playing_with_data

Franky

2023-11-20

1. From LEQ –> dose

1.1 verification of cumulatiive dose calculation (data source: Dosimeter sample raw data and notes.xlsx)

$$Dose = 100 \times \frac{T}{T_c} \times 2^{(L-L_c)/Q}$$

```
Dosimeter.sample <- read_excel("Dosimeter sample raw data and notes.xlsx", sheet = 2)
# criterion here to use:
# Lc <- 85
# Tc <- 8

dose_per <- function(Time, Leq, Lc = 85, Tc = 8, Q = 3){
  return(100*(Time/Tc)*2^((Leq - Lc)/Q))
}
# verify the calculation of sample
SampleDose.cum <- 0
time.seq <- Dosimeter.sample$`Time H`
leq.seq  <- Dosimeter.sample$`LEQ dB`
for(i in 1:(nrow(Dosimeter.sample)-1)){
  SampleDose.cum = SampleDose.cum + dose_per(Time = (time.seq[i+1] - time.seq[i]), Leq = leq.seq[i+1])
}
print(SampleDose.cum)
```

```
## [1] 16.96131
```

```
rm(time.seq, leq.seq)
```

1.2 verification of daily dose calculation (data source: DosimetryData_Example.xlsx$OriginalOutput)

```
OriginalOutput <- read_excel("DosimetryData_Example.xlsx", sheet = 2)
hours.map <- OriginalOutput[c("Day1_totalhrs", "Day2_totalhrs", "Day3_totalhrs", "Day4_totalhrs",
                              "Day5_totalhrs", "Day6_totalhrs", "Day7_totalhrs", "Day8_totalhrs")]
lep.map   <- OriginalOutput[c("Day1_LEQ", "Day2_LEQ", "Day3_LEQ", "Day4_LEQ",
                              "Day5_LEQ", "Day6_LEQ", "Day7_LEQ", "Day8_LEQ")]
dailydose.check <- OriginalOutput[c("Day1_Dose", "Day2_Dose", "Day3_Dose", "Day4_Dose",
                                    "Day5_Dose", "Day6_Dose", "Day7_Dose", "Day8_Dose")]
columns <- c()
for (i in 1:ncol(hours.map)) {columns[i]  = paste("Day", i, "_Dose", sep = "")}

dailydose.cal <- data.frame(matrix(ncol = ncol(hours.map), nrow = nrow(hours.map)))
colnames(dailydose.cal) <- columns
```

```
for(i in 1:nrow(hours.map)){
  for(j in 1:ncol(hours.map)){
    dailydose.cal[i, j] = dose_per(Time = hours.map[i, j], Leq = lep.map[i, j])
  }
}
# round(dailydose.cal, 2)
# dailydose.cal
# dailydose.check
```

Conclusion:

Until now, this step is good, which means the calculation way from LEQ(any time interval) to 8-hour-criterion dose is reliable.

2. From 3.75 minutes Leq –> daily Leq (data source: DosimetryData_Example.xlsx$LA_matri & $DayOfWeek)

2.1 Data loading and processing

```
# load .xlsx file and identify na with "NaN"
LA_matrix <- read_excel("DosimetryData_Example.xlsx", sheet = 4, col_names = TRUE, na = "NaN")
```

```
## New names:
## * `MC16_17, 11-18-16` -> `MC16_17, 11-18-16...67`
## * `MC16_17, 11-18-16` -> `MC16_17, 11-18-16...68`
```

```
# New names:
# • `MC16_17, 11-18-16` -> `MC16_17, 11-18-16...67`
# • `MC16_17, 11-18-16` -> `MC16_17, 11-18-16...68`
DayOfWeek <- read_excel("DosimetryData_Example.xlsx", sheet = 6, col_names = TRUE, na = "NaN")
```

```
## New names:
## * `MC16_17, 11-18-16` -> `MC16_17, 11-18-16...67`
## * `MC16_17, 11-18-16` -> `MC16_17, 11-18-16...68`
```

```
# New names:
# • `MC16_17, 11-18-16` -> `MC16_17, 11-18-16...67`
# • `MC16_17, 11-18-16` -> `MC16_17, 11-18-16...68`

# indicting that contains duplicated names
LA_matrix <- LA_matrix %>% select(-c(`MC16_17, 11-18-16...68`))
colnames(LA_matrix)[colnames(LA_matrix) == "MC16_17, 11-18-16...67"] <- "MC16_17, 11-18-16"
DayOfWeek <- DayOfWeek %>% select(-c(`MC16_17, 11-18-16...68`))
colnames(DayOfWeek)[colnames(DayOfWeek) == "MC16_17, 11-18-16...67"] <- "MC16_17, 11-18-16"

Week.Index <- data.frame(matrix(ncol = ncol(DayOfWeek), nrow = nrow(DayOfWeek)))
colnames(Week.Index) = colnames(DayOfWeek)

for(j in 1:ncol(DayOfWeek)){
  week_index = 1
  Week.Index[1, j] = 1
  for(i in 2:nrow(DayOfWeek)){
    week_index = ifelse(DayOfWeek[i-1,j] - DayOfWeek[i,j] > 1, week_index+1,  week_index) # identify th
    Week.Index[i, j] = week_index
  }
}
```

```r
# combine the Week.Index and DayOfWeek
MWeek.Index <- as.matrix(Week.Index)
MDayOfWeek  <- as.matrix(DayOfWeek)
Day.Index <- matrix(paste(MWeek.Index, MDayOfWeek, sep="-"), nrow=nrow(MWeek.Index),
                    dimnames=dimnames(MWeek.Index)) %>% as.data.frame()
# Day.Index$CT16_01[549] # looks good

# table(Day.Index$CT16_01)
#   1-6   1-7   2-1   2-2   2-3   2-4   2-5   2-6  2-NA NA-NA
#   164   384   384   384   384   384   384   208     1    11
# 164*0.0625 # =10.25
# which is the same as $OriginalOutput

# create list to store every student's data
LA.Seq <- list()
for (i in 1:length(colnames(LA_matrix))) {
  # "CT16_01"
  # "MC16_17, 11-18-16"
  LA.Seq[[i]] <- data.frame(Day.Index[colnames(LA_matrix)[i]],
                            LA_matrix[colnames(LA_matrix)[i]])
  colnames(LA.Seq[[i]]) <- c("Week-Day", "Leq(3.75min)")
}
# rename the list
names(LA.Seq) <- colnames(LA_matrix)
LA.names <- names(LA.Seq)

# remove NA records
for(key in LA.names){
  LA.Seq[[key]] = LA.Seq[[key]] %>% filter(!grepl("NA", `Leq(3.75min)`))
  LA.Seq[[key]] = LA.Seq[[key]] %>% filter(!grepl("NA", `Week-Day`))
}

# remove terms that no longer use
rm(Day.Index, MDayOfWeek, MWeek.Index, Week.Index, DayOfWeek, LA_matrix)

# table(LA.Seq[["CT16_01"]]["Week-Day"]) * 0.0625
#   1-6   1-7   2-1   2-2   2-3   2-4   2-5   2-6
# 10.25 24.00 24.00 24.00 24.00 24.00 24.00 13.00
# table(LA.Seq[["CT16_02, 4-15-16"]]["Week-Day"]) * 0.0625
# Week-Day
#     1-5     1-6     1-7     2-1     2-2     2-3     2-4     2-5
# 10.6875 24.0000 24.0000 24.0000 24.0000 24.0000 24.0000 13.3125

# perfect with OriginalOutput
```

2.2 Leq for each day calculation

$$L_{Aeq,T} = 10 * \log_{10}(\frac{1}{T_{duration}} * \sum_{i=1}^{n} t_i * 10^{\frac{L_i}{10}})$$

```r
DayLeq <- list()
for(key in LA.names){
  # key = "CT16_01"
```

```
  df <- LA.Seq[[key]]
  for(day in unique(df$`Week-Day`)){
    # day = "1-6"
    sub_df <- df %>% filter(`Week-Day` == day)
    exp_sum <- 0
    Time_dur <- nrow(sub_df) * 0.0625
    ti <- 0.0625
    exp_sum <- 0
    for(i in 1:nrow(sub_df)){
      exp_sum <- exp_sum + (1/Time_dur)*ti*10^(sub_df[i, 2]/10)
    }
    DayLeq[[key]][day] <- 10 * log(exp_sum, base = 10)
  }
}
LA.Seq[["CT16_01"]] %>% filter(`Week-Day` == "1-6")
```

```
##     Week-Day Leq(3.75min)
## 1        1-6       94.619
## 2        1-6       84.490
## 3        1-6       84.371
## 4        1-6       70.262
## 5        1-6        0.000
## 6        1-6        0.000
## 7        1-6       58.000
## 8        1-6        0.000
## 9        1-6        0.000
## 10       1-6       58.966
## 11       1-6       53.755
## 12       1-6       49.000
## 13       1-6       56.755
## 14       1-6        0.000
## 15       1-6        0.000
## 16       1-6        0.000
## 17       1-6       52.000
## 18       1-6       55.000
## 19       1-6       53.755
## 20       1-6        0.000
## 21       1-6        0.000
## 22       1-6        0.000
## 23       1-6        0.000
## 24       1-6        0.000
## 25       1-6       66.648
## 26       1-6        0.000
## 27       1-6        0.000
## 28       1-6        0.000
## 29       1-6       70.198
## 30       1-6       64.000
## 31       1-6        0.000
## 32       1-6        0.000
## 33       1-6       52.000
## 34       1-6       52.000
## 35       1-6        0.000
## 36       1-6       59.378
## 37       1-6        0.000
```

```
## 38       1-6        0.000
## 39       1-6        0.000
## 40       1-6        0.000
## 41       1-6        0.000
## 42       1-6        0.000
## 43       1-6        0.000
## 44       1-6        0.000
## 45       1-6        0.000
## 46       1-6        0.000
## 47       1-6        0.000
## 48       1-6       55.000
## 49       1-6        0.000
## 50       1-6        0.000
## 51       1-6        0.000
## 52       1-6        0.000
## 53       1-6        0.000
## 54       1-6        0.000
## 55       1-6        0.000
## 56       1-6        0.000
## 57       1-6        0.000
## 58       1-6        0.000
## 59       1-6        0.000
## 60       1-6        0.000
## 61       1-6       67.911
## 62       1-6       77.059
## 63       1-6       49.000
## 64       1-6        0.000
## 65       1-6        0.000
## 66       1-6       58.000
## 67       1-6       78.480
## 68       1-6       90.340
## 69       1-6       83.776
## 70       1-6       86.608
## 71       1-6       88.749
## 72       1-6       87.057
## 73       1-6       83.959
## 74       1-6       84.687
## 75       1-6       84.250
## 76       1-6       86.821
## 77       1-6       87.964
## 78       1-6       88.284
## 79       1-6       86.993
## 80       1-6       86.266
## 81       1-6       87.424
## 82       1-6       87.558
## 83       1-6       90.116
## 84       1-6       85.821
## 85       1-6       83.354
## 86       1-6       87.553
## 87       1-6       91.900
## 88       1-6       87.057
## 89       1-6       86.585
## 90       1-6       92.789
## 91       1-6       85.904
```

```
## 92       1-6       83.384
## 93       1-6       82.577
## 94       1-6       84.694
## 95       1-6       82.042
## 96       1-6       86.329
## 97       1-6       84.489
## 98       1-6       87.687
## 99       1-6       85.517
## 100      1-6       84.712
## 101      1-6       79.540
## 102      1-6       79.993
## 103      1-6       70.000
## 104      1-6       61.000
## 105      1-6        0.000
## 106      1-6        0.000
## 107      1-6       58.000
## 108      1-6        0.000
## 109      1-6       67.510
## 110      1-6        0.000
## 111      1-6        0.000
## 112      1-6        0.000
## 113      1-6        0.000
## 114      1-6        0.000
## 115      1-6       58.000
## 116      1-6        0.000
## 117      1-6        0.000
## 118      1-6        0.000
## 119      1-6        0.000
## 120      1-6        0.000
## 121      1-6        0.000
## 122      1-6        0.000
## 123      1-6        0.000
## 124      1-6        0.000
## 125      1-6        0.000
## 126      1-6        0.000
## 127      1-6        0.000
## 128      1-6        0.000
## 129      1-6        0.000
## 130      1-6        0.000
## 131      1-6        0.000
## 132      1-6        0.000
## 133      1-6        0.000
## 134      1-6        0.000
## 135      1-6        0.000
## 136      1-6        0.000
## 137      1-6        0.000
## 138      1-6        0.000
## 139      1-6        0.000
## 140      1-6        0.000
## 141      1-6        0.000
## 142      1-6        0.000
## 143      1-6        0.000
## 144      1-6        0.000
## 145      1-6        0.000
```

```
## 146      1-6       72.721
## 147      1-6       72.574
## 148      1-6        0.000
## 149      1-6        0.000
## 150      1-6        0.000
## 151      1-6        0.000
## 152      1-6        0.000
## 153      1-6        0.000
## 154      1-6        0.000
## 155      1-6        0.000
## 156      1-6        0.000
## 157      1-6        0.000
## 158      1-6        0.000
## 159      1-6       61.000
## 160      1-6       67.000
## 161      1-6        0.000
## 162      1-6       58.000
## 163      1-6        0.000
## 164      1-6        0.000
```

```
DayLeq[["CT16_01"]]
```

```
##      1-6      1-7      2-1      2-2      2-3      2-4      2-5      2-6
## 81.14258 83.87709 81.58470 78.19862 76.39172 77.93267 75.95544 80.40990
```

1.3 compare the daily dose via "Daily Leq" and "Cumulative (3.75 min) Leq"

```
# For each student, calculate cumulative dose for every single day
cumdose.cal <- list()
for(key in LA.names){
  # cat("key ready")
  df <- LA.Seq[[key]]
  days <- unique(LA.Seq[[key]]$`Week-Day`)
  index = 1
  for(day in days){
    # cat("day ready\n")
    sub_df <- df %>% filter(`Week-Day` == day)
    # cat("sub_df ready\n")
    Dose.cum <- c()
    Dose.cum[1] = 0
    for(i in 1:(nrow(sub_df)-1)){
      Dose.cum[i+1] = Dose.cum[i] + dose_per(Time = 0.0625, Leq = sub_df$`Leq(3.75min)`[i+1])
    }
    cumdose.cal[[key]][index] = tail(Dose.cum, 1)
    index = index + 1
  }
}
# cumdose.cal[["CT16_01"]]

cumdose.cal.df <- data.frame(matrix(ncol = ncol(dailydose.cal), nrow = nrow(dailydose.cal) - 1))
colnames(cumdose.cal.df) <- colnames(dailydose.cal)
for(i in 1:length(LA.names)){
  vec <- cumdose.cal[[i]]
  for(j in 1:length(vec)){
```

```
    cumdose.cal.df[i, j] = vec[j]
  }
}

dose.cal.diff <- dailydose.cal[-67, ] - cumdose.cal.df
print(round(dose.cal.diff, 2)[sample(1:79, 10), ])
```

```
##    Day1_Dose Day2_Dose Day3_Dose Day4_Dose Day5_Dose Day6_Dose Day7_Dose
## 14     -0.03     -0.07     -1.26     -0.02     -0.01     -3.04      0.16
## 24     -0.14     -0.28     -0.81     -0.02     -0.02     -0.01        NA
## 78     -0.73     -1.87    -34.20    -12.06     -0.13      0.16     -0.48
## 75     -0.18     -0.19    -26.72     -0.14     -0.25     -4.59     -2.69
## 49     -0.02      0.00     -0.10     -0.04     -0.10     -0.06     -0.02
## 66     -1.94     -1.87     -1.15     -0.93    -21.87    -17.26     -0.02
## 69     -0.01     -0.05     -0.03    -15.52     -1.36     -0.08     -0.02
## 72      0.00     -0.38     -1.51     -0.14     -0.11    -17.96     -0.02
## 12     -0.12     -0.27     -0.10     -0.11     -0.05     -0.25     -0.27
## 58     -5.60     -7.16     -0.02      0.10        NA        NA        NA
##    Day8_Dose
## 14      0.00
## 24        NA
## 78      0.00
## 75      0.09
## 49      0.00
## 66      0.00
## 69     -0.02
## 72        NA
## 12     -0.01
## 58        NA
```

Conclusion

The maximum difference is around 20% dose for each day. Majority of them are $< 1\%$. Using dayily Leq to estimate rather then calculate cumulations is reliable on the whole.