

Up-to-date with Dates – Imputation Rules in R Compared to SAS

Monika Ludwinska
PHUSE 2023 US Connect
Orlando, USA
5th–8th March 2023

Agenda

- ✓ Introduction
- ✓ Fundamentals of working with timing variables
- ✓ R packages: base, lubridate, hms
- ✓ Practical examples
- ✓ R and SAS comparison
- ✓ Summary



Introduction

- ✓ Timing variables in SDTM and ADaM
- ✓ R functions
- ✓ SAS and R for imputing timing variables



Fundamentals of working with timing variables

- SDTM Implementation Guide requires using the international standard **ISO 8601** format to store timing variables

2023-12-01T12:11:13

- Derived timing variables in ADaM datasets
- Partial dates

Base package

General information: specifying format of an input

- **Datetime variables**

```
> as.POSIXct("2022-09-13T12:40", format="%Y-%m-%dT%H:%M")  
[1] "2022-09-13 12:40:00 EDT"
```

- **Date variables**

```
> as.Date("2021-01-27")  
[1] "2021-01-27"
```

Lubridate package

General information: order of elements in a string, format not needed

- **Datetime variables**

```
> ymd_hm("2022-09-13T12:40")  
[1] "2022-09-13 12:40:00 UTC"
```

- **Date variables**

```
> x <- c("2021/01/27", "2021-02-24")  
> ymd(x)  
[1] "2021-01-27" "2021-02-24"
```

Hms package

General information: extracting time from derived datetime variables

- **Time variables**

```
> as_hms("10:12:00")  
10:12:00  
> class(as_hms("10:12:00"))  
[1] "hms"    "difftime"
```

Clinical trials examples

SDTM data overview: sdtm1 and sdtm2

USUBJID	TRTSDT	CMSTDTC	CMENDTC
0001	2021-09-21	2020-01	
0002	2021-10-06	2018	2021-08
0003	2021-11-16	2021-03-16	2021-03-16
0004	2021-12-14	2021-09	
0005	2022-01-04	2021-10-21	2021-10-21
0006	2022-01-14	2015	
0007	2021-11-04	2016-03	

USUBJID	LBDTC
0001	2021-09-13T16:40
0001	2021-09-21T09:20
0001	2021-09-22T10:00
0001	2021-09-24T08:38
0001	2021-10-05T08:35
0001	2021-10-26T09:07
0001	2021-11-08T09:02

Clinical trials examples

ADaM data overview: adam1 and adam2

USUBJID	TRTSDT	CMSTDTC	CMENDTC	ASTDT	AENDT	ASTDTF	ASTDY
0001	2021-09-21	2020-01		2020-01-01	NA	D	NA
0002	2021-10-06	2018	2021-08	2018-01-01	2021-08-31	M	NA
0003	2021-11-16	2021-03-16	2021-03-16	2021-03-16	2021-03-16		-245
0004	2021-12-14	2021-09		2021-09-01	NA	D	NA
0005	2022-01-04	2021-10-21	2021-10-21	2021-10-21	2021-10-21		-75
0006	2022-01-14	2015		2015-01-01	NA	M	NA
0007	2021-11-04	2016-03		2016-03-01	NA	D	NA

USUBJID	LBDTC	ADTM	ADT	ATM
0001	2021-09-13T16:40	2021-09-13 16:40:00	2021-09-13	16:40:00
0001	2021-09-21T09:20	2021-09-21 09:20:00	2021-09-21	09:20:00
0001	2021-09-22T10:00	2021-09-22 10:00:00	2021-09-22	10:00:00
0001	2021-09-24T08:38	2021-09-24 08:38:00	2021-09-24	08:38:00
0001	2021-10-05T08:35	2021-10-05 08:35:00	2021-10-05	08:35:00
0001	2021-10-26T09:07	2021-10-26 09:07:00	2021-10-26	09:07:00
0001	2021-11-08T09:02	2021-11-08 09:02:00	2021-11-08	09:02:00

Clinical trials examples

SDTM1: Base package

```
phuse1Base = sdtm1 %>%  
  mutate(ASTDT = as.Date(case_when(nchar(CMSTDTC) == 7 ~ paste0(CMSTDTC, "-01"),  
                                nchar(CMSTDTC) == 4 ~ paste0(CMSTDTC, "-01-01"),  
                                TRUE ~ CMSTDTC)),  
         ASTDTF = case_when(nchar(CMSTDTC) == 7 ~ "D",  
                           nchar(CMSTDTC) == 4 ~ "M",  
                           TRUE ~ ""),  
         ASTDY = case_when(ASTDTF == "" & ASTDT >= TRTSDT ~ as.numeric(difftime(ASTDT, TRTSDT, units = "days")+1),  
                           ASTDTF == "" & ASTDT < TRTSDT ~ as.numeric(difftime(ASTDT, TRTSDT, units = "days")),  
                           TRUE ~ as.numeric(NA)),  
         AENDT = case_when(nchar(CMENDTC) == 7 ~ as.Date(paste0(CMENDTC, "-01"), format = "%Y-%m-%d") + months(1) - 1,  
                           nchar(CMENDTC) == 4 ~ as.Date(paste0(CMENDTC, "-12-31"), format = "%Y-%m-%d"),  
                           TRUE ~ as.Date(CMENDTC, format = "%Y-%m-%d")))
```

```
> diffdf(adam1, phuse1Base)  
No issues were found!
```

Clinical trials examples

SDTM1: Lubridate package

```
phuse10ther = sdtm1 %>%
  mutate(ASTDT = ymd(case_when(nchar(CMSTDTC) == 7 ~ paste0(CMSTDTC, "-01"),
                              nchar(CMSTDTC) == 4 ~ paste0(CMSTDTC, "-01-01"),
                              TRUE ~ CMSTDTC)),
         ASTDTF = case_when(nchar(CMSTDTC) == 7 ~ "D",
                              nchar(CMSTDTC) == 4 ~ "M",
                              TRUE ~ ""),
         ASTDY = case_when(ASTDTF == "" & ASTDT >= TRTSDT ~ interval(TRTSDT, ASTDT) / days(1),
                              ASTDTF == "" & ASTDT < TRTSDT ~ interval(TRTSDT, ASTDT) / days(1),
                              TRUE ~ as.numeric(NA)),
         AENDT_ = ymd(case_when(nchar(CMENDTC) == 7 ~ paste0(CMENDTC, "-01"),
                              nchar(CMENDTC) == 4 ~ paste0(CMENDTC, "-12-31"),
                              TRUE ~ CMENDTC)),
         AENDT = case_when(nchar(CMENDTC) == 7 ~ ceiling_date(AENDT_, unit = "month")-1,
                              TRUE ~ AENDT_)) %>%
  select(-AENDT_)
```

```
> diffd(adam1, phuse10ther)
No issues were found!
```

Clinical trials examples

SDTM2: Base package

```
phuse2Base = sdtm2 %>%  
  mutate(ADTM = as.POSIXct(LBDTC, format="%Y-%m-%dT%H:%M", tz="UTC"),  
         ADT = as.Date(LBDTC),  
         ATM = difftime(strptime(LBDTC, "%Y-%m-%dT%H:%M") ,strptime(LBDTC, "%Y-%m-%d") , units="secs"))  
  
class(phuse2Base$ATM) = c("hms", "difftime")
```

```
> diffdf(adam2, phuse2Base)  
No issues were found!
```

Clinical trials examples

SDTM2: Lubridate and Hms packages

```
phuse20ther = sdtm2 %>%  
  mutate(ADTM = ymd_hm(LBDTC),  
         ADT = date(ADTM),  
         ATM = as_hms(ADTM))
```

```
> diffdf(adam2, phuse20ther)  
No issues were found!
```

SAS and R – comparison of functions

Calculating timing variables

Timing variable	SAS function	R function
Datetime	input(*DTC, is8601dt.)	as.POSIXct(*DTC, format="%Y-%m-%dT%H:%M") ymd_hm(*DTC)
Date	datepart(datetime) input(*DTC, is8601da.)	date(datetime) as.Date(*DTC) or date(*DTC)
Time	timepart(datetime) input(scan(*DTC, 2, "T"), is8601tm.)	as_hms(datetime)

Summary

- ✓ Working with dates – inevitable part of work
- ✓ Application of R in validation of timing variables
- ✓ SAS and R – perfect duo
- ✓ Upskilling



THANK YOU

Monika Ludwińska

Intego Group LLC

Grzybowska 78

00-844 Warsaw, Poland

Work Phone: +48 222 500 032 ext. 2583

Email: monika.ludwinska@intego-group.com

Web: www.intego-group.com