R语言生成CDISC递交数据集的优势及挑战

北京方胜远临医药科技有限公司
ClinFunction Co.,Ltd

方胜远临
ClinFunction

CONTENTS

PART
01

R语言背景介绍

Background Introduction of R Language

背景介绍

R语言的起源和发展

R 语言最初由罗斯·伊哈卡（Ross Ihaka）和罗伯特·根特尔曼（Robert Gentleman）在新西兰奥克兰大学开发，于1993年首次发布。它是一种基于S语言的开源编程语言，旨在提供灵活的统计计算和数据分析工具。随着学术界和业界对数据分析需求的增长，R 在全球得到了广泛的应用和支持，发展成为强大的统计编程语言。

R语言的特点和应用

R 是一门以数据分析和统计为核心的编程语言，拥有丰富的包生态系统，涵盖统计建模、机器学习、数据可视化等众多领域。它的包管理系统 CRAN 拥有数千个开源扩展包，使用户能够快速解决复杂的统计和数据处理问题。R 的可视化功能也是其一大优势，可以轻松创建精美且复杂的图表，这让它在科研、金融、生物统计等领域广受欢迎。

R语言在现代数据科学中的地位

随着数据科学的兴起，R 语言已成为数据分析师和统计学家不可或缺的工具之一。它不仅在学术研究中发挥着重要作用，而且在商业应用中被广泛用于数据分析、预测建模和报告生成。RStudio 等集成开发环境（IDE）的出现，使得 R 的使用更加便捷和高效，进一步巩固了它在数据科学和统计计算中的重要地位。

与SAS比较

R的优势

1、开源免费：完全开源的，用户可以免费使用并根据需要定制代码

2、社区活跃：R 拥有一个庞大的全球用户社区，用户可以通过论坛和社交媒体获取丰富的教程、包更新和技术支持

3、灵活性强：R 提供了灵活的数据分析和建模工具，可以通过自定义脚本处理复杂的数据集，满足各类科研和商业需求

SAS的优势

1、有商业支持：用户可以获得官方的技术支持、培训和咨询服务

2、稳定性高：SAS 在数据处理和分析的稳定性方面表现出色，尤其适合大型企业和政府机构处理重要的业务数据

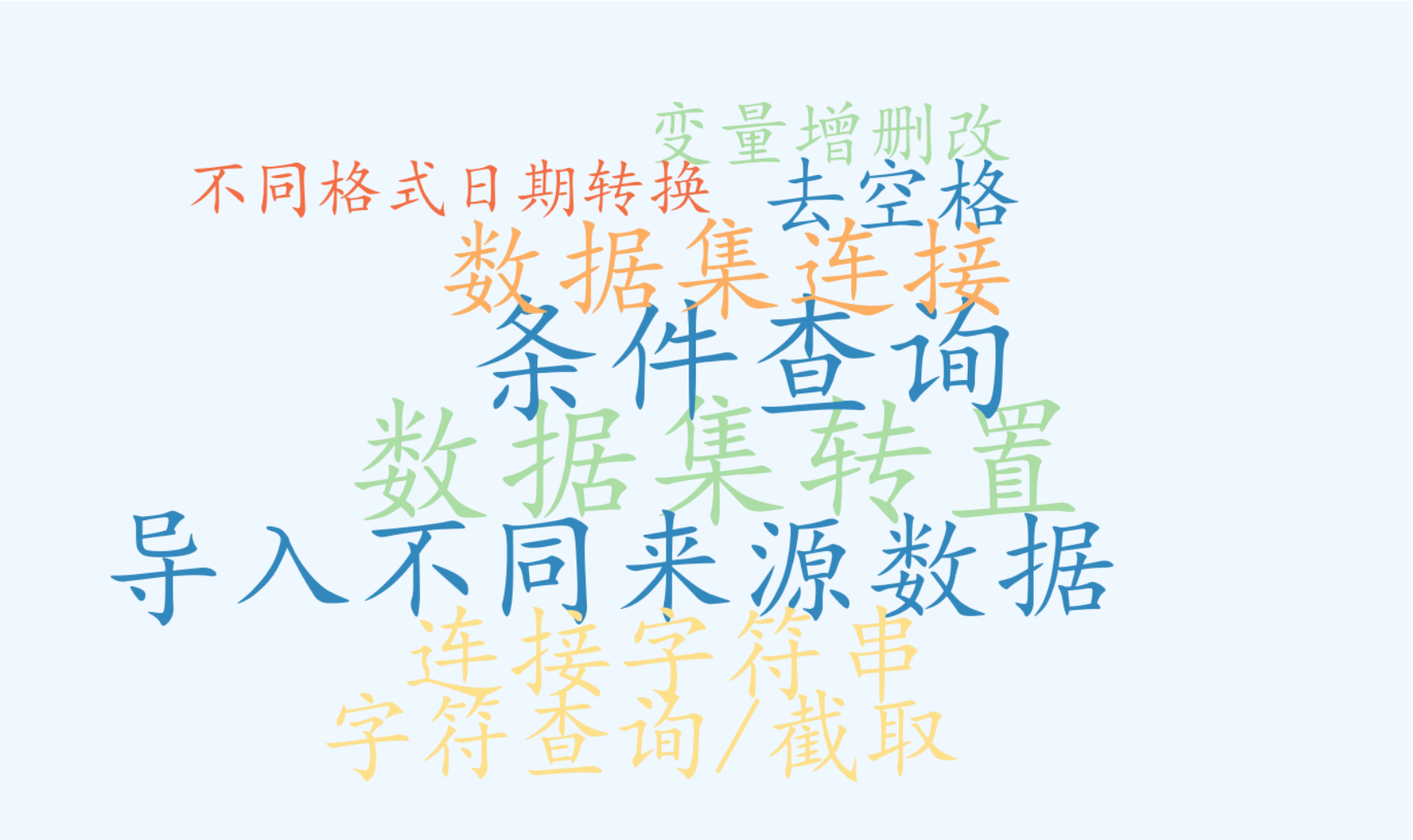
3、在统计方面有很大优势：SAS 拥有强大的统计分析功能和标准化的过程，许多制药公司目前使用。

PART
02

R生成SDTM数据集

Generating SDTM Datasets with R

SDTM数据标准化编程常用技巧



常见数据处理及转换 – 去空格

SAS

```
data b;
   set a;
   Name_trimmed = strip(Name); /* 去除 Name 字段前后的空格 */
run;

proc print data=b;
run;
```

R

```
# 假设有原始数据框 a
a <- data.frame(
   ID = 1:3,
   Name = c("  John Doe  ", " Jane Smith ", "  Alice Johnson  ")
)

# 去除 Name 列前后的空格
    a$Name <- trimws(a$Name)

# 查看结果
    print(a)
```

常见数据处理及转换 – 连接字符

SAS

```
data b;
   set a;

   /* 使用 CATX() 函数连接字符串，指定分隔符 */
   FullName = catx(' ', FirstName, LastName); /* 使用空格作为分隔符 */

   /* 使用 CATX() 函数连接字符串，指定逗号和空格为分隔符 */
   FullNameComma = catx(', ', FirstName, LastName); /* 使用逗号和空格作为分隔符 */

   /* 直接连接，不加分隔符 */
   FullNameNoSpace = catt(FirstName, LastName);
run;

proc print data=b;
run;
```

R

```
# 示例数据框 a
a <- data.frame(
  ID = 1:3,
  FirstName = c("John", "Jane", "Alice"),
  LastName = c("Doe", "Smith", "Johnson")
)

# 使用 paste() 连接 FirstName 和 LastName
a$FullName <- paste(a$FirstName, a$LastName)

# 使用 paste() 连接 FirstName 和 LastName，且以逗号和空格为分隔符
a$FullNameComma <- paste(a$FirstName, a$LastName, sep = ", ")

# 使用 paste0() 直接连接（没有空格）
a$FullNameNoSpace <- paste0(a$FirstName, a$LastName)

# 查看结果
print(a)
```

常见数据处理及转换 – 连接字符

```
>
> # 查看结果
> print(a)
  ID FirstName LastName      FullName  FullNameComma FullNameNoSpace
1  1      John      Doe      John Doe      John, Doe          JohnDoe
2  2      Jane    Smith    Jane Smith    Jane, Smith        JaneSmith
3  3     Alice  Johnson Alice Johnson Alice, Johnson     AliceJohnson
>
```

常见数据处理及转换 – 字符串拆分

SAS

```
data b;
   set a;
   /* 示例字符串 */
   Name = "John Doe";

   /* 使用 scan() 拆分字符串 */
   FirstName = scan(Name, 1, ' '); /* 提取第一个单词（名字）
*/
   LastName = scan(Name, 2, ' ');  /* 提取第二个单词（姓氏）
*/
run;

proc print data=b;
run;
```

R

```
# 示例字符串向量
names <- c("John Doe", "Jane Smith", "Alice Johnson")

# 使用 strsplit() 拆分字符串
split_names <- strsplit(names, split = " ")

# 查看结果
print(split_names)
```

常见数据处理及转换 – 数据格式转换

SAS

```
data b;
   set a;

   /* 使用 input() 将字符型 Score 转换为数值型 */
   Score_numeric = input(Score, 8.);

   /* 如果 Grade 是字符型变量，可以用 input() 转换为数值型 */
   Grade_numeric = input(Grade, 8.);
run;

proc print data=b;
run;
```

R

```
# 示例数据框 a
a <- data.frame(
  ID = 1:4,
  Score = c("85", "90", "78", "88"),
  Grade = factor(c("1", "2", "3", "2"))
)

# 使用 as.numeric() 将字符型和因子型数据转换为数值型
a$Score <- as.numeric(a$Score)    # 将字符型 Score 转换为数值型
a$Grade <- as.numeric(as.character(a$Grade)) # 因子转换为字符再转换为数值

# 查看结果
print(a)
```

常见数据处理及转换 – ISO8601日期转换

SAS

```
data iso_format;
   set parsed_date;

   /* 提取年、月、日、小时和分钟 */
   Year = put(year(date_obj), 4.);
   Month = put(month(date_obj), z2.); /* 月份为两位数，用 z2. 格式 */
   Day = put(day(date_obj), z2.);    /* 日期为两位数，用 z2. 格式 */
   Hour = put(hour(date_obj), z2.);
   Minute = put(minute(date_obj), z2.);
   /* 拼接成 ISO 8601 格式 */
   iso_date = catx('T', catx('-', Year, Month, Day), catx(':', Hour, Minute));
   /* 输出 ISO 8601 格式 */
   format date_obj datetime20.;
run;

proc print data=iso_format;
run;
```

R

```
# 安装 lubridate 包（如果尚未安装）
# install.packages("lubridate")

# 加载 lubridate 包
library(lubridate)

# 示例日期字符串
dates <- c("18-05-2021 10:10", "2021/05/18 10:10 AM", "May 18, 2021 10:10")

# 使用 parse_date_time() 解析多种日期格式
date_objs <- parse_date_time(dates, orders = c("dmy HM", "ymd HMp", "b d, Y HM"))

# 使用 format() 转换为 ISO 8601 格式
iso_dates <- sapply(date_objs, function(x) format(x, "%Y-%m-%dT%H:%M"))

# 查看结果
print(iso_dates)
```

常见数据处理及转换 – 条件筛选

SAS

```
data b;
   set a;
   where Age > 25; /* 筛选 Age 大于 25 的行 */
run;

proc print data=b;
run;
```

R

```
# 加载 dplyr 包
library(dplyr)

# 示例数据框 a
a <- data.frame(
  ID = 1:5,
  Age = c(25, 30, 22, 28, 35),
  Weight = c(70, 80, 65, 85, 60)
)

# 使用 filter() 筛选 Age 大于 25 的行
filtered_data <- a %>%
  filter(Age > 25)

# 查看结果
print(filtered_data)
```

常见数据处理及转换 – 条件判断

SAS

```
data a_with_category;
   set a;
   if Age > 30 then Category = "Old";
   else if Age > 25 then Category = "Middle-aged";
   else Category = "Young";
run;

proc print data=a_with_category;
run;
```

R

```
# 示例数据框 a
a <- data.frame(
  ID = 1:5,
  Age = c(25, 30, 22, 28, 35)
)

# 使用 ifelse() 创建新列 Category
a$Category <- ifelse(a$Age > 30, "Old",
ifelse(a$Age > 25, "Middle-aged", "Young"))

# 查看结果
print(a)
```

常见数据处理及转换 – 修改增加新列

SAS

```
data b;
   set a;
   Height = 1.75;
   if ID = 2 then Height = 1.80;
   if ID = 3 then Height = 1.68;

   /* 计算 BMI 并创建新变量 */
   BMI = Weight / (Height ** 2);
run;

proc print data=b;
run;
```

R

```
# 加载 dplyr 包
library(dplyr)

# 示例数据框 a
a <- data.frame(
  ID = 1:3,
  Age = c(25, 30, 22),
  Weight = c(70, 80, 65)
)

# 使用 mutate() 添加新列 BMI
a <- a %>%
  mutate(Height = c(1.75, 1.80, 1.68),  # 添加 Height 列
         BMI = Weight / (Height ^ 2))   # 计算 BMI 值

# 查看结果
print(a)
```

常见数据处理及转换 – 变量排序

SAS

```
proc sort data=a out=b;
   by Age descending Weight; /* 先按 Age 升序，再按 Weight
降序 */
run;

proc print data=b;
run;
```

R

```
# 加载 dplyr 包
library(dplyr)

# 示例数据框 a
a <- data.frame(
  ID = 1:5,
  Age = c(25, 30, 22, 28, 35),
  Weight = c(70, 80, 65, 85, 60)
)

# 使用 arrange() 对 Age 进行升序排序
sorted_data <- a %>%
   arrange(Age, desc(Weight))
# 查看结果
print(sorted_data)
```

常见数据处理及转换 – 数据集合并

SAS

```
/* 使用 merge 和 in= 语句来求交集 */
data common;
   merge a(in=in_a) b(in=in_b);
   by ID Value;
   if in_a and in_b; /* 只保留在两个数据集中都存在的行 */
run;

proc print data=common;
run;
```

R

```
# 示例数据框 df1 和 df2
df1 <- data.frame(
  ID = c(1, 2, 3, 4),
  Value = c("A", "B", "C", "D")
)

df2 <- data.frame(
  ID = c(3, 4, 5, 6),
  Value = c("C", "D", "E", "F")
)

# 使用 intersect() 找到共同行
common_rows <- intersect(df1, df2)

# 查看结果
print(common_rows)
```

常见数据处理及转换 – 数据集连接

SAS

```
/* 使用 PROC SQL 进行左连接 */
proc sql;
   create table merged as
   select a.*, b.Age
   from a
   left join b
   on a.ID = b.ID;
quit;

proc print data=merged;
run;
```

R

```
# 加载 dplyr 包
library(dplyr)

# 示例数据框 a 和 b
a <- data.frame(
  ID = 1:5,
  Name = c("John", "Jane", "Paul", "Kate", "Tom")
)

b <- data.frame(
  ID = c(2, 3, 5, 6),
  Age = c(30, 22, 35, 40)
)

# 使用 left_join() 进行左连接
merged_data <- a %>%
  left_join(b, by = "ID")

# 查看结果
print(merged_data)
```

常见数据处理及转换 – 数据集转置

SAS

```
/* 示例数据集 a */
data a;
   input ID Year_2020 Year_2021 Year_2022;
   datalines;
   1 100 150 180
   2 200 250 280
   3 300 350 380
   ;
run;

/* 使用 PROC TRANSPOSE 将宽格式转换为长格式 */
proc transpose data=a out=long_data;
   by ID; /* 保留 ID 变量 */
   var Year_2020 Year_2021 Year_2022; /* 转换的列 */
run;

proc print data=long_data;
run;
```

R

```
# 示例数据框 a
a <- data.frame(
  ID = 1:3,
  Year_2020 = c(100, 200, 300),
  Year_2021 = c(150, 250, 350),
  Year_2022 = c(180, 280, 380)
)

# 使用 pivot_longer() 将宽格式转换为长格式
long_data <- a %>%
  pivot_longer(cols = starts_with("Year"),
               names_to = "Year",
               values_to = "Value")

# 查看结果
print(long_data)
```

常见数据处理及转换 – 数据集转置

R 4.4.1 · D:/ClinFunction/R/SARIS/1231/11x-statpr/testdir/program/sdtm_r/

```
> a <- data.frame(
+   ID = 1:3,
+   Year_2020 = c(100, 200, 300),
+   Year_2021 = c(150, 250, 350),
+   Year_2022 = c(180, 280, 380)
+ )
> print(a)
  ID Year_2020 Year_2021 Year_2022
1  1       100       150       180
2  2       200       250       280
3  3       300       350       380
>
```

R 4.4.1 · D:/ClinFunction/R/SARIS/1231/11x-statpr/testdir/program/sdtm_r/

```
> long_data <- a %>%
+   pivot_longer(cols = starts_with("Year"),
+                names_to = "Year",
+                values_to = "Value")
>
> # 查看结果
> print(long_data)
# A tibble: 9 × 3
     ID Year       Value
  <int> <chr>      <dbl>
1     1 Year_2020   100
2     1 Year_2021   150
3     1 Year_2022   180
4     2 Year_2020   200
5     2 Year_2021   250
6     2 Year_2022   280
7     3 Year_2020   300
8     3 Year_2021   350
9     3 Year_2022   380
>
```

常见数据处理及转换 – 列表循环

SAS

```
%macro square_vars(ds_in, ds_out);
   /* 创建输出数据集，平方所有变量 */
   data &ds_out.;
      set &ds_in.;
      array nums[*] _numeric_; /* 定义一个数组包含所有数值
变量 */
      do i = 1 to dim(nums);
         nums[i] = nums[i]**2; /* 将每个变量平方 */
      end;
      drop i; /* 删除辅助变量 i */
   run;
%mend;


/* 调用宏对数据集的每个数值变量进行平方操作 */
%square_vars(a, b);

proc print data=b;
run;
```

R

```
# 示例列表
text_list <- list(
  c("Hello", "world"),
  c("R", "is", "great"),
  c("Learning", "sapply")
)

# 使用 sapply() 计算每个字符串的长度
lengths <- sapply(text_list, function(x) nchar(x))

# 查看结果
print(lengths)
```

常见数据处理及转换 – 读取外部Excel

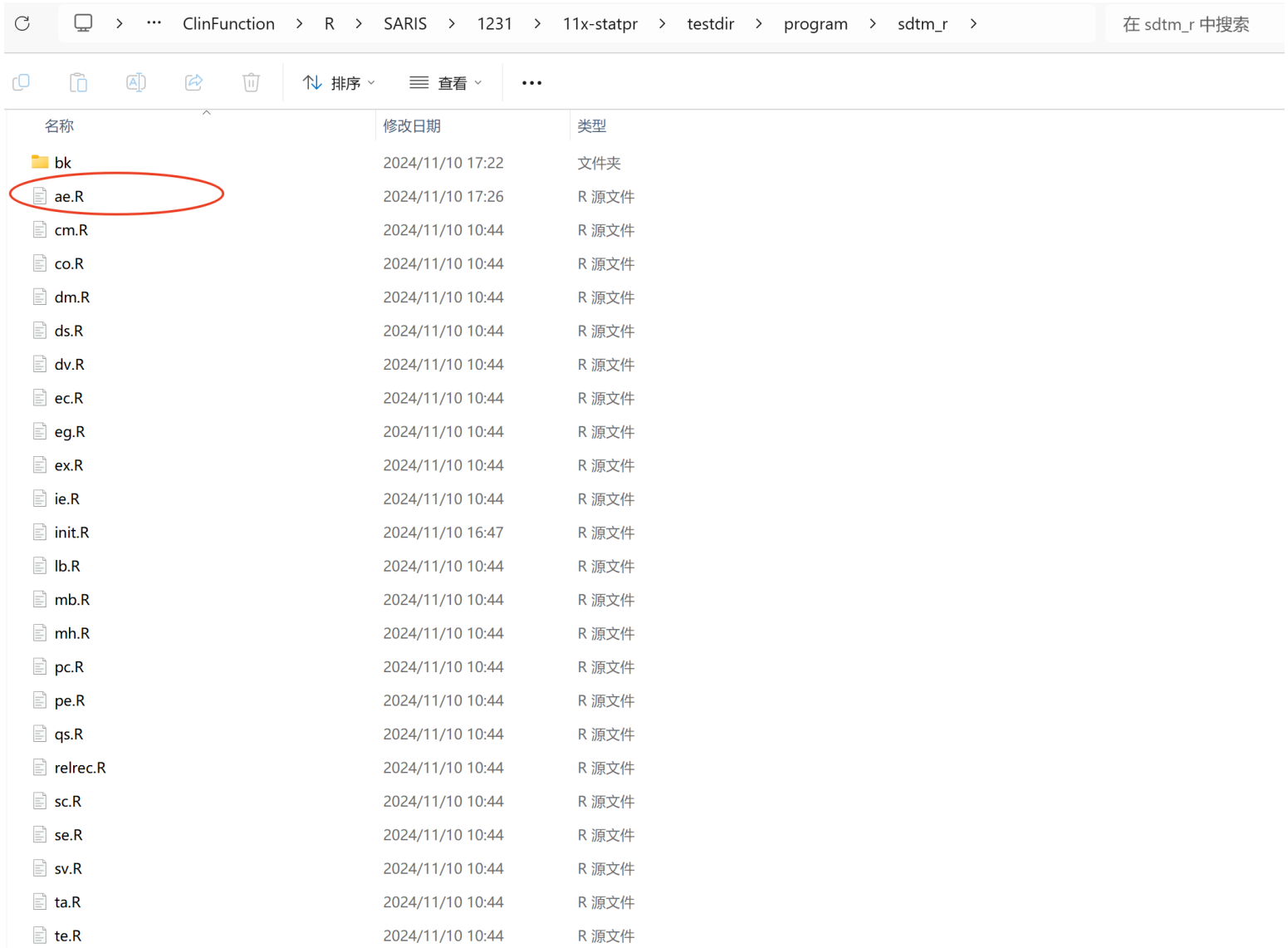
SAS

```
PROC IMPORT DATAFILE="data.xlsx"
   OUT=a
   DBMS=xlsx
   REPLACE;
   SHEET="Sheet1";  /* 指定工作表 */
   GETNAMES=YES;    /* 获取列名 */
RUN;

proc print data=a;
run;
```

R

```
# 读取 Excel 文件中特定范围的数据
data_range <- read_excel("data.xlsx", sheet = 1, range =
"A1:C10")

# 查看结果
print(data_range)
```

SDTM程序撰写 – AE为例



SDTM程序撰写 – AE为例

ae.R

D: > ClinFunction > R > SARIS > 1231 > 11x-statpr > testdir > program > sdtm_r > ae.R

```r
# 引用初始化脚本并读取原始数据
source("init.r")
library(haven)
library(readxl)
library(dplyr)

# 读取数据并创建 AE01 数据框
aec <- read_sas(file.path(raw_dataset_path, "AECODE.sas7bdat"))
dm <- read_sas(file.path(sdtm_dataset_path, "DM.sas7bdat"))
AE01 <- aec %>% left_join(dm, by = "SUBJID") %>% mutate(DOMAIN = "AE")

# 创建 AE02 数据框，重命名并计算新列
AE02 <- AE01 %>%
  rename(
    AETERM1 = AETERM,
    AESER1 = AESER,
    AEACN_ = AEACN,
    AEREL1 = AEREL,
    AEOUT1 = AEOUT,
    AESTDTC1 = AESTDTC,
    AEENDTC1 = AEENDTC
  ) %>%
  mutate(
    STUDYID = PSTUDYID,  # 添加 STUDYID 列
    AESPID = RECREP,
    AETERM = trimws(as.character(AETERM1)),
    AESER = ifelse(AESER1 == "是", "是", "否"),
    AEACN = trimws(as.character(AEACN_)),
    AEACNOTH = apply(select(., AEACN1, AEACN2, AEACN3), 1, function(x) paste(na.omit(ifelse(x == '1', c('不采取措施', '药物治疗', '其他治疗')[which(x == '1')], NA)), collapse = '+')),
    AEREL = trimws(as.character(AEREL1)),
    AEOUT = trimws(as.character(AEOUT1)),
    AEPRESP = ifelse(AENYN == "否", "是", NA),
    AESCONG = ifelse(SAE3 == "是", "是", NA),
    AESDISAB = ifelse(SAE4 == "是", "是", NA),
    AESDTH = ifelse(SAE1 == "是", "是", NA),
    AESHOSP = ifelse(SAE5 == "是", "是", NA),
    AESLIFE = ifelse(SAE2 == "是", "是", NA),
    AESMIE = ifelse(SAE6 == "是", "是", NA),
    AESTDTC = ifelse(!is.na(AESTDTC1) & !is.na(AESTTIM), paste(AESTDTC1, AESTTIM, sep = 'T'), AESTDTC1) %>% na_if('T'),
    AEENDTC = ifelse(!is.na(AEENDTC1) & !is.na(AEENTIM), paste(ifelse(grepl('UK', toupper(AEENDTC1)), gsub('UK', '-', toupper(AEENDTC1)), AEENDTC1), AEENTIM, sep = 'T'), AEENDTC1) %>%
      na_if('T') %>% na_if("-----"),
    RFSTDTC_d = sapply(strsplit(RFSTDTC, 'T'), `[`, 1),
    AESTDY = ifelse(nchar(AESTDTC1) == 10 & nchar(RFSTDTC_d) == 10,
                    as.Date(AESTDTC1) - as.Date(RFSTDTC_d) + as.integer(AESTDTC1 >= RFSTDTC_d), NA),
    AEENDY = ifelse(nchar(AEENDTC1) == 10 & nchar(RFSTDTC_d) == 10,
                    as.Date(AEENDTC1) - as.Date(RFSTDTC_d) + as.integer(AEENDTC1 >= RFSTDTC_d), NA),
    AELLT = trimws(as.character(LLT_Name)),
```

SDTM程序撰写 – AE为例

ae.R

D: > ClinFunction > R > SARIS > 1231 > 11x-statpr > testdir > program > sdtm_r > ae.R

```r
AE02 <- AE01 %>%
  mutate(
    AESOC = AEBODSYS,
    AESOCCD = AEBDSYCD
  ) %>%
  filter(!is.na(AETERM) & AETERM != "")

# 创建 AE02_EP 数据框并生成 EPF 标志
se <- read_sas(file.path(sdtm_dataset_path, "SE.sas7bdat"))
AE02_EP <- AE02 %>%
  left_join(se, by = "USUBJID", suffix = c("", ".se")) %>%
  arrange(USUBJID, AETERM, SESTDTC, SEENDTC) %>%
  group_by(USUBJID, AETERM) %>%
  mutate(
    is_last = row_number() == n(),
    EPF = case_when(
      is_last & (!is.na(SESTDTC) & AESTDTC1 >= SESTDTC & AESTDTC1 <= SEENDTC) ~ 'Y',
      EPOCH == '筛选期' & AESTDTC1 <= SEENDTC ~ 'Y',
      SESTDTC <= AESTDTC1 & AESTDTC1 < SEENDTC ~ 'Y',
      TRUE ~ NA_character_
    )
  ) %>%
  filter(!is.na(EPF)) %>%
  ungroup() %>%
  arrange(USUBJID, AETERM, AESTDTC1, SESTDTC, SEENDTC) %>%
  distinct(USUBJID, AETERM, AESTDTC1, SESTDTC, SEENDTC, .keep_all = TRUE)

# 保持原有行数，生成 AESEQ
ae <- AE02_EP %>%
  group_by(STUDYID, USUBJID) %>%
  mutate(AESEQ = row_number()) %>%
  ungroup()

# 读取并筛选 Excel 数据
excel_data <- read_excel(vdt_excel_path, sheet = "AE", skip = 10) %>% filter(`Mapping Action` == 'x')

# 获取要保留的列名并筛选 ae 中的合法列
ae <- ae %>% select(all_of(intersect(excel_data$`Variable Name`, names(.))))

# 添加列标签
ae <- ae %>% mutate(across(everything(), ~ {attr(., "label") <- excel_data$`Variable Label`[excel_data$`Variable Name` == cur_column()]; .}))

# 定义保存路径并保存为 SAS 数据集和 SAS Transport 格式
dir.create(sdtm_dataset_path, recursive = TRUE, showWarnings = FALSE)
write_sas(ae, file.path(sdtm_dataset_path, "ae.sas7bdat"))
write_xpt(ae, file.path(sdtm_dataset_path, "ae.xpt"))
cat("文件已保存到:", file.path(sdtm_dataset_path, "ae.sas7bdat"), "\n")
```

SDTM程序撰写 – AE为例

adsl.R × ae.R × ae × init.R ×

Filter

	STUDYID 研究标识符	DOMAIN 域名缩写	USUBJID 受试者唯一标识符	AESEQ 序号	AESPID 自定义标识符	AETERM 不良事件报告名称	AELLT 低位语	AELLTCD 低位语编码	AEDECOD 标准化名称	AEPTCD 首选语编码	AEHLT 高位语	AEHLTCD 高位语编码	AEHLGT 高位组语	AEHLGTCD 高位组语编码	AEPRES 预设
1	[redacted]A103	AE	-S001	1	2	丙氨酸氨基转移酶升高	丙氨酸氨基转移酶升高	10001551	丙氨酸氨基转移酶升高	10001551	各种肝功能分析	10024689	肝胆系统检查	10019809	是
2	[redacted]A103	AE	-S001	2	1	天门冬氨酸氨基转移酶升高	天门冬氨酸氨基转移酶升高	10003481	天门冬氨酸氨基转移酶升高	10003481	各种肝功能分析	10024689	肝胆系统检查	10019809	是
3	[redacted]A103	AE	-S002	1	1	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
4	[redacted]A103	AE	-S002	2	2	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
5	[redacted]A103	AE	-S002	3	3	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
6	[redacted]A103	AE	-S002	4	4	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
7	[redacted]A103	AE	-S002	5	5	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
8	[redacted]A103	AE	-S003	1	2	肌酐升高	肌酐增加	10011368	血肌酐升高	10005483	各种肾功能分析	10038454	肾脏、泌尿道检查及尿液分析	10038362	是
9	[redacted]A103	AE	-S003	2	3	血胆红素增高	血胆红素升高	10005364	血胆红素升高	10005364	各种肝功能分析	10024689	肝胆系统检查	10019809	是
10	[redacted]A103	AE	-S003	3	1	血胆红素增高	血胆红素升高	10005364	血胆红素升高	10005364	各种肝功能分析	10024689	肝胆系统检查	10019809	是
11	[redacted]A103	AE	-S008	1	1	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
12	[redacted]A103	AE	-S009	1	1	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
13	[redacted]A103	AE	-S009	2	2	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
14	[redacted]A103	AE	-S009	3	4	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
15	[redacted]A103	AE	-S010	1	1	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
16	[redacted]A103	AE	-S010	2	2	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
17	[redacted]A103	AE	-S012	1	1	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
18	[redacted]A103	AE	-S019	1	1	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
19	[redacted]A103	AE	-S019	2	2	血胆红素升高	血胆红素升高	10005364	血胆红素升高	10005364	各种肝功能分析	10024689	肝胆系统检查	10019809	是
20	[redacted]A103	AE	-S023	1	1	低钾血症	低钾血症	10021015	低钾血症	10021015	钾平衡失调	10036451	电解质及液体平衡类体况	10014412	是
21	[redacted]A103	AE	-S023	2	2	贫血	贫血	10002034	贫血	10002034	各种贫血（不另分类）	10002067	非溶血性贫血及骨髓抑制	10002086	是
22	[redacted]A103	AE	-S023	3	3	贫血	贫血	10002034	贫血	10002034	各种贫血（不另分类）	10002067	非溶血性贫血及骨髓抑制	10002086	是
23	[redacted]A103	AE	-S025	1	1	偶发房性期前收缩	房性早搏	10036591	室上性期外收缩	10042602	各种室上性心律失常	10042600	心律失常类疾病	10007521	NA
24	[redacted]A103	AE	-S025	2	3	血中性粒细胞降低	血中性粒细胞计数降低	10005673	中性粒细胞计数降低	10029366	各种白细胞分析	10047938	血液类检查（包括血型）	10018851	是
25	[redacted]A103	AE	-S025	3	2	血白细胞降低	白细胞计数降低	10047942	白细胞计数降低	10047942	各种白细胞分析	10047938	血液类检查（包括血型）	10018851	是
26	[redacted]A103	AE	-S025	4	4	贫血	贫血	10002034	贫血	10002034	各种贫血（不另分类）	10002067	非溶血性贫血及骨髓抑制	10002086	NA
27	[redacted]A103	AE	-S032	1	1	低血压	低血压	10021097	低血压	10021097	血管低血压疾病	10057181	血压降低和性质不明的血压异常及休克	10011954	是
28	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Showing 1 to 28 of 80 entries, 35 total columns

Console Terminal × Background Jobs ×

R 4.4.1 · D:/ClinFunction/R/SARIS/1231/11x-statpr/testdir/program/sdtm_r/

文件已保存到: D:/ClinFunction/R/SARIS/1231/11x-statpr/testdir/dataset/sdtm_r/ae.sas7bdat

SDTM程序撰写 – AE为例

```r

# 保持原有行数，生成 AESEQ
ae <- AE02_EP %>%
  group_by(STUDYID, USUBJID) %>%
  mutate(AESEQ = row_number()) %>%
  ungroup()

# 读取并筛选 Excel 数据
excel_data <- read_excel(vdt_excel_path, sheet = "AE", skip = 10) %>% filter(`Mapping Action` == 'x')

# 获取要保留的列名并筛选 ae 中的合法列
ae <- ae %>% select(all_of(intersect(excel_data$`Variable Name`, names(.))))

# 添加列标签
ae <- ae %>% mutate(across(everything(), ~ {attr(., "label") <- excel_data$`Variable Label`[excel_data$`Variable Name` == cur_column()]; .}))

# 定义保存路径并保存为 SAS 数据集和 SAS Transport 格式
dir.create(sdtm_dataset_path, recursive = TRUE, showWarnings = FALSE)
write_sas(ae, file.path(sdtm_dataset_path, "ae.sas7bdat"))
```

SDTM程序撰写 – AE为例

Protocol Number:	SAL086A103
SDTM Dataset Description:	不良事件
SDTM Dataset Structure:	每个受试者每例次不良事件一条记录
Program Name:	AE.sas
Creation Order:	4
Source Data Used:	AE
Key Variables:	STUDYID, USUBJID, AEDECOD, AETERM, AESTDTC, AEENDTC
Sort Order:	USUBJID, AESEQ

Specifications for AE Dataset

Content

Domain	Variable Order	Variable Name	Variable Label	Type	Length	Role	Core	Controlled Terminology	Origin	CRF No	Page	Mapping Action	Raw Variable or Hardcode Value	Mapping Rule	Programmer Notes	CDISC Notes
AE	1	STUDYID	研究标识符	Char		Identifier	Req		Protocol			x		"SAL086A103"		Unique identifier for a study.
AE	2	DOMAIN	域名缩写	Char	2	Identifier	Req	AE	Assigned			x		"AE"		Two-character abbreviation f
AE	3	USUBJID	受试者唯一标识符	Char		Identifier	Req		Derived			x	SDTM.DM.USUBJID	给值SDTM.DM.USUBJID		Identifier used to uniquely ide studies for all applications or product.
AE	4	AESEQ	序号	Num	8	Identifier	Req		Derived			x		按关键变量排序，为每个USUBJID分配从1开始的序列号。		Sequence Number given to e records within a domain. May
AE	5	AEGRPID	组合ID	Char		Identifier	Perm					NA				Used to tie together a block c domain for a subject.
AE	6	AEREFID	参考ID	Char		Identifier	Perm					NA				Internal or external identifier SAE reporting form.
AE	7	AESPID	自定义标识符	Char		Identifier	Perm		Assigned			x		记录号		Sponsor-defined identifier. It CRF as an explicit line identif operational database. Examp Events page.
AE	8	AETERM	不良事件报告名称	Char		Topic	Req		CRF	90		x		不良事件页面，不良事件名称		Verbatim name of the event.
AE	9	AEMODIFY	修正词	Char		Synonym Qualifier	Perm		Assigned			NA				If AETERM is modified to fac will contain the modified text.
AE	10	AELLT	低位语	Char		Variable Qualifier	Exp	MedDRA	Assigned			x	AECODE.LLT_NAME	低位语		Dictionary-derived text descr Term.
AE	11	AELLTCD	低位语编码	Num	8	Variable Qualifier	Exp	MedDRA	Assigned			x	AECODE.LLT_CODE	低位语编码		Dictionary-derived code for t
AE	12	AEDECOD	标准化名称	Char		Synonym Qualifier	Req	MedDRA	Assigned			x	AECODE.PT_NAME	标准化名称		Dictionary-derived text descr AEMODIFY. Equivalent to the MedDRA). The sponsor is ex dictionary name and version the define.xml external codeli
AE	13	AEPTCD	首选语编码	Num	8	Variable Qualifier	Exp	MedDRA	Assigned			x	AECODE.PT_CODE	首选语编码		Dictionary-derived code for t
AE	14	AEHLT	高位语	Char		Variable Qualifier	Exp	MedDRA	Assigned			x	AECODE.HLT_NAME	高位语		Dictionary-derived text descr for the primary System Orga
AE	15	AEHLTCD	高位语编码	Num	8	Variable Qualifier	Exp	MedDRA	Assigned			x	AECODE.HLT_CODE	高位语编码		Dictionary-derived code for t primary System Organ Class
AE	16	AEHLGT	高位组语	Char		Variable Qualifier	Exp	MedDRA	Assigned			x	AECODE.HLGT_NAME	高位组语		Dictionary-derived text descr Term for the primary System
AE	17	AEHLGTCD	高位组语编码	Num	8	Variable Qualifier	Exp	MedDRA	Assigned			x	AECODE.HLGT_CODE	高位组语编码		Dictionary-derived code for t the primary System Organ C
AE	18	AECAT	类别	Char		Grouping Qualifier	Perm	AECAT				NA				Used to define a category of BLEEDING, NEUROPSYCHI
AE	19	AESCAT	子类	Char		Grouping Qualifier	Perm	AESCAT				NA				A further categorization of ac NEUROLOGIC.
AE	20	AEPRESP	预设	Char	1	Variable Qualifier	Perm	(NY)	CRF	91		x		不良事件页面，是否为非预期不良反应勾选否，给值"是"。		A value of "Y" indicates that t specified on the CRF. Values reported events (i.e., those c terms)
AE	21	AEBODSYS	系统器官分类	Char		Record Qualifier	Exp	MedDRA	Assigned			x	AECODE.SOC_NAME	系统器官分类		Dictionary derived. Body sys the sponsor from the coding When using a multi-axial dict should contain the SOC used and summary tables which m primary SOC.
AE	22	AEBDSYCD	系统器官分类编码	Num	8	Variable Qualifier	Exp	MedDRA	Assigned			x	AECODE.SOC_CODE	系统器官分类编码		Dictionary-derived. Code for

... TV_detail | CO | DM | SE | SV | SV_detail | CM | EX | PR | AE | DS | DV | MH | EG | IE | LB | LB_Metadata | MB | PC | PE | PP | PP_Detail | QS | SC | VS | XU | XU_Metadata | RELREC | SUPPQUA ...

SDTM程序撰写 – AE为例

```
%dsoutput(type      = sdtm
         ,domain    = AE
         ,inds      = AE05
         ,outds     = ae
         ,outxpt    =
         ,compress  = y
         );
```

```
31 > 11x-statpr > testdir > program > macros > dsoutput-v2-6.
in =
ase(&domain.)="AVITALS" %then %let dslabel =Vital
ase(&domain.)="APE" %then %let dslabel =Physical
ase(&domain.)="ALAB_RAW" %then %let dslabel =Raw

&domain;

eq 0 %then %do;
macro parameter(domain) can not be null;
1;
```

```
%if %length(&OutEmptySDTM)=0 %then %let OutEmptySDTM=1;

%put Macro(&SYSMACRONAME.) : ============================================
    Begin domain=&domain.;

%if &Debug>0 %then %do;
%put ;
%put Debug=&debug. (&SYSMACRONAME.) : ================================
    1. local or global macro variables;
%end;

    %local _datalib _xptpath _outxpt outds _ContentDsInDomain;

%if &Debug>0 %then %do;
%put ;
%put Debug=&debug. (&SYSMACRONAME.) : ================================
    2. Check validate of macro parameters, and set default value for s
%end;

    %if "%upcase(&outxpt.)" eq "Y" or "%upcase(&outxpt.)" eq "1" %ther
        %else %let _outxpt = N;
    %if "%upcase(&compress.)" eq "Y" or "%upcase(&compress.)" eq "1" %
        %else %let _compress=N;

    %if %length(&role.)=0 and %symexist(_role) %then %let role=&_role.
    %if %length(&role.)=0 %then %do;
        %put ERROR: Please define the parameter(role=);
        %goto exit;
    %end;
    %if %length(&type.)=0 and %symexist(_type) %then %let type=&_type.
    %if %length(&type.)=0 %then %do;
        %put ERROR: Please define the parameter(type=);
        %goto exit;
    %end;
```

```
ae.sas   dsoutput-v2-6.sas
D: > ClinFunction > R > SARIS > 1231 > 11x-statpr > testdir > program > macros > dsoutput-v2-6.sas > {} %MACRO
%macro dsoutput(domain =
    libname xptout xport "&_xptpath.\%lowcase(&outds.).xpt";
    proc copy in=&_datalib. out=xptout ;select %lowcase(&outds.);quit;
    libname xptout clear;
%end;

/**  output the data set with CSV format **/
%if &excelyn=1 or "%upcase(&excelyn)"="Y" %then %do;
    %if %sysfunc(fileexist(&_xptpath.\&outds..csv)) ^= 0 %then %do;
        options noxwait XSYNC;
        %local rc filrf;
        %let rc = %sysfunc(filename(filrf,&_xptpath.\&outds..csv));
        %if &rc = 0 and %sysfunc(fexist(&filrf.)) %then %let rc = %sysfunc(fdelete(&filrf.));
        %let rc = %sysfunc(filename(&filrf.));
    %end;
    %ds2csv (data=&_datalib..&outds., csvfile=&_xptpath.\&outds..csv, runmode=b,openmode=REPLACE,labels=N);
%end;

%exit:
%if &error_num ne 0 %then %do;
    %put ERROR: Exit macro(&SYSMACRONAME.) because of error;
    %put ERROR: &error_msg;
%end;

/**  Check Dataset name, Dataset label, Variable name, Variable label **/
%if &chk_vdt=1 %then %do;
    %if %upcase(&type) eq SDTM or %upcase(&type) eq ADAM %then %do;
        %check_vdt(vdt=CDISC_&type., tocdataset=&type.spec_content, dataset= &type.spec_&domain. ,select=&select,exclude=&exclude);
    %end;
    %else %if %upcase(&type) eq ISDB %then %do;
        %check_vdt(vdt=&type._ds,tocdataset=, dataset=&type.spec_&domain. ,select=&select,exclude=&exclude);
    %end;
    %else %do;
        %if &_pktype eq NONMEM %then %let exclude=dsnamelen dslabellen varnamelen varlabellen varlen;
        %check_vdt(vdt=PK,tocdataset=, dataset= &type.spec_&domain. ,select=&select,exclude=&exclude);
    %end;
%end;
%**delete temp datasets;

%if &Debug<=0 %then %do;
    proc datasets nolist nodetails NoWarn lib=work;
        delete &type.spec_&domain. _spec_con  _spec_labid _ds_inds _ds_outds1 &outds &type.spec_content/memtype=data;
    quit;
%end;

%put Macro(&SYSMACRONAME.) : ============================================
    End domain=%upcase(&outds.);
```

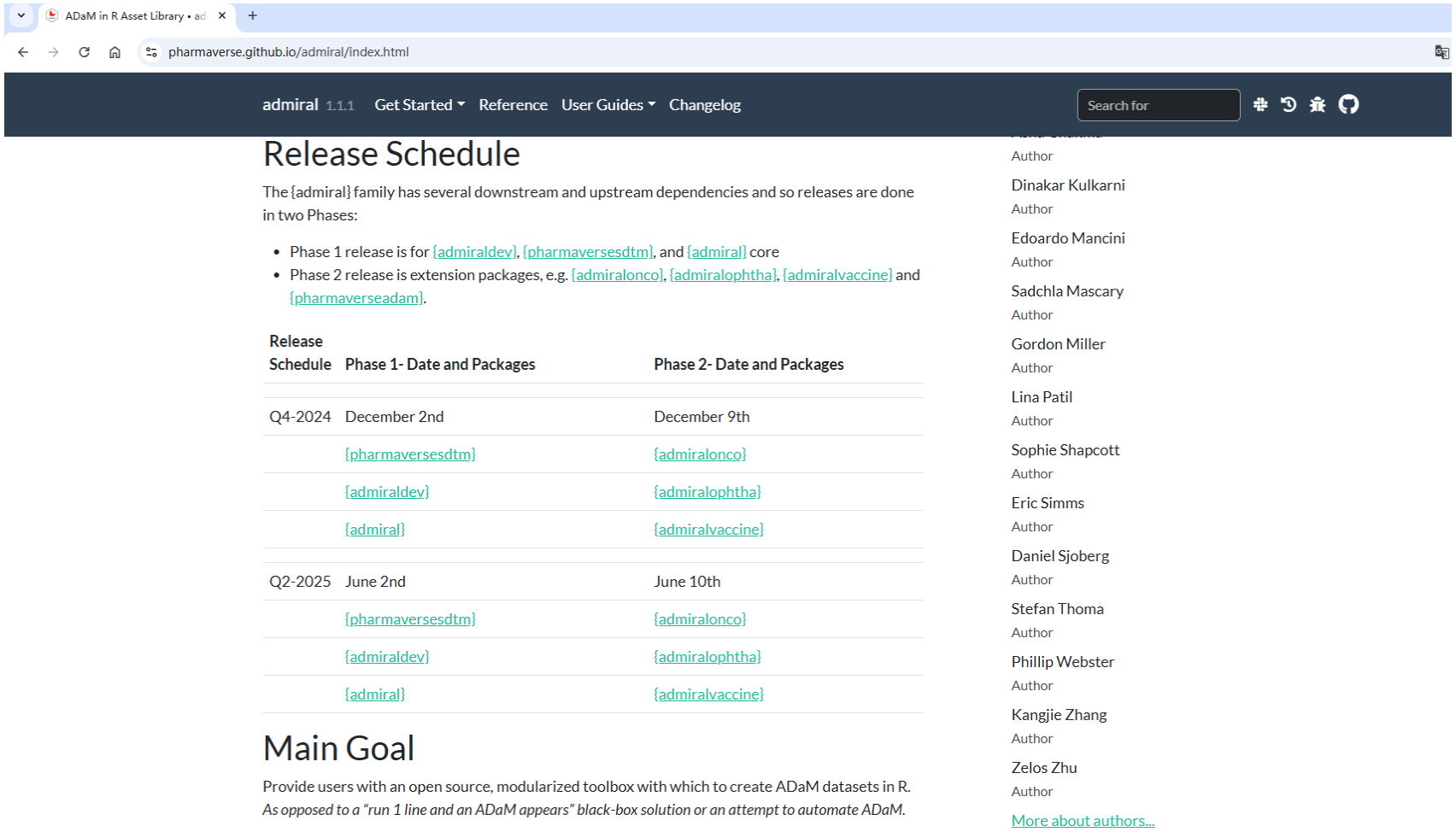
PART
03

R生成ADaM数据集

Generating ADaM Datasets with R

Admiral包介绍

专门用于支持制药行业和临床研究中 **ADaM（Analysis Data Model）** 数据集的生成。它由Roche和Atorus Research的团队开发。 https://pharmaverse.github.io/admiral/index.html



功能和优势

- ✓ **标准化和模块化的ADaM数据生成：**如自动生成ADSL、ADAE，每个TA一个包
- ✓ **简化的编程流程:** 提供生成ADaM数据集常见操作的一系列函数
- ✓ **自动化和可重用性：**如用于生成关键变量（基线特征、疗效终点等）模版
- ➢ 同时也鼓励各个制药公司和CRO按照模块化的方法去Admiral上递交程序

Admiral包-生成ADSL

```r
# 2. 从ex数据集中派生治疗开始日期（TRTSDTM）和结束日期（TRTEDTM）
ex_ext <- ex %>%
  derive_vars_dtm(dtc = EXSTDTC, new_vars_prefix = "EXST") %>%
  derive_vars_dtm(dtc = EXENDTC, new_vars_prefix = "EXEN")

adsl <- adsl %>%
  derive_vars_merged(
    dataset_add = ex_ext,
    filter_add = EXDOSE > 0 & !is.na(EXSTDTM),
    new_vars = exprs(TRTSDTM = EXSTDTM),
    order = exprs(EXSTDTM, EXSEQ),
    mode = "first",
    by_vars = exprs(STUDYID, USUBJID)
  ) %>%
  derive_vars_merged(
    dataset_add = ex_ext,
    filter_add = EXDOSE > 0 & !is.na(EXENDTM),
    new_vars = exprs(TRTEDTM = EXENDTM),
    order = exprs(EXENDTM, EXSEQ),
    mode = "last",
    by_vars = exprs(STUDYID, USUBJID)
  )

# 3. 将日期时间变量转换为日期变量
adsl <- adsl %>%
  derive_vars_dtm_to_dt(source_vars = exprs(TRTSDTM, TRTEDTM))

# 4. 从ds数据集中提取研究结束日期（EOSDT）和状态（EOSSTT）
ds_ext <- ds %>%
  derive_vars_dt(dtc = DSSTDTC, new_vars_prefix = "DSST")

adsl <- adsl %>%
  derive_vars_merged(
    dataset_add = ds_ext,
    by_vars = exprs(STUDYID, USUBJID),
    new_vars = exprs(EOSDT = DSSTDT),
    filter_add = DSCAT == "DISPOSITION EVENT" & DSDECOD != "SCREEN FAILURE"
  ) %>%
  derive_vars_merged(
    dataset_add = ds_ext,
    by_vars = exprs(STUDYID, USUBJID),
    filter_add = DSCAT == "DISPOSITION EVENT",
    new_vars = exprs(EOSSTT = case_when(
      DSDECOD == "COMPLETED" ~ "COMPLETED",
      DSDECOD == "SCREEN FAILURE" ~ NA_character_,
      !is.na(DSDECOD) ~ "DISCONTINUED",
      TRUE ~ "ONGOING"
    )),
    missing_values = exprs(EOSSTT = "ONGOING")
```

Admiral包-生成ADSL

init.R | adsl.R | adsl

Filter

	STUDYID 研究标识符	USUBJID 受试者唯一标识符	AGE 年龄	SEX 性别	RACE 种族	TRTSDTM	TRTEDTM	TRTSDT	TRTEDT	EOSDT	EOSSTT
1	[redacted]6A103	[redacted]-01-S001	49	女性	亚裔	2021-05-12 08:00:00	2021-05-21 20:00:00	2021-05-12	2021-05-21	NA	ONGOING
2	[redacted]6A103	[redacted]-01-S002	36	女性	亚裔	2021-05-12 08:02:00	2021-05-21 20:02:00	2021-05-12	2021-05-21	NA	ONGOING
3	[redacted]6A103	[redacted]-01-S003	21	男性	亚裔	2021-05-12 08:40:00	2021-05-21 20:40:00	2021-05-12	2021-05-21	NA	ONGOING
4	[redacted]6A103	[redacted]-01-S004	32	男性	亚裔	NA	NA	NA	NA	NA	ONGOING
5	[redacted]6A103	[redacted]-01-S005	36	男性	亚裔	NA	NA	NA	NA	NA	ONGOING
6	[redacted]6A103	[redacted]-01-S006	36	女性	亚裔	NA	NA	NA	NA	NA	ONGOING
7	[redacted]6A103	[redacted]-01-S007	31	女性	亚裔	2021-05-12 08:04:00	2021-05-21 20:04:00	2021-05-12	2021-05-21	NA	ONGOING
8	[redacted]6A103	[redacted]-01-S008	42	女性	亚裔	2021-05-12 08:06:00	2021-05-21 20:06:00	2021-05-12	2021-05-21	NA	ONGOING
9	[redacted]6A103	[redacted]-01-S009	39	女性	亚裔	2021-05-12 08:08:00	2021-05-21 20:08:00	2021-05-12	2021-05-21	NA	ONGOING
10	[redacted]6A103	[redacted]-01-S010	50	女性	亚裔	2021-05-12 08:10:00	2021-05-21 20:10:00	2021-05-12	2021-05-21	NA	ONGOING
11	[redacted]6A103	[redacted]-01-S011	35	女性	亚裔	2021-05-12 08:12:00	2021-05-21 20:12:00	2021-05-12	2021-05-21	NA	ONGOING
12	[redacted]6A103	[redacted]-01-S012	44	女性	亚裔	2021-05-12 08:14:00	2021-05-21 20:15:00	2021-05-12	2021-05-21	NA	ONGOING
13	[redacted]6A103	[redacted]-01-S013	31	男性	亚裔	NA	NA	NA	NA	NA	ONGOING
14	[redacted]6A103	[redacted]-01-S014	43	女性	亚裔	NA	NA	NA	NA	NA	ONGOING
15	[redacted]6A103	[redacted]-01-S015	31	女性	亚裔	NA	NA	NA	NA	NA	ONGOING
16	[redacted]6A103	[redacted]-01-S016	46	女性	亚裔	NA	NA	NA	NA	NA	ONGOING
17	[redacted]6A103	[redacted]-01-S017	42	男性	亚裔	2021-05-12 08:42:00	2021-05-21 20:42:00	2021-05-12	2021-05-21	NA	ONGOING
18	[redacted]6A103	[redacted]-01-S018	27	女性	亚裔	NA	NA	NA	NA	NA	ONGOING
19	[redacted]6A103	[redacted]-01-S019	40	男性	亚裔	2021-05-12 08:44:00	2021-05-21 20:44:00	2021-05-12	2021-05-21	NA	ONGOING
20	[redacted]6A103	[redacted]-01-S020	34	女性	亚裔	NA	NA	NA	NA	NA	ONGOING
21	[redacted]6A103	[redacted]-01-S021	36	男性	亚裔	2021-05-12 08:46:00	2021-05-21 20:46:00	2021-05-12	2021-05-21	NA	ONGOING
22	[redacted]6A103	[redacted]-01-S022	36	男性	亚裔	2021-05-12 08:48:00	2021-05-21 20:48:00	2021-05-12	2021-05-21	NA	ONGOING

Showing 1 to 22 of 205 entries, 11 total columns

Console | Terminal | Background Jobs

R 4.4.1 · D:/ClinFunction/R/SARIS/1231/11x-statpr/testdir/program/adam_r/

```
# ℹ Use `print(n = ...)` to see more rows
> View(adsl)
```

Admiral包-生成ADAE

adsl.R | init.R | adae.R

Source on Save

```r
# 加载必要的库
source("init.r")
library(admiral)
library(dplyr)
library(haven)


# 读取SDTM数据集
ae <- read_sas(file.path(sdtm_dataset_path, "ae.sas7bdat")) # 不良事件数据
dm <- read_sas(file.path(sdtm_dataset_path, "dm.sas7bdat")) # 人口统计学数据

# 1. 提取受试者的人口统计学信息
adsl <- dm %>%
  select(STUDYID, USUBJID, AGE, SEX, RACE)

# 基于admiral派生ADAE
adae <- ae %>%
  # 添加是否为严重不良事件（AESER）
  mutate(
    AESER = case_when(AESER == "Y" ~ "Y", TRUE ~ "N"),
    # 添加是否治疗相关（AERELFL）
    AERELFL = case_when(AEREL == "RELATED" ~ "Y", TRUE ~ "N"),
    # 添加是否导致治疗中断（AESTDISP）
    AESTDISP = case_when(AEACN == "DRUG WITHDRAWN" ~ "Y", TRUE ~ "N")
  ) %>%
  # 派生开始日期（ASTDT）和结束日期（AENDT）
  derive_vars_dtm(dtc = AESTDTC, new_vars_prefix = "AST") %>%
  derive_vars_dtm(dtc = AEENDTC, new_vars_prefix = "AEN") %>%
  derive_vars_dtm_to_dt(source_vars = exprs(ASTDTM, AENDTM)) %>%
  # 合并人口统计信息
  left_join(adsl, by = c("STUDYID", "USUBJID")) %>%
  # 标注首次事件
  derive_var_extreme_flag(
    by_vars = exprs(STUDYID, USUBJID, AEDECOD),
    order = exprs(ASTDTM),
    new_var = FAEFL,
    mode = "first"
  )


# 3. 检查输出
print(adae)

```

Admiral包-生成ADAE

adsl.R × init.R × adae.R × adae ×

Filter

DSP 院或住院时间延长	AESLIFE 危及生命	AESMIE 其他重要的医学事件	EPOCH 时期	AESTDTC 开始日期/时间	AEENDTC 结束日期/时间	AESTDY 开始日	AEENDY 结束日	AERELFL	AESTDISP	ASTDTM	ASTTMF	AENDTM	AENTMF	ASTDT	AENDT	AGE 年龄	SEX 性别	RACE 种族	FAEFL
	NA	NA	治疗期	2021-05-18T07:50	2021-05-23T08:00	7	12	N	N	2021-05-18 07:50:00	S	2021-05-23 08:00:00	S	2021-05-18	2021-05-23	49	女性	亚裔	Y
	NA	NA	治疗期	2021-05-18T07:50	2021-05-23T08:00	7	12	N	N	2021-05-18 07:50:00	S	2021-05-23 08:00:00	S	2021-05-18	2021-05-23	49	女性	亚裔	Y
	NA	NA	治疗期	2021-05-12T16:14	2021-05-12T17:54	1	1	N	N	2021-05-12 16:14:00	S	2021-05-12 17:54:00	S	2021-05-12	2021-05-12	36	女性	亚裔	Y
	NA	NA	治疗期	2021-05-17T07:42	2021-05-17T10:02	6	6	N	N	2021-05-17 07:42:00	S	2021-05-17 10:02:00	S	2021-05-17	2021-05-17	36	女性	亚裔	NA
	NA	NA	治疗期	2021-05-18T16:14	2021-05-18T16:22	7	7	N	N	2021-05-18 16:14:00	S	2021-05-18 16:22:00	S	2021-05-18	2021-05-18	36	女性	亚裔	NA
	NA	NA	治疗期	2021-05-20T16:12	2021-05-20T16:40	9	9	N	N	2021-05-20 16:12:00	S	2021-05-20 16:40:00	S	2021-05-20	2021-05-20	36	女性	亚裔	NA
	NA	NA	治疗期	2021-05-21T10:12	2021-05-21T10:26	10	10	N	N	2021-05-21 10:12:00	S	2021-05-21 10:26:00	S	2021-05-21	2021-05-21	36	女性	亚裔	NA
	NA	NA	治疗期	2021-05-23T08:40	2021-05-31T08:43	12	20	N	N	2021-05-23 08:40:00	S	2021-05-31 08:43:00	S	2021-05-23	2021-05-31	21	男性	亚裔	Y
	NA	NA	治疗期	2021-05-14T08:40	2021-05-18T08:30	3	7	N	N	2021-05-14 08:40:00	S	2021-05-18 08:30:00	S	2021-05-14	2021-05-18	21	男性	亚裔	Y
	NA	NA	治疗期	2021-05-23T08:40	2021-05-31T08:43	12	20	N	N	2021-05-23 08:40:00	S	2021-05-31 08:43:00	S	2021-05-23	2021-05-31	21	男性	亚裔	NA
	NA	NA	治疗期	2021-05-21T10:16	2021-05-21T10:23	10	10	N	N	2021-05-21 10:16:00	S	2021-05-21 10:23:00	S	2021-05-21	2021-05-21	42	女性	亚裔	Y
	NA	NA	治疗期	2021-05-14T08:18	2021-05-14T08:27	3	3	N	N	2021-05-14 08:18:00	S	2021-05-14 08:27:00	S	2021-05-14	2021-05-14	39	女性	亚裔	Y
	NA	NA	治疗期	2021-05-16T16:18	2021-05-16T16:33	5	5	N	N	2021-05-16 16:18:00	S	2021-05-16 16:33:00	S	2021-05-16	2021-05-16	39	女性	亚裔	NA
	NA	NA	治疗期	2021-05-17T07:48	2021-05-17T08:03	6	6	N	N	2021-05-17 07:48:00	S	2021-05-17 08:03:00	S	2021-05-17	2021-05-17	39	女性	亚裔	NA
	NA	NA	治疗期	2021-05-12T16:25	2021-05-12T17:43	1	1	N	N	2021-05-12 16:25:00	S	2021-05-12 17:43:00	S	2021-05-12	2021-05-12	50	女性	亚裔	Y
	NA	NA	治疗期	2021-05-18T10:10	2021-05-18T10:19	7	7	N	N	2021-05-18 10:10:00	S	2021-05-18 10:19:00	S	2021-05-18	2021-05-18	50	女性	亚裔	NA
	NA	NA	治疗期	2021-05-12T16:30	2021-05-12T17:51	1	1	N	N	2021-05-12 16:30:00	S	2021-05-12 17:51:00	S	2021-05-12	2021-05-12	44	女性	亚裔	Y
	NA	NA	治疗期	2021-05-21T16:54	2021-05-21T17:25	10	10	N	N	2021-05-21 16:54:00	S	2021-05-21 17:25:00	S	2021-05-21	2021-05-21	40	男性	亚裔	Y
	NA	NA	治疗期	2021-05-23T08:44	2021-05-31T08:38	12	20	N	N	2021-05-23 08:44:00	S	2021-05-31 08:38:00	S	2021-05-23	2021-05-31	40	男性	亚裔	Y
	NA	NA	治疗期	2021-05-14T08:16	2021-05-15T06:25	3	4	N	N	2021-05-14 08:16:00	S	2021-05-15 06:25:00	S	2021-05-14	2021-05-15	29	女性	亚裔	Y
	NA	NA	治疗期	2021-05-14T08:16	2021-05-18T08:06	3	7	N	N	2021-05-14 08:16:00	S	2021-05-18 08:06:00	S	2021-05-14	2021-05-18	29	女性	亚裔	Y
	NA	NA	治疗期	2021-05-23T08:16	-----Tuk:uk	12	NA	N	N	2021-05-23 08:16:00	S	NA	NA	2021-05-23	NA	29	女性	亚裔	NA
	NA	NA	治疗期	2021-05-23T08:20	2021-06-08T08:06	12	28	N	N	2021-05-23 08:20:00	S	2021-06-08 08:06:00	S	2021-05-23	2021-06-08	35	女性	亚裔	Y
	NA	NA	治疗期	2021-05-23T08:30	2021-05-31T08:48	12	20	N	N	2021-05-23 08:30:00	S	2021-05-31 08:48:00	S	2021-05-23	2021-05-31	35	女性	亚裔	Y
	NA	NA	治疗期	2021-05-23T08:20	2021-06-08T08:06	12	28	N	N	2021-05-23 08:20:00	S	2021-06-08 08:06:00	S	2021-05-23	2021-06-08	35	女性	亚裔	Y

Admiral包-疗效性分析

```
      RECIST = case_when(
        RSSTRESC == "CR" ~ "Complete Response",
        RSSTRESC == "PR" ~ "Partial Response",
        RSSTRESC == "SD" ~ "Stable Disease",
        RSSTRESC == "PD" ~ "Progressive Disease",
        TRUE ~ "Not Evaluable"
      )
    )

# 3. 合并 SLD 和 RS 数据集，并标记进展事件
adrs <- rs %>%
  derive_vars_merged(
    dataset_add = sld,
    by_vars = exprs(USUBJID, RSDTC = TRDTC),
    new_vars = exprs(SLD)
  ) %>%
  derive_vars_event(
    dataset_source = rs,
    filter_source = RECIST == "Progressive Disease" | SLD > 1.2 * lag(SLD) + 5,
    by_vars = exprs(USUBJID),
    set_values_to = exprs(PROG_EVENT = TRUE, PROG_DATE = ymd(RSDTC))
  )

# 4. 计算无进展生存期 (PFS)
adpfs <- adrs %>%
  derive_var_extreme_flag(
    by_vars = exprs(USUBJID),
    order = exprs(PROG_DATE),
    new_var = FIRST_PROG_DATE, # 首次进展时间
    mode = "first"
  ) %>%
  derive_vars_duration(
    new_var = exprs(PFS_TIME),
    start_date = exprs(RSDTC),
    end_date = exprs(FIRST_PROG_DATE),
    unit = "days"
  ) %>%
  mutate(
    PFS_CENSOR = if_else(is.na(FIRST_PROG_DATE), 1, 0) # 1 表示删失，0 表示事件发生
  )
```

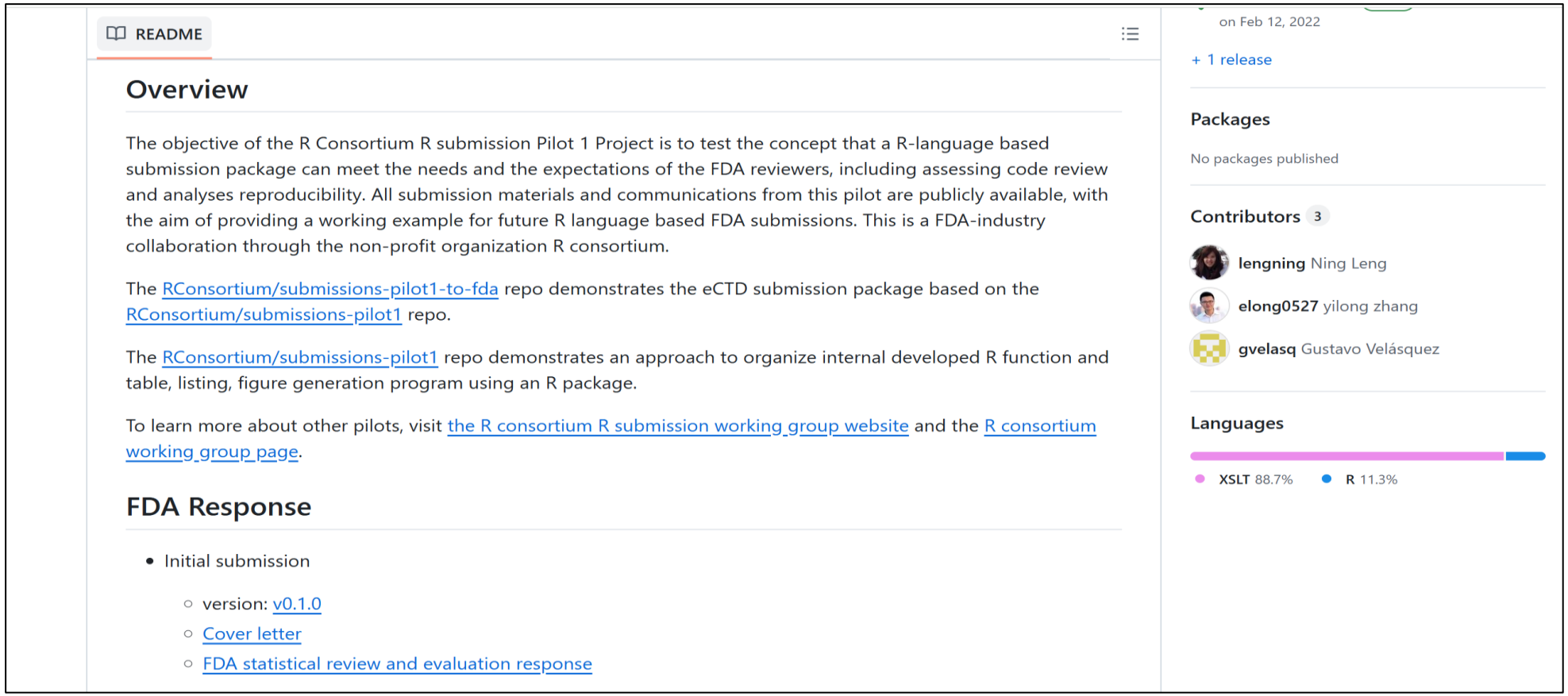
PART
04

R监管递交进展

Regulatory Submission Progress for R

FDA递交 – Pilot 1

R递交工作组最初于2021年11月22日提交Pilot 1通过eCTD网关。2021年12月3日收到了FDA的书面答复函。FDA的回复对两个次要的发现进行了评论，并包括一些最佳实践建议。更新后的提交包解决了所有这些问题，并于2022年2月11日提交。2022年3月14日收到了FDA的最终答复函。



FDA递交 – Pilot 2

旨在验证：使用R语言开发的Shiny应用程序可以被打包到递交材料中并成功传输给FDA审评。该应用程序是基于R递交试点项目1（Pilot 1 Project）中包含的原始数据集和分析构建。

Overview

The objective of the R Consortium R submission Pilot 2 Project is to test the concept that a Shiny application created with the R-language can be bundled into a submission package and transferred successfully to FDA reviewers. The application was built using the source data sets and analyses contained in the R submission Pilot 1 Project, with materials available on the RConsortium/submissions-pilot1 repository, All submission materials and communications from this pilot are publicly available, with the aim of providing a working example for future R language based FDA submissions. This is a FDA-industry collaboration through the non-profit organization R consortium.

While the intent of the project is to enable execution of the Shiny application in a reviewer's local R environment, a deployed version of the application is available in open access through the Shinyapps.io service at rconsortium.shinyapps.io/submissions-pilot2.

The RConsortium/submissions-pilot1-to-fda repository demonstrates the eCTD submission package based on the RConsortium/submissions-pilot1 repo.

The RConsortium/submissions-pilot1 repository demonstrates an approach to organize internal developed R function and table, listing, figure generation program using an R package.

The RConsortium/submissions-pilot2 repository demonstrates an approach to organize a Shiny application as an R package.

To learn more about other pilots, visit the R consortium R submission working group website and the R consortium working group page.

Submission Pilot 2 eCTD v0.10.0 Latest
on Jul 19, 2023

+ 4 releases

Packages

No packages published

Contributors 2

rpodcast Eric Nantz

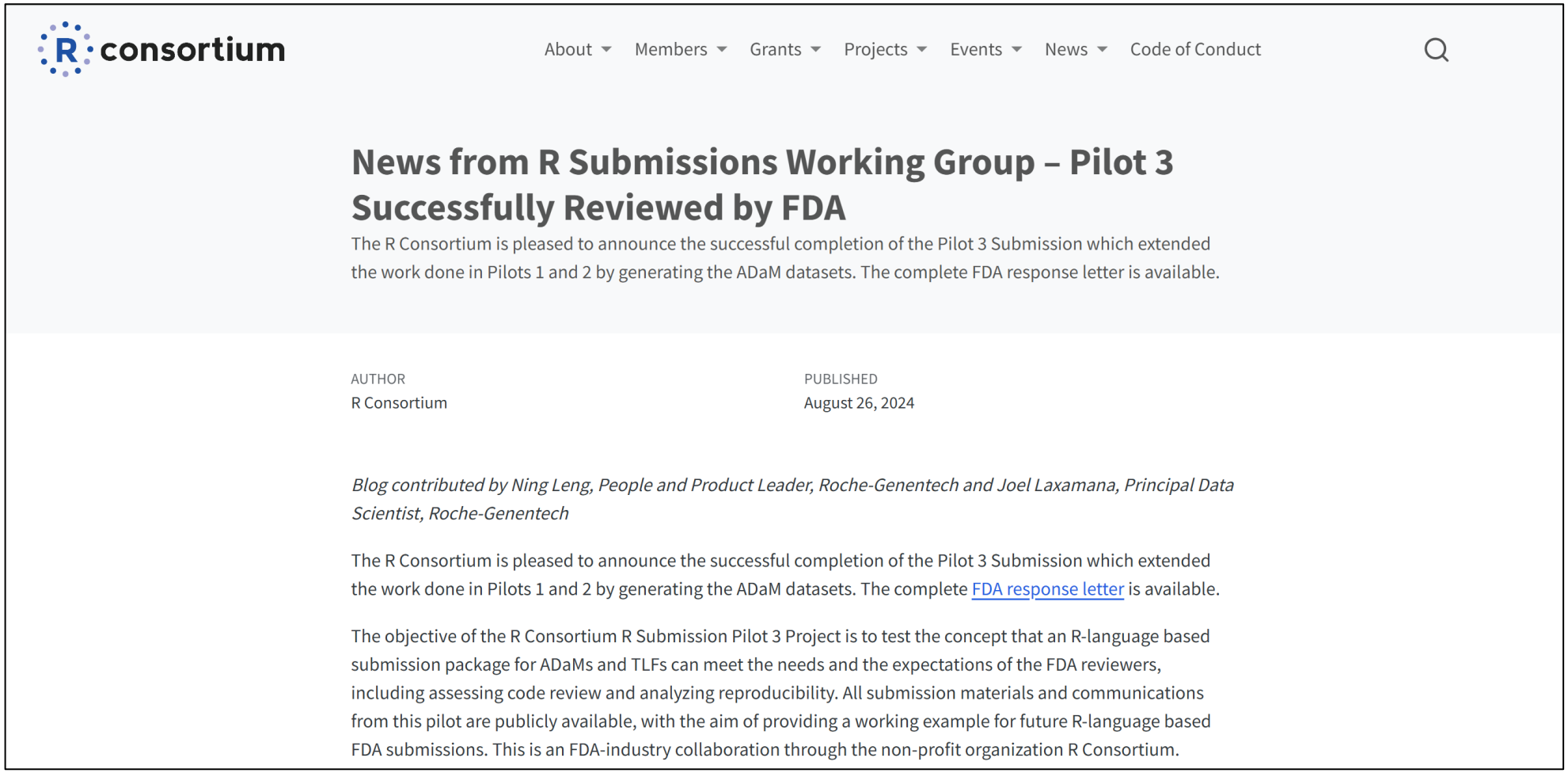
lengning Ning Leng

Languages

XSLT 100.0%

FDA递交 – Pilot 3

首次递交于2023年8月28日通过eCTD网关完成，最终的FDA回复信于2024年8月8日收到。首个公开的遵循eCTD规范的全R提交包示例，其中包括生成ADaM数据集和TFL的R脚本。



R consortium

About Members Grants Projects Events News Code of Conduct

News from R Submissions Working Group – Pilot 3 Successfully Reviewed by FDA

The R Consortium is pleased to announce the successful completion of the Pilot 3 Submission which extended the work done in Pilots 1 and 2 by generating the ADaM datasets. The complete FDA response letter is available.

AUTHOR
R Consortium

PUBLISHED
August 26, 2024

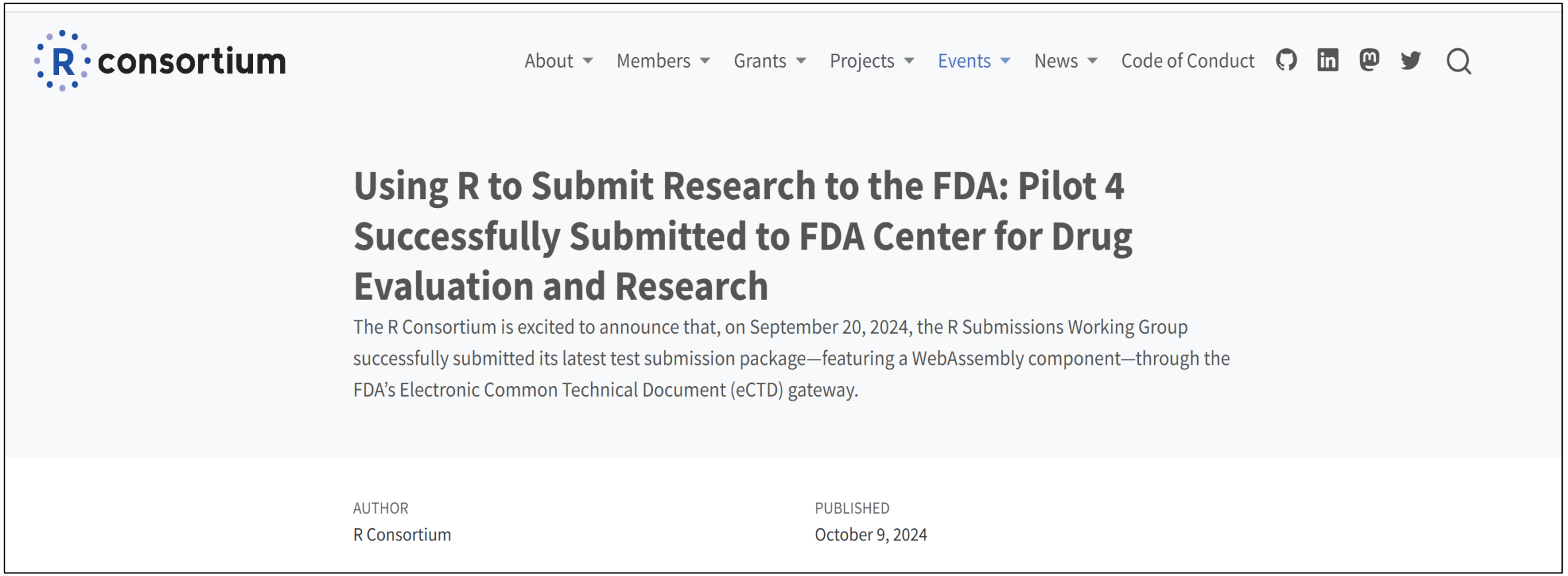
Blog contributed by Ning Leng, People and Product Leader, Roche-Genentech and Joel Laxamana, Principal Data Scientist, Roche-Genentech

The R Consortium is pleased to announce the successful completion of the Pilot 3 Submission which extended the work done in Pilots 1 and 2 by generating the ADaM datasets. The complete FDA response letter is available.

The objective of the R Consortium R Submission Pilot 3 Project is to test the concept that an R-language based submission package for ADaMs and TLFs can meet the needs and the expectations of the FDA reviewers, including assessing code review and analyzing reproducibility. All submission materials and communications from this pilot are publicly available, with the aim of providing a working example for future R-language based FDA submissions. This is an FDA-industry collaboration through the non-profit organization R Consortium.

FDA递交 – Pilot 4

2024年9月20日，R递交工作组（R Submissions Working Group）成功通过FDA的电子通用技术文档（eCTD）网关递交了其最新的测试递交包，其中包含一个WebAssembly组件。



数据驱动

助力产品成功

Drive product success with data-driven approach

联系我们
公司地址：北京市丰台区金泽路161号院1号楼远洋锐中心2911室
商务合作：杨总 13911798010
骆总 15201237869