Mortality descriptive

Esther Chung

March 6, 2018

Goal: Identify and explore mortality variables in the GHAP nutritional intervention trials.

# According to kikm\_variables, the following nutritional intervention studies have AGEDTH as mortality variables   
 # DIVIDS, JIVITA3 and 4, iDOSE, iDYAD, ZVITAMBO, TanzaniaChild2, PROVIDE  
  
 d$AGEDTH <- as.numeric(d$AGEDTH)  
  
# Pull out data from each study  
 divids <- d %>% filter(STUDYID=="ki1000301-DIVIDS")  
 jvit3 <- d %>% filter(STUDYID=="kiGH5241-JiVitA-3")  
 jvit4 <- d %>% filter(STUDYID=="kiGH5241-JiVitA-4")  
 idose <- d %>% filter(STUDYID=="ki1148112-iLiNS-DOSE")  
 idyadm <- d %>% filter(STUDYID=="ki1148112-iLiNS-DYAD-M")  
 idyadg <- d %>% filter(STUDYID=="ki1033518-iLiNS-DYAD-G")  
 zvit <- d %>% filter(STUDYID=="ki1126311-ZVITAMBO")  
 tanzchild <- d %>% filter(STUDYID=="ki1066203-TanzaniaChild2")  
 provide <- d %>% filter(STUDYID=="ki1017093b-PROVIDE")

DIVIDS

# Make sure data arranged by subject id and age days  
 divids <- divids %>% arrange(SUBJID, AGEDAYS)  
  
# Check if age of death is recorded multiple times for each child who died  
 divids %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 16 x 2  
## SUBJID N  
## <dbl> <int>  
## 1 101 1  
## 2 124 12  
## 3 143 4  
## 4 259 3  
## 5 260 14  
## 6 331 7  
## 7 401 6  
## 8 487 3  
## 9 529 3  
## 10 714 3  
## 11 801 2  
## 12 1098 1  
## 13 1208 4  
## 14 1418 1  
## 15 1467 1  
## 16 1666 7

# Yes, recorded multiple times per subject  
  
# Create variable with maximum age per child  
 divids <- divids %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
# create indicator for last observation for a child using age   
 divids <- divids %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge)  
   
# Create flag for the last time point for each subject if the child died   
 divids$death <- NA  
 divids <- divids %>% group\_by(SUBJID) %>% mutate(death = any(!is.na(AGEDTH)) & last==T)   
   
# Count total number of reported deaths using last age of the child  
 table(divids$death==T) # 16 deaths

##   
## FALSE TRUE   
## 18976 16

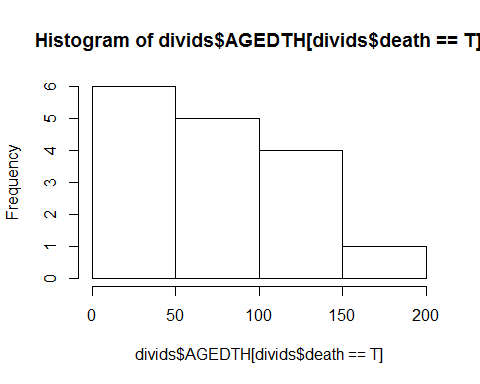
table(divids$DEAD[divids$death==T]) # 16 deaths

##   
## 1   
## 16

summary(divids$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 14.00 74.25 114.00 110.33 142.00 180.00 18920

hist(divids$AGEDTH[divids$death==T])



# Check that the age of death is later than the oldest measurement on each child  
 divids <- divids %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(divids$deathErrorFlag[!is.na(divids$AGEDTH)]) # 0

##   
## FALSE   
## 72

#0 children have measurements after their recorded age of death  
   
# subset data to children with death error flags  
 df\_error <- divids %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 0 x 6  
## # Groups: SUBJID [0]  
## # ... with 6 variables: SUBJID <dbl>, last <lgl>, AGEDTH <dbl>,  
## # AGEDAYS <int>, maxAge <dbl>, deathErrorFlag <lgl>

#make indicator for any death (died)  
 divids <- divids %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>%  
 arrange(SUBJID, AGEDAYS)  
   
#subset data set to 1) children who died AND 2) do not have death error flags  
 df\_death <- divids %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died, deathErrorFlag))  
   
 head(df\_death, 20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [4]  
## SUBJID AGEDAYS WHZ AGEDTH died deathErrorFlag  
## <dbl> <int> <dbl> <dbl> <lgl> <lgl>   
## 1 101 1 - 3.48 22.0 T F   
## 2 124 1 - 2.46 180 T F   
## 3 124 30 NA 180 T F   
## 4 124 72 NA 180 T F   
## 5 124 74 - 5.60 180 T F   
## 6 124 86 NA 180 T F   
## 7 124 100 NA 180 T F   
## 8 124 105 - 5.35 180 T F   
## 9 124 126 - 5.85 180 T F   
## 10 124 128 NA 180 T F   
## 11 124 156 NA 180 T F   
## 12 124 158 - 3.35 180 T F   
## 13 124 179 - 5.21 180 T F   
## 14 143 1 - 4.10 78.0 T F   
## 15 143 43 NA 78.0 T F   
## 16 143 71 NA 78.0 T F   
## 17 143 75 - 5.20 78.0 T F   
## 18 259 1 - 1.47 60.0 T F   
## 19 259 44 - 3.32 60.0 T F   
## 20 259 58 NA 60.0 T F

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

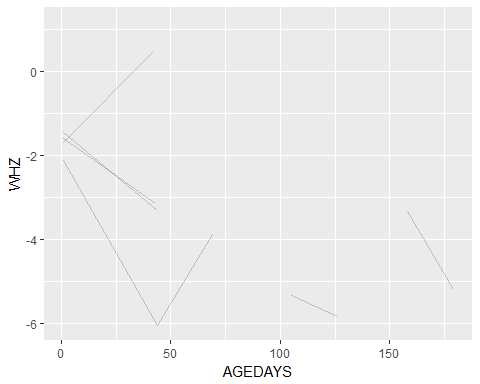
## [1] -3.48 -2.46 NA NA -5.60 NA NA -5.35 -5.85 NA NA  
## [12] -3.35 -5.21 -4.10 NA NA -5.20 -1.47 -3.32 NA -1.35 NA  
## [23] NA NA -4.27 NA NA -4.48 NA -4.98 NA NA -4.69  
## [34] NA NA -2.09 NA NA NA -3.31 NA NA 0.66 NA  
## [45] NA -0.65 NA NA -0.68 NA NA NA 1.16 -0.10 NA  
## [56] NA -1.61 -3.14 NA -2.13 -6.06 -3.90 NA NA -0.79 -1.70  
## [67] 0.48 NA -1.58 NA NA -1.36

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 33 missing WHZ

##   
## FALSE TRUE   
## 39 33

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

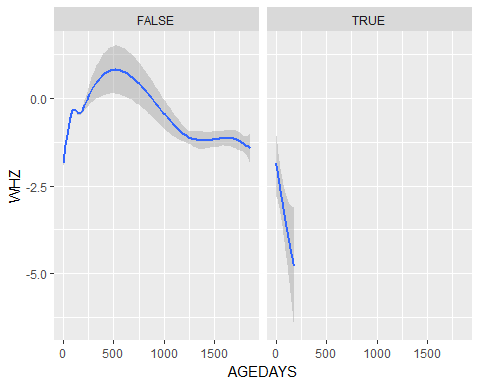
## Warning: Removed 13 rows containing missing values (geom\_path).



# warning indicates that 13 rows removed b/c had missing values   
  
#Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(divids) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) + facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 8137 rows containing non-finite values (stat\_smooth).



# warning indicates that 8137 rows contained non-finite values

JiVitA-3

# Make sure data arranged by subject id and age days  
 jvit3 <- jvit3 %>% arrange(SUBJID, AGEDAYS)  
  
# Check if age of death is recorded multiple times for each child who died  
 jvit3 %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 1,226 x 2  
## SUBJID N  
## <dbl> <int>  
## 1 21.0 1  
## 2 26.0 2  
## 3 48.0 2  
## 4 59.0 2  
## 5 61.0 2  
## 6 73.0 3  
## 7 211 2  
## 8 232 2  
## 9 245 2  
## 10 250 3  
## 11 251 3  
## 12 271 3  
## 13 275 3  
## 14 291 3  
## 15 296 2  
## 16 297 1  
## 17 329 1  
## 18 365 2  
## 19 369 3  
## 20 389 6  
## 21 426 2  
## 22 433 4  
## 23 471 4  
## 24 504 5  
## 25 506 2  
## 26 512 2  
## 27 537 2  
## 28 581 1  
## 29 586 2  
## 30 605 5  
## 31 608 4  
## 32 643 1  
## 33 682 2  
## 34 686 2  
## 35 695 7  
## 36 739 6  
## 37 759 3  
## 38 791 9  
## 39 792 5  
## 40 815 4  
## # ... with 1,186 more rows

# Recorded multiple times per subject  
   
# Create variable with maximum age per child  
 jvit3 <- jvit3 %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
# create indicator for last observation for a child using age   
 jvit3 <- jvit3 %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge)  
  
# Create flag for the last time point for each subject if the child died   
 jvit3$death <- NA  
 jvit3 <- jvit3 %>% group\_by(SUBJID) %>% mutate(death = any(!is.na(AGEDTH)) & last==T)   
# Count total number of reported deaths using last age of the child  
 table(jvit3$death==T) # 1226 deaths

##   
## FALSE TRUE   
## 234470 1226

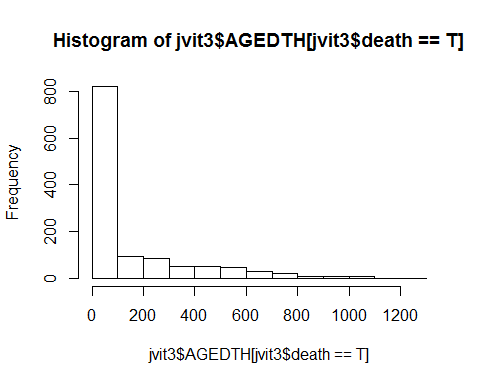
table(jvit3$DEAD[jvit3$death==T]) # 0 deaths

## < table of extent 0 >

# DEAD and AGEDTH variables don't match  
   
 summary(jvit3$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1.0 19.0 176.0 257.6 421.0 1226.0 231167

hist(jvit3$AGEDTH[jvit3$death==T])



# Check that the age of death is later than the oldest measurement on each child  
 jvit3 <- jvit3 %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(jvit3$deathErrorFlag[!is.na(jvit3$AGEDTH)]) # 331

##   
## FALSE TRUE   
## 4198 331

#ISSUE- 331 children have measurements after their recorded age of death  
   
# subset data to children with death error flags  
 df\_error <- jvit3 %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [20]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <dbl> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 73.0 T 7.00 27 27.0 T   
## 2 211 T 14.0 19 19.0 T   
## 3 245 T 3.00 9 9.00 T   
## 4 250 T 5.00 13 13.0 T   
## 5 251 T 4.00 35 35.0 T   
## 6 271 T 36.0 166 166 T   
## 7 275 T 3.00 23 23.0 T   
## 8 329 T 2.00 46 46.0 T   
## 9 369 T 7.00 32 32.0 T   
## 10 426 T 2.00 144 144 T   
## 11 759 T 8.00 108 108 T   
## 12 852 T 2.00 21 21.0 T   
## 13 948 T 4.00 11 11.0 T   
## 14 1053 T 6.00 25 25.0 T   
## 15 1064 T 1.00 40 40.0 T   
## 16 1288 T 3.00 28 28.0 T   
## 17 1423 T 19.0 32 32.0 T   
## 18 1648 T 4.00 12 12.0 T   
## 19 1769 T 3.00 14 14.0 T   
## 20 1890 T 3.00 134 134 T

#make indicator for any death (died)  
 jvit3 <- jvit3 %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>%  
 arrange(SUBJID, AGEDAYS)  
   
#subset data set to 1) children who died AND 2) do not have death error flags  
 df\_death <- jvit3 %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died, deathErrorFlag))  
   
 head(df\_death, 20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [12]  
## SUBJID AGEDAYS WHZ AGEDTH died deathErrorFlag  
## <dbl> <int> <dbl> <dbl> <lgl> <lgl>   
## 1 21.0 1 NA 2.00 T F   
## 2 26.0 1 NA 22.0 T F   
## 3 26.0 22 NA 22.0 T F   
## 4 48.0 1 NA 10.0 T F   
## 5 48.0 10 NA 10.0 T F   
## 6 59.0 1 NA 3.00 T F   
## 7 59.0 3 NA 3.00 T F   
## 8 61.0 1 NA 31.0 T F   
## 9 61.0 31 NA 31.0 T F   
## 10 73.0 1 - 0.440 7.00 T F   
## 11 73.0 7 NA 7.00 T F   
## 12 211 14 NA 14.0 T F   
## 13 232 2 NA 26.0 T F   
## 14 232 26 NA 26.0 T F   
## 15 245 1 NA 3.00 T F   
## 16 250 1 - 2.48 5.00 T F   
## 17 250 5 NA 5.00 T F   
## 18 251 1 NA 4.00 T F   
## 19 251 4 NA 4.00 T F   
## 20 271 1 - 1.98 36.0 T F

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

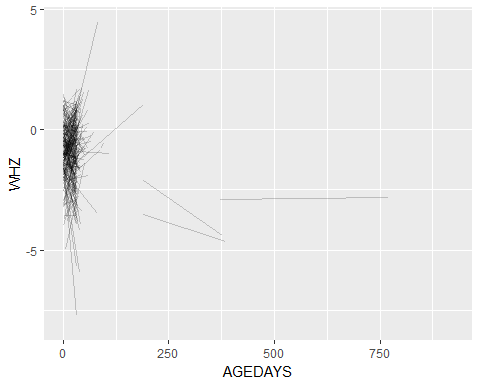
## [1] NA NA NA NA NA NA NA NA NA -0.44 NA  
## [12] NA NA NA NA -2.48 NA NA NA -1.98 NA NA  
## [23] NA -1.10 4.45 NA NA NA NA NA NA NA NA  
## [34] NA -0.22 NA 1.63 NA NA NA -1.08 -2.00 NA NA  
## [45] -0.41 -0.98 NA NA -0.15 0.07 NA NA NA NA NA  
## [56] NA NA NA NA NA -1.91 NA -1.61 NA -3.44 NA  
## [67] NA -0.99 -1.20 NA NA NA NA NA NA NA -1.73  
## [78] -0.13 NA -0.04 NA 0.19 NA NA -1.32 NA NA NA  
## [89] NA NA NA -0.96 -2.59 NA -1.60 NA NA -2.08 NA  
## [100] NA NA -1.63 NA -2.71 NA -0.64 -1.40 NA NA NA  
## [111] NA NA NA NA NA -1.81 NA NA NA -1.15 -1.49  
## [122] NA NA -2.29 NA NA -1.78 NA NA NA -0.75 0.13  
## [133] NA -0.89 NA -1.14 NA NA -0.70 NA -0.24 -1.12 NA  
## [144] NA NA 1.37 NA -0.47 NA -3.18 NA -2.54 NA -1.02  
## [155] -0.82 NA NA NA NA NA -0.83 NA -1.24 NA NA  
## [166] -1.60 -0.82 NA -0.29 NA NA NA NA NA NA 0.27  
## [177] NA NA -0.11 NA -2.87 NA -1.85 NA NA -2.75 NA  
## [188] NA -2.16 NA -2.02 NA NA NA NA NA -5.29 NA  
## [199] NA NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 1471 missing WHZ

##   
## FALSE TRUE   
## 2727 1471

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

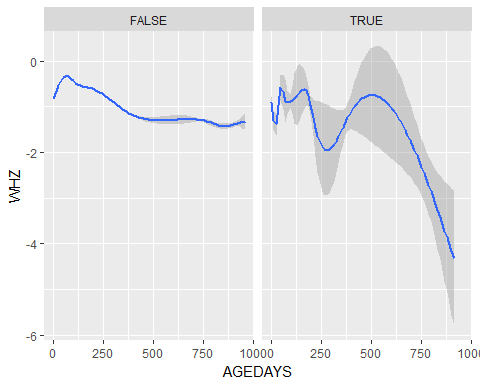
## Warning: Removed 1105 rows containing missing values (geom\_path).



# warning indicates that 1105 rows had missing values   
  
#Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(jvit3) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) + facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 120153 rows containing non-finite values (stat\_smooth).



# warning indicates that 120153 rows contained non-finite values

JiVitA-4

# Make sure data arranged by subject id and age days  
 jvit4 <- jvit4 %>% arrange(SUBJID, AGEDAYS)  
  
# Check if age of death is recorded multiple times for each child who died  
 jvit4 %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 50 x 2  
## SUBJID N  
## <dbl> <int>  
## 1 23.0 90  
## 2 263 333  
## 3 323 34  
## 4 704 163  
## 5 823 373  
## 6 849 370  
## 7 861 153  
## 8 902 268  
## 9 922 88  
## 10 1021 237  
## 11 1282 377  
## 12 1474 338  
## 13 1591 123  
## 14 1718 226  
## 15 1768 69  
## 16 1835 64  
## 17 1843 335  
## 18 1864 140  
## 19 1872 323  
## 20 1934 314  
## 21 2220 291  
## 22 2303 346  
## 23 2363 147  
## 24 2395 311  
## 25 2501 112  
## 26 2619 67  
## 27 2630 376  
## 28 2774 285  
## 29 2862 61  
## 30 2924 340  
## 31 2946 110  
## 32 3069 118  
## 33 3327 242  
## 34 3349 294  
## 35 3421 345  
## 36 3921 162  
## 37 4020 155  
## 38 4245 68  
## 39 4377 315  
## 40 4483 256  
## # ... with 10 more rows

# Recorded multiple times per subject  
   
# Create variable with maximum age per child  
 jvit4 <- jvit4 %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
# create indicator for last observation for a child using age   
 jvit4 <- jvit4 %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge, TRUE, FALSE)  
  
# Create flag for the last time point for each subject if the child died   
 jvit4$death <- NA  
 jvit4 <- jvit4 %>% group\_by(SUBJID) %>% mutate(death = any(!is.na(AGEDTH)) & last==T)   
   
# Count total number of reported deaths using last age of the child  
 table(jvit4$death==T) # 50 deaths

##   
## FALSE TRUE   
## 1883767 50

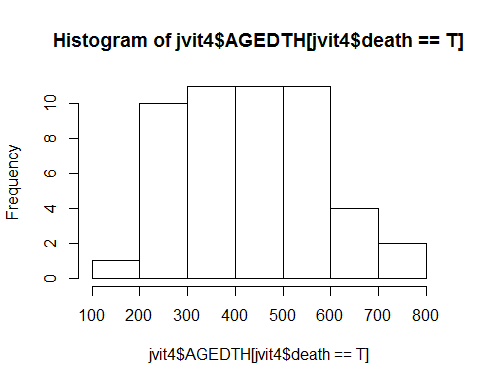
table(jvit4$DEAD[jvit4$death==T]) # 50 deaths

##   
## 1   
## 50

summary(jvit4$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 187.0 452.0 514.0 501.4 590.0 735.0 1872890

hist(jvit4$AGEDTH[jvit4$death==T])



# Check that the age of death is later than the oldest measurement on each child  
 jvit4 <- jvit4 %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(jvit4$deathErrorFlag[!is.na(jvit4$AGEDTH)]) # 1

##   
## FALSE TRUE   
## 10926 1

# ISSUE: 1 person has measurements after recorded age of death  
   
# subset data to children with death error flags  
 df\_error <- jvit4 %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 1 x 6  
## # Groups: SUBJID [1]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <dbl> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 3069 T 296 297 297 T

#make indicator for any death (died)  
 jvit4 <- jvit4 %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>%  
 arrange(SUBJID, AGEDAYS)  
   
#subset data set to 1) children who died AND 2) do not have death error flags  
 df\_death <- jvit4 %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died))  
   
 head(df\_death, 20)

## # A tibble: 20 x 5  
## # Groups: SUBJID [1]  
## SUBJID AGEDAYS WHZ AGEDTH died   
## <dbl> <int> <dbl> <dbl> <lgl>  
## 1 23.0 177 - 0.660 287 T   
## 2 23.0 196 NA 287 T   
## 3 23.0 197 NA 287 T   
## 4 23.0 198 NA 287 T   
## 5 23.0 199 NA 287 T   
## 6 23.0 200 NA 287 T   
## 7 23.0 201 NA 287 T   
## 8 23.0 202 NA 287 T   
## 9 23.0 203 NA 287 T   
## 10 23.0 204 NA 287 T   
## 11 23.0 205 NA 287 T   
## 12 23.0 206 NA 287 T   
## 13 23.0 207 NA 287 T   
## 14 23.0 208 NA 287 T   
## 15 23.0 209 NA 287 T   
## 16 23.0 210 NA 287 T   
## 17 23.0 211 NA 287 T   
## 18 23.0 212 NA 287 T   
## 19 23.0 213 NA 287 T   
## 20 23.0 214 NA 287 T

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

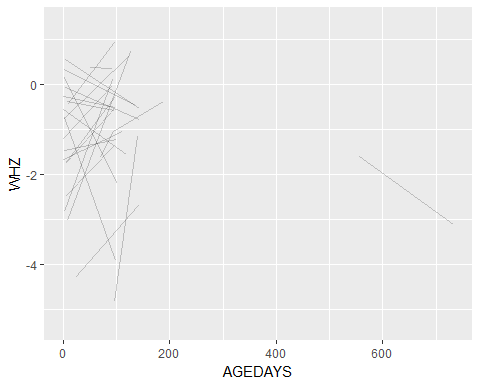
## [1] -0.66 NA NA NA NA NA NA NA NA NA NA  
## [12] NA NA NA NA NA NA NA NA NA -1.95 NA  
## [23] NA NA NA NA NA NA NA NA NA NA NA  
## [34] NA NA NA NA NA NA NA NA NA NA NA  
## [45] NA NA NA NA NA NA NA NA NA NA NA  
## [56] NA NA NA NA NA NA NA NA NA NA NA  
## [67] NA NA NA NA NA NA NA NA NA NA NA  
## [78] NA NA NA NA NA -3.49 NA NA NA NA NA  
## [89] NA NA -3.03 0.72 NA NA NA -1.38 NA NA NA  
## [100] NA NA NA NA NA NA NA NA NA NA NA  
## [111] NA NA NA NA NA NA NA NA NA NA NA  
## [122] NA NA NA NA NA NA NA NA NA NA NA  
## [133] NA NA NA NA NA NA NA NA NA NA NA  
## [144] NA NA NA NA NA NA NA NA NA NA NA  
## [155] NA NA NA NA NA NA NA NA NA NA NA  
## [166] NA NA NA NA NA NA NA NA NA NA NA  
## [177] NA NA NA NA NA NA NA NA NA NA NA  
## [188] NA NA NA NA NA NA NA NA NA NA NA  
## [199] NA -1.65

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 213 missing WHZ

##   
## FALSE TRUE   
## 10713 213

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

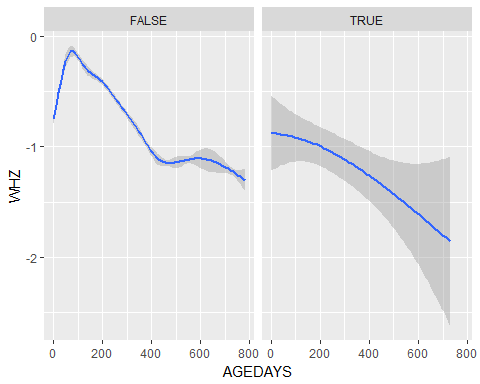
## Warning: Removed 1726 rows containing missing values (geom\_path).



# warning 1726 rows containing missing values  
   
#Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(jvit4) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) + facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 1844998 rows containing non-finite values (stat\_smooth).



# warning Removed 1,844,998 rows containing non-finite values

iLiNS-DOSE

# Make sure data arranged by subject id and age days  
 idose <- idose %>% arrange(SUBJID, AGEDAYS)  
  
# Check if age of death is recorded multiple times for each child who died  
 idose %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 53 x 2  
## SUBJID N  
## <dbl> <int>  
## 1 19.0 15  
## 2 53.0 6  
## 3 77.0 9  
## 4 79.0 3  
## 5 132 5  
## 6 183 8  
## 7 197 7  
## 8 212 1  
## 9 217 8  
## 10 249 9  
## 11 379 11  
## 12 390 4  
## 13 419 4  
## 14 461 8  
## 15 488 9  
## 16 494 7  
## 17 532 13  
## 18 544 11  
## 19 557 10  
## 20 613 9  
## 21 711 21  
## 22 799 5  
## 23 801 9  
## 24 840 5  
## 25 845 3  
## 26 866 9  
## 27 880 11  
## 28 898 2  
## 29 985 6  
## 30 1002 4  
## 31 1016 7  
## 32 1051 7  
## 33 1064 17  
## 34 1074 7  
## 35 1093 2  
## 36 1134 6  
## 37 1136 2  
## 38 1204 4  
## 39 1211 9  
## 40 1215 5  
## # ... with 13 more rows

# Recorded multiple times per subject  
   
# Create variable with maximum age per child  
 idose <- idose %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
# create indicator for last observation for a child using age   
 idose <- idose %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge)  
   
 # Create flag for the last time point for each subject if the child died   
 idose$death <- NA  
 idose <- idose %>% group\_by(SUBJID) %>% mutate(death = any(!is.na(AGEDTH)) & last==T)   
   
  
# Count total number of reported deaths using last age of the child  
 table(idose$death==T) # 53 deaths

##   
## FALSE TRUE   
## 23626 53

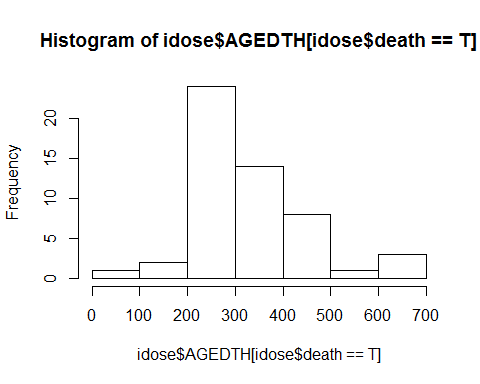
table(idose$DEAD[idose$death==T]) # 129 deaths

##   
## 1   
## 53

# DEAD and AGEDTH don't match!  
   
 summary(idose$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 7.0 273.0 358.0 378.3 459.0 667.0 23319

hist(idose$AGEDTH[idose$death==T])



# Check that the age of death is later than the oldest measurement on each child  
 idose <- idose %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(idose$deathErrorFlag[!is.na(idose$AGEDTH)]) # 38

##   
## FALSE TRUE   
## 322 38

#ISSUE - 38 children have measurements after their recorded age of death  
  
 # subset data to children with death error flags  
 df\_error <- idose %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [20]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <dbl> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 19.0 T 646 675 675 T   
## 2 53.0 T 323 337 337 T   
## 3 77.0 T 265 286 286 T   
## 4 132 T 257 326 326 T   
## 5 183 T 283 428 428 T   
## 6 197 T 285 300 300 T   
## 7 217 T 254 262 262 T   
## 8 249 T 483 1145 1145 T   
## 9 379 T 468 498 498 T   
## 10 419 T 224 255 255 T   
## 11 461 T 409 1441 1441 T   
## 12 544 T 417 439 439 T   
## 13 557 T 374 457 457 T   
## 14 613 T 385 399 399 T   
## 15 711 T 645 838 838 T   
## 16 799 T 315 664 664 T   
## 17 801 T 358 371 371 T   
## 18 845 T 205 233 233 T   
## 19 866 T 357 1742 1742 T   
## 20 898 T 342 352 352 T

#make indicator for any death (died)  
 idose <- idose %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>%  
 arrange(SUBJID, AGEDAYS)  
   
 #subset data set to children who died AND do not have death error flags  
 df\_death <- idose %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died, DEAD))  
   
 head(df\_death, 20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [3]  
## SUBJID AGEDAYS WHZ AGEDTH died DEAD  
## <dbl> <int> <dbl> <dbl> <lgl> <int>  
## 1 19.0 198 - 1.38 646 T 1  
## 2 19.0 255 NA 646 T 1  
## 3 19.0 281 NA 646 T 1  
## 4 19.0 283 NA 646 T 1  
## 5 19.0 306 NA 646 T 1  
## 6 19.0 311 NA 646 T 1  
## 7 19.0 330 NA 646 T 1  
## 8 19.0 353 NA 646 T 1  
## 9 19.0 381 0.0600 646 T 1  
## 10 19.0 395 NA 646 T 1  
## 11 19.0 408 NA 646 T 1  
## 12 19.0 479 NA 646 T 1  
## 13 19.0 563 - 2.25 646 T 1  
## 14 19.0 640 NA 646 T 1  
## 15 53.0 198 - 0.850 323 T 1  
## 16 53.0 228 NA 323 T 1  
## 17 53.0 256 NA 323 T 1  
## 18 53.0 296 NA 323 T 1  
## 19 53.0 317 NA 323 T 1  
## 20 77.0 175 - 0.420 265 T 1

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

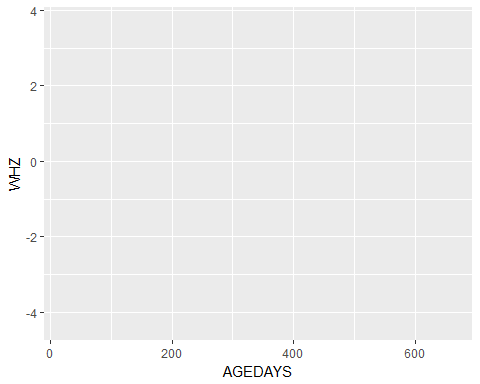
## [1] -1.38 NA NA NA NA NA NA NA 0.06 NA NA  
## [12] NA -2.25 NA -0.85 NA NA NA NA -0.42 NA NA  
## [23] NA NA NA NA NA 1.05 NA NA 0.12 NA NA  
## [34] NA -1.15 NA NA NA NA NA NA 2.30 NA NA  
## [45] NA NA NA -0.68 0.91 NA NA NA NA NA NA  
## [56] -0.25 NA NA NA NA NA NA NA -0.17 NA NA  
## [67] NA NA NA NA NA NA NA -0.51 NA NA NA  
## [78] 0.34 NA NA -2.00 NA NA NA NA -2.93 NA -0.35  
## [89] NA NA NA NA NA NA NA NA NA -0.25 NA  
## [100] NA NA NA NA 0.67 NA NA NA NA NA NA  
## [111] -1.41 NA NA NA NA NA NA -1.82 NA NA NA  
## [122] NA NA -4.35 NA NA -1.52 NA NA NA NA NA  
## [133] -1.85 NA NA -2.42 NA NA NA NA -1.34 NA NA  
## [144] 2.58 NA NA NA NA 3.07 NA NA NA NA NA  
## [155] 3.68 NA NA NA NA NA NA NA NA 1.10 NA  
## [166] NA NA NA 0.26 NA NA NA NA NA NA -1.37  
## [177] NA NA NA NA 1.13 NA 0.42 NA NA NA NA  
## [188] NA NA NA 1.01 NA NA NA NA NA NA NA  
## [199] -2.02 NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 66 missing WHZ

##   
## FALSE TRUE   
## 256 66

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

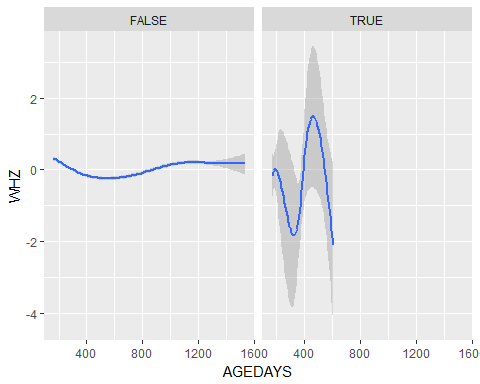
## Warning: Removed 178 rows containing missing values (geom\_path).



# Warning removed 178 rows containing missing values  
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(idose) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) + facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 14248 rows containing non-finite values (stat\_smooth).



# warning removed 14248 rows containing non-finite values

iLiNS-DYAD-M

# Make sure data arranged by subject id and age days  
 idyadm <- idyadm %>% arrange(SUBJID, AGEDAYS)  
  
# Check if age of death is recorded multiple times for each child who died  
 idyadm %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 57 x 2  
## SUBJID N  
## <dbl> <int>  
## 1 30291 17  
## 2 30331 20  
## 3 30401 1  
## 4 31041 18  
## 5 31081 1  
## 6 31281 8  
## 7 31661 1  
## 8 31911 7  
## 9 31971 14  
## 10 32081 21  
## 11 32461 10  
## 12 32661 10  
## 13 32721 9  
## 14 32781 8  
## 15 33021 3  
## 16 33091 9  
## 17 33281 13  
## 18 33381 14  
## 19 33451 3  
## 20 33462 5  
## 21 33511 4  
## 22 33611 3  
## 23 33631 2  
## 24 33791 13  
## 25 33901 18  
## 26 34081 7  
## 27 34101 1  
## 28 34111 7  
## 29 34161 1  
## 30 34291 9  
## 31 34461 2  
## 32 34581 2  
## 33 34621 1  
## 34 34921 1  
## 35 40031 15  
## 36 40321 6  
## 37 40361 6  
## 38 40461 2  
## 39 40781 9  
## 40 50651 11  
## # ... with 17 more rows

# Recorded multiple times per subject  
   
# Create variable with maximum age per child  
 idyadm <- idyadm %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
# create indicator for last observation for a child using age   
 idyadm <- idyadm %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge)  
   
# Create flag for the last time point for each subject if the child died   
 idyadm$death <- NA  
 idyadm <- idyadm %>% group\_by(SUBJID) %>% mutate(death = any(!is.na(AGEDTH)) & last==T)   
  
# Count total number of reported deaths using last age of the child  
 table(idyadm$death==T) # 57 deaths

##   
## FALSE TRUE   
## 20531 57

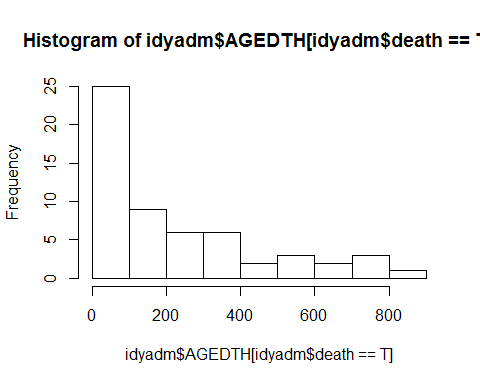
table(idyadm$DEAD[idyadm$death==T]) # 59 deaths

##   
## 1   
## 57

# DEAD and AGEDTH don't match!  
   
 summary(idyadm$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 0.0 187.0 336.0 369.2 572.0 824.0 20151

hist(idyadm$AGEDTH[idyadm$death==T])



# Check that the age of death is later than the oldest measurement on each child  
 idyadm <- idyadm %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(idyadm$deathErrorFlag[!is.na(idyadm$AGEDTH)]) #10

##   
## FALSE TRUE   
## 427 10

#ISSUE - 10 children have measurements after their recorded age of death  
   
# subset data to children with death error flags  
 df\_error <- idyadm %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 10 x 6  
## # Groups: SUBJID [10]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <dbl> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 31661 T 0 1 1.00 T   
## 2 31971 T 301 438 438 T   
## 3 33021 T 2.00 13 13.0 T   
## 4 33462 T 30.0 918 918 T   
## 5 33611 T 2.00 95 95.0 T   
## 6 34101 T 1.00 126 126 T   
## 7 34111 T 196 400 400 T   
## 8 50651 T 382 587 587 T   
## 9 81021 T 6.00 203 203 T   
## 10 83201 T 10.0 13 13.0 T

#make indicator for any death  
 idyadm <- idyadm %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>% arrange(SUBJID, AGEDAYS)  
   
#subset data set to children who died AND do not have death error flags  
 df\_death <- idyadm %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died, DEAD))  
   
 head(df\_death, 20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [2]  
## SUBJID AGEDAYS WHZ AGEDTH died DEAD  
## <dbl> <int> <dbl> <dbl> <lgl> <int>  
## 1 30291 1 NA 336 T 1  
## 2 30291 3 NA 336 T 1  
## 3 30291 34 1.21 336 T 1  
## 4 30291 95 NA 336 T 1  
## 5 30291 110 NA 336 T 1  
## 6 30291 111 NA 336 T 1  
## 7 30291 146 NA 336 T 1  
## 8 30291 148 NA 336 T 1  
## 9 30291 152 NA 336 T 1  
## 10 30291 174 NA 336 T 1  
## 11 30291 187 - 0.760 336 T 1  
## 12 30291 188 NA 336 T 1  
## 13 30291 208 NA 336 T 1  
## 14 30291 214 NA 336 T 1  
## 15 30291 287 NA 336 T 1  
## 16 30291 299 NA 336 T 1  
## 17 30291 308 NA 336 T 1  
## 18 30331 1 NA 447 T 1  
## 19 30331 10 NA 447 T 1  
## 20 30331 11 0.110 447 T 1

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

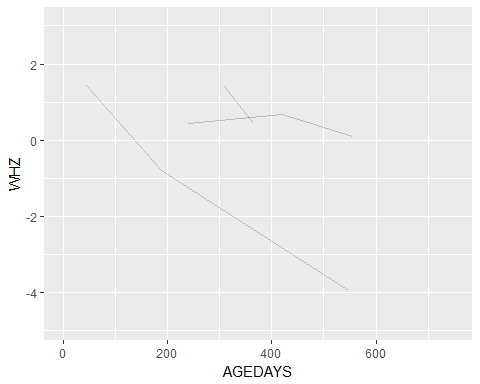
## [1] NA NA 1.21 NA NA NA NA NA NA NA -0.76  
## [12] NA NA NA NA NA NA NA NA 0.11 NA NA  
## [23] NA NA 0.45 NA NA NA NA NA -1.84 NA NA  
## [34] NA NA NA NA NA NA NA -1.90 NA NA NA  
## [45] NA NA -4.70 NA NA NA -1.69 NA NA -1.35 NA  
## [56] NA NA NA NA -1.68 NA NA NA NA -1.07 NA  
## [67] 0.20 NA NA NA NA NA NA NA -0.42 NA NA  
## [78] NA NA NA NA NA NA NA NA NA -0.49 NA  
## [89] NA NA NA NA NA NA NA NA 0.52 NA NA  
## [100] NA 3.09 NA NA NA -0.71 NA 2.30 NA NA NA  
## [111] NA NA NA NA NA NA NA NA NA NA NA  
## [122] NA NA NA NA NA NA NA NA NA NA -1.42  
## [133] NA NA -4.26 -0.09 NA NA NA NA NA NA -2.87  
## [144] NA NA NA NA -2.15 NA NA NA NA NA NA  
## [155] NA NA -0.73 NA NA NA NA NA NA NA NA  
## [166] -0.86 NA NA NA NA NA NA NA 1.97 NA NA  
## [177] 1.40 0.44 NA NA 1.55 NA 1.31 NA NA -0.95 NA  
## [188] NA NA -4.86 NA NA NA 1.75 NA -2.09 NA -0.74  
## [199] NA NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 81 missing WHZ

##   
## FALSE TRUE   
## 346 81

#Plot WHZ trajectories before death   
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

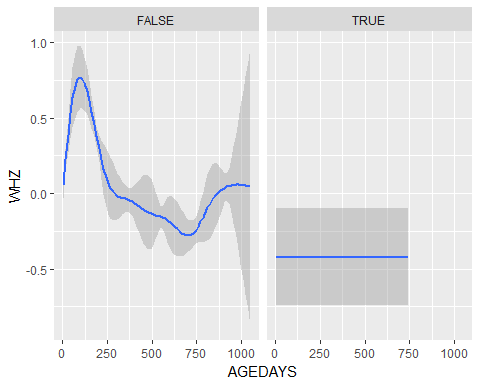
## Warning: Removed 178 rows containing missing values (geom\_path).



# warning removed 178 rows containing missing values   
   
#Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(idyadm) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) + facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 15346 rows containing non-finite values (stat\_smooth).



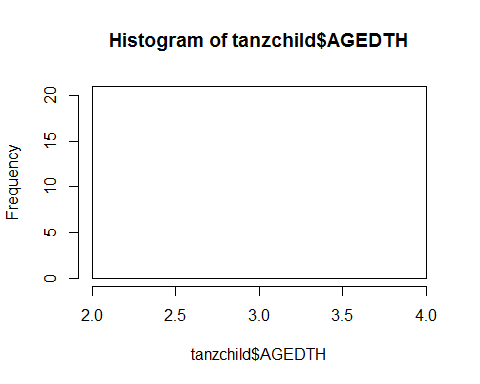
# warning Removed 15346 rows containing non-finite values  
   
 # Looks strange - horizontal line at roughly -0.45 for those who died

iLiNS-DYAD-G

# Make sure data arranged by subject id and age days  
 # idyadg <- idyadg %>% arrange(SUBJID, AGEDAYS)  
  
# DEATH VAR NOT INCLUDED YET, but published paper says there were 27 deaths  
  
# Check if age of death is recorded multiple times for each child who died  
 # idyadg %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)  
   
  
 # Recorded multiple times per subject  
   
# Create variable with maximum age per child  
 # idyadg <- idyadg %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
# create indicator for last observation for a child using age   
 # idyadg <- idyadg %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge, TRUE, FALSE)  
  
# Create flag for the last time point for each subject if the child died   
 # idyadg$death <- NA  
 # idyadg <- idyadg %>% group\_by(SUBJID) %>% mutate(death = any(!is.na(AGEDTH)) & last==T)   
   
  
# Count total number of reported deaths using last age of the child  
 # table(idyadg$death==T) # 57 deaths  
 # table(idyadg$DEAD[idyadg$death==T]) # 59 deaths  
   
 # DEAD and AGEDTH don't match!  
   
 # hist(idyadg$AGEDTH[idyadm$death==T])   
   
# Check that the age of death is later than the oldest measurement on each child  
 # idyadg <- idyadg %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 # table(idyadg$deathErrorFlag[!is.na(idyadg$AGEDTH)]) #39  
   
 #ISSUE - 39 children have measurements after their recorded age of death  
   
 # subset data to children with death error flags  
 # df\_error <- idyadg %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 # head(df\_error,20)  
   
 #make indicator for any death  
 # idyadg <- idyadg %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>% arrange(SUBJID, AGEDAYS)  
   
 #subset data set to children who died AND do not have death error flags  
 # df\_death <- idyadg %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died, DEAD))  
   
 # head(df\_death, 20)  
   
 #Check if anthropometry is measured when the age of death is recorded  
 # head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)  
 # table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 1396 missing WHZ  
   
 #Plot WHZ trajectories before death   
 # ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)   
 # warning removed 632 rows containing missing values   
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 # ggplot(idyadg) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) + facet\_wrap(~died)  
 # warning Removed 78964 rows containing non-finite values

TanzaniaChild2

# Make sure data arranged by subject id and age days  
 tanzchild <- tanzchild %>% arrange(SUBJID, AGEDAYS)  
  
# Something weird here... Only 1 child is marked as having died in the dataset, but the main paper said at least 45 children died  
 hist(tanzchild$AGEDTH)



# Check if age of death is recorded multiple times for each child who died  
 tanzchild %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 1 x 2  
## SUBJID N  
## <dbl> <int>  
## 1 60212 21

# Recorded 21 times per subject but only for SUBJID 60212   
   
# Create variable with maximum age per child  
 tanzchild <- tanzchild %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
# create indicator for last observation for a child using age   
 tanzchild <- tanzchild %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge)  
   
# Create flag for the last time point for each subject if the child died   
 tanzchild$death <- NA  
 tanzchild <- tanzchild %>% group\_by(SUBJID) %>% mutate(death = any(!is.na(AGEDTH)) & last==T)   
  
# Count total number of reported deaths using last age of the child  
 table(tanzchild$death==T) # 1 death

##   
## FALSE TRUE   
## 36517 1

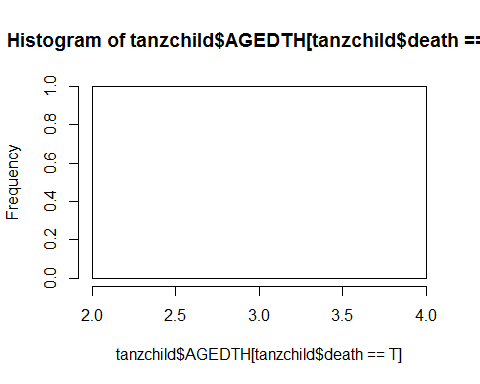
table(tanzchild$DEAD[tanzchild$death==T]) # 0 deaths

## < table of extent 0 >

# DEAD and AGEDTH don't match!  
 summary(tanzchild$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 3 3 3 3 3 3 36497

hist(tanzchild$AGEDTH[tanzchild$death==T])



ZVITAMBO

# Also measured cause of death, CAUSEDTH  
  
# Make sure data arranged by subject id and age days  
 zvit <- zvit %>% arrange(SUBJID, AGEDAYS)  
   
# Check if age of death is recorded multiple times for each child who died  
 zvit %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 1,067 x 2  
## SUBJID N  
## <dbl> <int>  
## 1 100101 3  
## 2 100241 10  
## 3 100671 16  
## 4 100751 1  
## 5 100781 2  
## 6 100791 8  
## 7 100881 3  
## 8 100981 8  
## 9 101011 1  
## 10 101051 11  
## 11 101481 6  
## 12 101571 13  
## 13 101741 14  
## 14 101851 7  
## 15 101921 10  
## 16 101941 13  
## 17 102041 11  
## 18 102121 10  
## 19 102271 6  
## 20 102331 2  
## 21 102601 3  
## 22 102861 7  
## 23 103141 14  
## 24 103231 3  
## 25 103511 7  
## 26 103601 4  
## 27 103701 6  
## 28 103771 3  
## 29 103831 5  
## 30 104191 14  
## 31 104281 5  
## 32 104381 7  
## 33 104431 3  
## 34 104461 4  
## 35 104561 13  
## 36 104651 4  
## 37 104711 22  
## 38 104761 8  
## 39 104851 5  
## 40 105151 1  
## # ... with 1,027 more rows

# Recorded multiple times per subject  
   
# Create variable with maximum age per child  
 zvit <- zvit %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
# create indicator for last observation for a child using age   
 zvit <- zvit %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge)  
   
# Create flag for the last time point for each subject if the child died   
 zvit$death <- NA  
 zvit <- zvit %>% group\_by(SUBJID) %>% mutate(death = any(!is.na(AGEDTH)) & last==T)   
  
# Count total number of reported deaths using last age of the child  
 table(zvit$death==T) # 1067 death

##   
## FALSE TRUE   
## 99414 1067

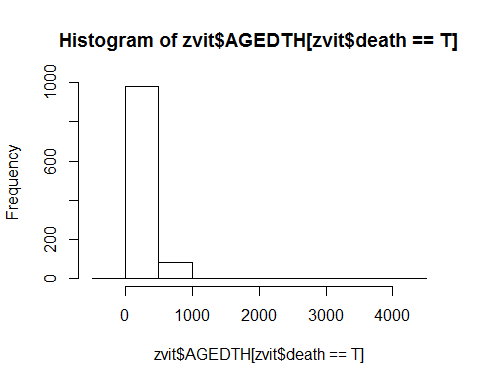
table(zvit$DEAD[zvit$death==T]) # 1123 deaths

##   
## 1   
## 1067

# DEAD and AGEDTH don't match!  
   
 summary(zvit$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## -8.0 105.0 237.0 288.3 418.0 4161.0 94812

hist(zvit$AGEDTH[zvit$death==T])



# Check that the age of death is later than the oldest measurement on each child  
 zvit <- zvit %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(zvit$deathErrorFlag[!is.na(zvit$AGEDTH)]) #85

##   
## FALSE TRUE   
## 5584 85

#ISSUE - 85 children have measurements after their recorded age of death  
   
 # subset data to children with death error flags  
 df\_error <- zvit %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 20 x 6  
## # Groups: SUBJID [20]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <dbl> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 101571 T 248 249 249 T   
## 2 102121 T 202 378 378 T   
## 3 103511 T 557 663 663 T   
## 4 104461 T 2.00 641 641 T   
## 5 105591 T 169 275 275 T   
## 6 107881 T 8.00 222 222 T   
## 7 109301 T 222 427 427 T   
## 8 109571 T 545 554 554 T   
## 9 110001 T 91.0 106 106 T   
## 10 113211 T 659 727 727 T   
## 11 113281 T 466 1054 1054 T   
## 12 115631 T 96.0 644 644 T   
## 13 116991 T 60.0 495 495 T   
## 14 119141 T 135 368 368 T   
## 15 119241 T 319 320 320 T   
## 16 121531 T 124 142 142 T   
## 17 121611 T 126 331 331 T   
## 18 122061 T 91.0 641 641 T   
## 19 122831 T 88.0 367 367 T   
## 20 123251 T 135 350 350 T

#make indicator for any death  
 zvit <- zvit %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>% arrange(SUBJID, AGEDAYS)  
   
 #subset data set to children who died AND do not have death error flags  
 df\_death <- zvit %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died))  
   
 head(df\_death, 20)

## # A tibble: 20 x 5  
## # Groups: SUBJID [3]  
## SUBJID AGEDAYS WHZ AGEDTH died   
## <dbl> <int> <dbl> <dbl> <lgl>  
## 1 100101 1 - 0.470 74.0 T   
## 2 100101 43 0.160 74.0 T   
## 3 100101 61 NA 74.0 T   
## 4 100241 1 0.100 514 T   
## 5 100241 61 2.82 514 T   
## 6 100241 107 NA 514 T   
## 7 100241 145 3.64 514 T   
## 8 100241 168 1.97 514 T   
## 9 100241 307 NA 514 T   
## 10 100241 347 0.210 514 T   
## 11 100241 495 NA 514 T   
## 12 100241 496 NA 514 T   
## 13 100241 497 NA 514 T   
## 14 100671 1 - 3.85 483 T   
## 15 100671 26 NA 483 T   
## 16 100671 111 3.64 483 T   
## 17 100671 189 1.79 483 T   
## 18 100671 367 NA 483 T   
## 19 100671 368 NA 483 T   
## 20 100671 369 NA 483 T

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

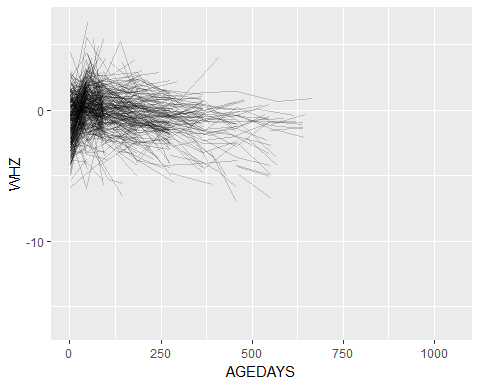
## [1] -0.47 0.16 NA 0.10 2.82 NA 3.64 1.97 NA 0.21 NA  
## [12] NA NA -3.85 NA 3.64 1.79 NA NA NA NA NA  
## [23] NA NA NA NA NA -7.32 NA -2.69 -0.30 NA 0.84  
## [34] 0.48 0.90 NA -0.58 NA NA NA -9.11 NA NA -2.76  
## [45] NA NA NA NA NA NA 0.19 -1.85 NA NA 1.46  
## [56] NA NA NA 1.58 0.34 NA NA NA -0.80 -1.37 -0.83  
## [67] NA NA NA -0.48 NA 0.88 -1.16 NA NA NA NA  
## [78] NA NA NA NA -4.87 2.13 NA 0.18 0.34 NA NA  
## [89] NA -0.85 NA NA NA -4.41 NA -0.06 NA -1.28 -0.81  
## [100] NA 0.24 NA 0.52 NA NA NA NA NA NA -0.77  
## [111] NA NA 1.22 NA 1.57 -1.05 NA NA NA -1.38 0.15  
## [122] -1.48 -1.22 -0.46 NA 2.77 NA NA NA NA 0.22 2.00  
## [133] -1.40 NA NA NA -0.81 NA 1.20 -1.82 -2.24 NA NA  
## [144] NA NA -0.51 NA -1.40 -2.31 -2.50 NA 1.22 NA 0.88  
## [155] NA NA -0.44 NA NA NA NA NA NA 1.91 NA  
## [166] 1.03 -0.06 -0.11 -0.31 0.06 -0.36 NA NA NA NA NA  
## [177] NA 1.22 NA NA 0.85 NA 1.81 NA NA NA -1.76  
## [188] NA 1.11 1.57 -0.79 NA 0.43 -1.34 NA NA 0.13 NA  
## [199] NA -0.11

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 2690 missing WHZ at time of death

##   
## FALSE TRUE   
## 2894 2690

#Plot WHZ trajectories before death among those who died AND w/o death errors  
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

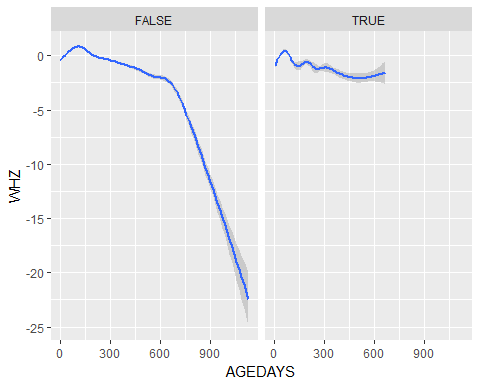
## Warning: Removed 1717 rows containing missing values (geom\_path).



# warning Removed 1717 rows containing missing values  
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(zvit) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) + facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 36476 rows containing non-finite values (stat\_smooth).



# warning Removed 36476 rows containing non-finite values

PROVIDE

# Make sure data arranged by subject id and age days  
 provide <- provide %>% arrange(SUBJID, AGEDAYS)  
  
# Check if age of death is recorded multiple times for each child who died  
 provide %>% group\_by(SUBJID) %>% filter(!is.na(AGEDTH)) %>% summarize(N=n()) %>% print(n=40)

## # A tibble: 5 x 2  
## SUBJID N  
## <dbl> <int>  
## 1 1317 763  
## 2 2757 762  
## 3 5871 762  
## 4 8800 762  
## 5 9707 762

# Recorded multiple times per subject  
   
# Create variable with maximum age per child  
 provide <- provide %>% group\_by(SUBJID) %>% mutate(maxAge=max(AGEDAYS))  
   
# create indicator for last observation for a child using age   
 provide <- provide %>% group\_by(SUBJID) %>% mutate(last = AGEDAYS==maxAge)  
   
# Create flag for the last time point for each subject if the child died   
 provide$death <- NA  
 provide <- provide %>% group\_by(SUBJID) %>% mutate(death = any(!is.na(AGEDTH)) & last==T)   
   
# Count total number of reported deaths using last age of the child  
 table(provide$death==T) # 5 death

##   
## FALSE TRUE   
## 535140 5

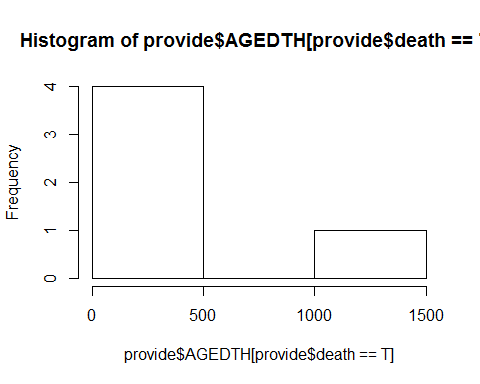
table(provide$DEAD[provide$death==T]) # 5 deaths

##   
## 1   
## 5

summary(provide$AGEDTH)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 11.0 33.0 183.0 327.6 194.0 1216.0 531334

hist(provide$AGEDTH[provide$death==T])



# Check that the age of death is later than the oldest measurement on each child  
 provide <- provide %>% group\_by(SUBJID) %>% mutate(deathErrorFlag= AGEDTH < maxAge & last==TRUE & !is.na(AGEDTH))  
   
 table(provide$deathErrorFlag[!is.na(provide$AGEDTH)]) # 4

##   
## FALSE TRUE   
## 3807 4

#ISSUE - 4 children have measurements after their recorded age of death  
 # A LOT of children were followed up to 762 days.   
   
 # subset data to children with death error flags  
 df\_error <- provide %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 4 x 6  
## # Groups: SUBJID [4]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <dbl> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 2757 T 194 762 762 T   
## 2 5871 T 11.0 762 762 T   
## 3 8800 T 183 762 762 T   
## 4 9707 T 33.0 762 762 T

# subset data to children with death error flags  
 df\_error <- provide %>% filter(deathErrorFlag) %>% subset(., select=c(SUBJID, last, AGEDTH, AGEDAYS, maxAge, deathErrorFlag))  
   
 head(df\_error,20)

## # A tibble: 4 x 6  
## # Groups: SUBJID [4]  
## SUBJID last AGEDTH AGEDAYS maxAge deathErrorFlag  
## <dbl> <lgl> <dbl> <int> <dbl> <lgl>   
## 1 2757 T 194 762 762 T   
## 2 5871 T 11.0 762 762 T   
## 3 8800 T 183 762 762 T   
## 4 9707 T 33.0 762 762 T

#make indicator for any death  
 provide <- provide %>% group\_by(SUBJID) %>% mutate(died= any(!is.na(AGEDTH))) %>% arrange(SUBJID, AGEDAYS)  
   
 #subset data set to children who died AND do not have death error flags  
 df\_death <- provide %>% filter(died, deathErrorFlag==FALSE) %>% subset(., select=c(SUBJID, AGEDAYS, WHZ, AGEDTH, died))  
   
 head(df\_death, 20)

## # A tibble: 20 x 5  
## # Groups: SUBJID [1]  
## SUBJID AGEDAYS WHZ AGEDTH died   
## <dbl> <int> <dbl> <dbl> <lgl>  
## 1 1317 1 NA 1216 T   
## 2 1317 2 NA 1216 T   
## 3 1317 3 NA 1216 T   
## 4 1317 4 NA 1216 T   
## 5 1317 5 NA 1216 T   
## 6 1317 6 NA 1216 T   
## 7 1317 7 - 1.66 1216 T   
## 8 1317 8 NA 1216 T   
## 9 1317 9 NA 1216 T   
## 10 1317 10 NA 1216 T   
## 11 1317 11 NA 1216 T   
## 12 1317 12 NA 1216 T   
## 13 1317 13 NA 1216 T   
## 14 1317 14 NA 1216 T   
## 15 1317 15 NA 1216 T   
## 16 1317 16 NA 1216 T   
## 17 1317 17 NA 1216 T   
## 18 1317 18 NA 1216 T   
## 19 1317 19 NA 1216 T   
## 20 1317 20 NA 1216 T

#Check if anthropometry is measured when the age of death is recorded  
 head(df\_death$WHZ[!is.na(df\_death$AGEDTH)], 200)

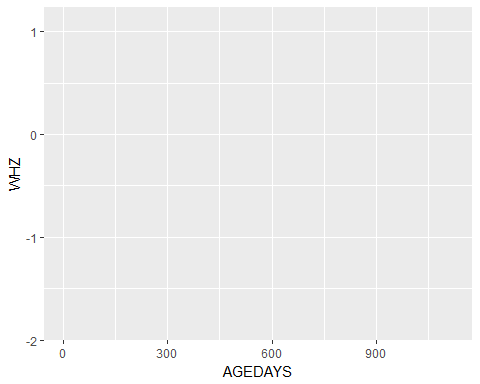
## [1] NA NA NA NA NA NA -1.66 NA NA NA NA  
## [12] NA NA NA NA NA NA NA NA NA NA NA  
## [23] NA NA NA NA NA NA NA NA NA NA NA  
## [34] NA NA NA NA NA NA NA NA NA NA NA  
## [45] NA NA NA NA NA NA 1.09 NA NA NA NA  
## [56] NA NA NA NA NA NA NA NA NA NA NA  
## [67] NA NA NA NA NA NA NA NA NA NA NA  
## [78] NA NA NA NA NA 1.05 NA NA NA NA NA  
## [89] NA NA NA NA NA NA NA NA NA NA 0.84  
## [100] NA NA NA NA NA NA NA NA NA NA NA  
## [111] NA -0.32 NA NA NA NA NA NA NA NA NA  
## [122] NA NA NA NA NA NA NA NA NA NA NA  
## [133] NA -0.52 NA NA NA NA NA -0.87 NA NA NA  
## [144] NA NA NA NA NA NA NA NA NA NA NA  
## [155] NA NA NA NA NA NA NA NA NA NA NA  
## [166] NA NA NA NA NA NA NA NA NA NA NA  
## [177] NA NA NA NA NA NA -1.13 NA NA NA NA  
## [188] NA NA NA NA NA NA NA NA NA NA NA  
## [199] NA NA

table(!is.na(df\_death$WHZ) & !is.na(df\_death$AGEDTH)) # 35 missing WHZ

##   
## FALSE TRUE   
## 3772 35

#Plot WHZ trajectories before death among those who died and w/o death errors  
 ggplot(df\_death) + geom\_line(aes(x=AGEDAYS, y=WHZ, group=SUBJID), alpha=0.2)

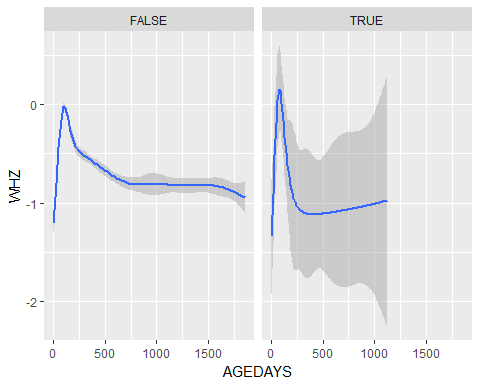
## Warning: Removed 2705 rows containing missing values (geom\_path).



# ISSUE: plot is blank.   
   
 #Smoothed WHZ trajectories before death and compare to those who didn't die  
 ggplot(provide) + geom\_smooth(aes(x=AGEDAYS, y=WHZ)) + facet\_wrap(~died)

## `geom\_smooth()` using method = 'gam'

## Warning: Removed 523542 rows containing non-finite values (stat\_smooth).



# warning Removed 523542 rows containing non-finite values