BMI (kg/m^2) at Baseline")
```

| TRT01PN | n  | Mean   | SD     | Median | Min.  | Max.  |
|---------|----|--------|--------|--------|-------|-------|
| 1       | 9  | 24.534 | 5.1625 | 24.570 | 16.67 | 35.51 |
| 2       | 8  | 26.461 | 5.6305 | 27.020 | 19.11 | 35.38 |
| 9       | 17 | 25.441 | 5.3083 | 25.070 | 16.67 | 35.51 |

initial dataset

| TRT01PN | stats  | value  |
|---------|--------|--------|
| 1       | n      | 9      |
| 1       | Mean   | 24.534 |
| 1       | SD     | 5.1625 |
| 1       | Median | 24.570 |
| 1       | Min.   | 16.67  |
| 1       | Max.   | 35.51  |
| ...     | ...    | ...    |

pivot\_longer result

| stats  | col1   | col2   | col9   |
|--------|--------|--------|--------|
| n      | 9      | 8      | 17     |
| Mean   | 24.534 | 26.461 | 25.441 |
| SD     | 5.1625 | 5.6305 | 5.3083 |
| Median | 24.570 | 27.020 | 25.070 |
| Min.   | 16.67  | 19.11  | 16.67  |
| Max.   | 35.51  | 35.38  | 35.51  |

pivot\_wider result

# Tables

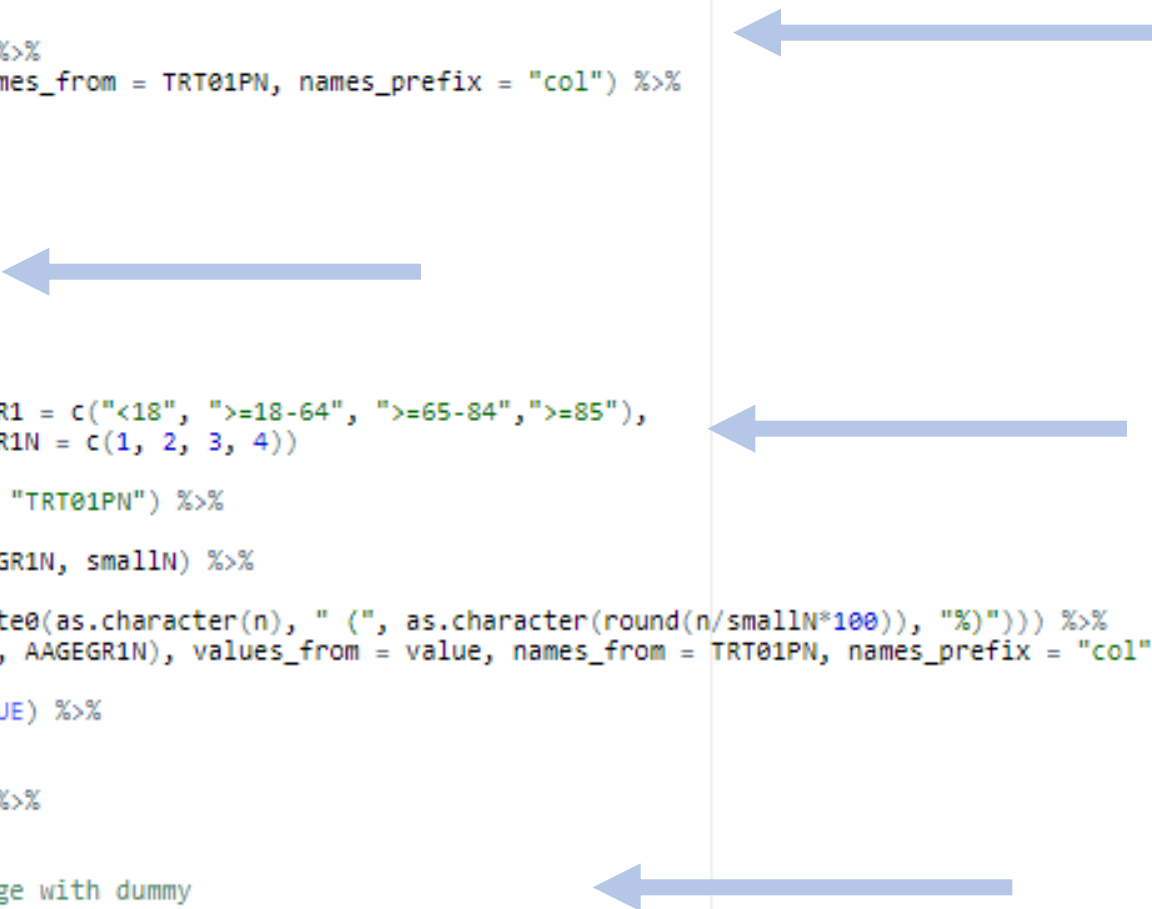
```
# Second part of the table
qc2_1 <- tab %>%
  filter(!is.na(AAGEGR1N)) %>%
  group_by(TRT01PN) %>%
  summarise(n=as.character(n())) %>%
  pivot_wider(values_from = n, names_from = TRT01PN, names_prefix = "col") %>%
  mutate(char = "Age group (%)",
         stats = 'n')

#Denominator
smallN <- tab %>%
  filter(!is.na(AAGEGR1N)) %>%
  group_by(TRT01PN) %>%
  summarise(smallN=n())

#Dummy dataset
dummy_agegr1 <- data.frame(AAGEGR1 = c("<18", ">=18-64", ">=65-84", ">=85"),
                          AAGEGR1N = c(1, 2, 3, 4))

qc2_2 <- merge(tab, smallN, by = "TRT01PN") %>%
  filter(!is.na(AAGEGR1N)) %>%
  group_by(TRT01PN, AAGEGR1, AAGEGR1N, smallN) %>%
  summarise(n = n()) %>%
  mutate(value = as.character(paste0(as.character(n), " (", as.character(round(n/smallN*100)), "%)"))) %>%
  pivot_wider(id_cols = c(AAGEGR1, AAGEGR1N), values_from = value, names_from = TRT01PN, names_prefix = "col") %>%
  # Merge with dummy
  merge(dummy_agegr1, all.y = TRUE) %>%
  #Renaming
  rename(stats = AAGEGR1) %>%
  mutate(char = "Age group (%)") %>%
  select(-AAGEGR1N)

# Replacing NA with "0" after merge with dummy
qc2_2[is.na(qc2_2)] <- "0"
```



# Tables

```
# Setting dynamic labels - into vector
labels <- tab %>%
  distinct(label, TRT01PN) %>%
  arrange(TRT01PN) %>%
  pull(label)

attr(qc$col1, "label") <- labels[1]
attr(qc$col2, "label") <- labels[2]
attr(qc$col9, "label") <- labels[3]

prod <- as.data.frame(lapply(prod, function(y) gsub("\\s+", "", y)))
qc <- as.data.frame(lapply(qc, function(y) gsub("\\s+", "", y)))

difffdf(base = prod,
        compare = qc)
```

```
library(stringr)
```

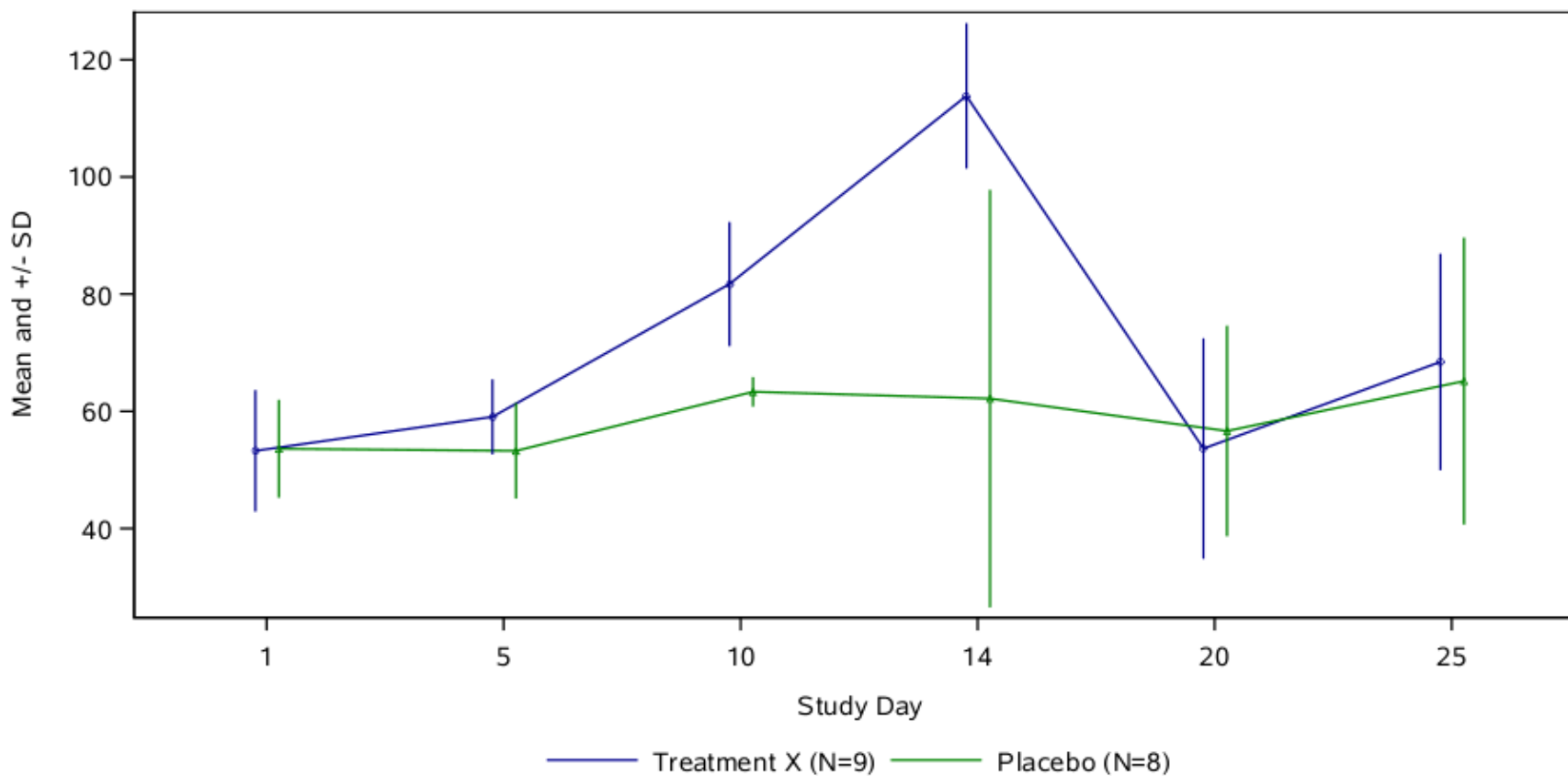
```
lst <- data.frame(text = str_trim(readLines("~/Desktop/Phuse/EU 2023/SAS/t
  filter(row_number() == 3 | row_number() == n() | row_number() == as.numeric(n())-1))
```

| text                                |
|-------------------------------------|
| Summary of Baseline Characteristics |
| Note: This is footnote 1.           |
| Note: This is footnote 2.           |

# Figures

Population: Safety

Figure 1  
Mean (SD) Creatinine Values Over Time by Treatment Group



# Figures

```
bign <- adsl %>%
  filter(SAFFL == "Y") %>%
  #BigN
  group_by(TRT01A) %>%
  summarise(bign = n()) %>%
  mutate(trtlabel = paste0(TRT01A, " (N=", bign, ")"))

fig <- merge(adlb, bign, by = "TRT01A", all.x = TRUE) %>%
  filter(SAFFL == "Y")

qc <- fig %>%
  group_by(TRT01AN, TRT01A, trtlabel, PARAM, ADY, bign) %>%
  summarise(mean = mean(AVAL),
            sd = sd(AVAL)) %>%
  mutate(upper = ifelse(!is.na(mean) & !is.na(sd), mean + sd, NA),
         lower = ifelse(!is.na(mean) & !is.na(sd), mean - sd, NA)) %>%
  mutate(mean = round(mean,8),
         upper = round(upper,8),
         lower = round(lower,8)) %>%
  select(-sd, -bign)

diffdf(base = prod,
       compare = qc,
       file = "qc_figure_1.lst")
```

# Figures

```
# Axes ranges
y <- qc %>%
  group_by(PARAM, TRT01A) %>%
  summarise(ymin = min(lowersd),
            ymax = max(upperd))

# Specific colors for eg. Treatment groups
library(imager)

colors_ <- load.image("figure_1.png")

colors1 <- data.frame(
  R = as.vector(colors_[,,1]),
  G = as.vector(colors_[,,2]),
  B = as.vector(colors_[,,3]))

colors <- data.frame(RGB = rgb(colors1)) %>%
  group_by(RGB) %>%
  summarise(n = n()) %>%
  arrange(desc(n)) %>%
  head(4)
```

| PARAM      | TRT01A      | ymin     | ymax      |
|------------|-------------|----------|-----------|
| Creatinine | Placebo     | 26.53478 | 97.80922  |
| Creatinine | Treatment X | 34.81289 | 126.23201 |

| RGB     | n      | color |
|---------|--------|-------|
| #FFFFFF | 564684 |       |
| #000000 | 4842   |       |
| #008000 | 775    |       |
| #00008B | 579    |       |

# Listings

Population: Enrolled

## Listing 1 Listing of Demographic Characteristics

Site Id.: XYZ

| Treatment | Unique Subject Identifier | Date of Birth | Age<br>(YEARS) [1] | Sex | Height (cm) | Weight (kg) |
|-----------|---------------------------|---------------|--------------------|-----|-------------|-------------|
| Active    | 001                       | 1961          | 30                 | F   | 160         | 51.5        |

```
qc <- adsl %>%
  filter(SAFFL == "Y") %>%
  mutate(group = paste0("Site Id.: ", SITEID)) %>%
  select(TRT01P, USUBJID, BRTHDTC, AAGE, SEX, HEIGHTBL, WEIGHTBL, group, TRT01PN, SITEID) %>%
  arrange(SITEID, TRT01PN)
```

|         |     |      |    |   |     |      |
|---------|-----|------|----|---|-----|------|
|         | 011 | 1959 | 62 | F | 159 | 89.2 |
|         | 012 | 1964 | 56 | F | 163 | 56.5 |
|         | 013 | 1971 | 49 | M | 173 | 89.5 |
|         | 014 | 1969 | 51 | M | 169 | 75.8 |
| Placebo | 003 | 1956 | 70 | F | 188 | 75.5 |
|         | 005 | 1978 | 42 | M | 152 | 45.6 |

# R traps

## ✓ SAS vs R rounding

| Number | SAS | R  |
|--------|-----|----|
| 1.5    | 2   | 2  |
| 2.5    | 3   | 2  |
| 3.5    | 4   | 4  |
| 4.5    | 5   | 4  |
| 12.5   | 13  | 12 |

## ✓ Attributes comparison

There are columns in BASE and COMPARE with differing attributes !!  
All rows are shown in table below

```
=====
VARIABLE  ATTR_NAME  VALUES.BASE  VALUES.COMP
-----
TRT01P    label      Treatment      XXX
-----
```

# Summary

- ✓ R as a Quality Control engine
- ✓ Tidyverse and Pharmaverse packages
- ✓ R traps
- ✓ Upskilling



# THANK YOU

Monika Ludwinska

Intego Group LLC

Grzybowska 78

00-844 Warsaw, Poland

Work Phone: +48 222 500 032 ext. 2583

Email: [monika.ludwinska@intego-group.com](mailto:monika.ludwinska@intego-group.com)

Web: [www.intego-group.com](http://www.intego-group.com)