Malaria incidence in children, a recurrent events model

Jordache Ramjith, Andreas Bender, Kit C.B. Roes & Marianne A. Jonker

Introduction

We illustrate application of PAMMs in the analysis of the effect of covariates on the time to malaria incidence in children, with possible recurrences. This guide is setup to make the code used for the manuscript results available. The children are anonymized and random `id`s were made. The data, unfortunately, can not be shared, but we show the structure of the data and detailed analysis below, such that application to new data should be straight forward. For further information on access to the data, please see the data availability section of the published study.

Packages

The PAMM application requires the installation and loading of two specific user-written packages. These are the **`pammtools`** and **`mgcv`** packages. The pammtools package includes the data augmentation function that can restructure your data into the required structure for piece-wise exponential models, and some utility functions that can easily provide estimates of the hazards, cumulative hazards and survival probabilities, which can be used for visualization. See the package page for more details. The package **`mgcv`** is used for building the actual PAMM. We also use additional packages for data wrangling and visualization.

```
library(pammtools)
library(mgcv)
library(purrr) #data wrangling
library(dplyr) #data wrangling
library(ggplot2) # visualization
theme_set(theme_bw() +
  theme(legend.position = "bottom"))
library(patchwork)
```

Example data

The data set (`malaria`), is read in below and has 1796 observations from 678 children with a maximum of 10 recurrences. The `malaria` data is in longitudinal format reflecting the recurrences for children over different rows. The data come from a Malaria study conducted in Busia district, south-eastern Uganda, an area of high malaria transmission intensity where the goal was to study the effect of intermittent preventive treatment of malaria in pregnancy with dihydroartemisinin-piperaquine (DP) versus sulfadoxine-pyrimethamine (SP) on the incidence of malaria in infancy.

An excerpt of the data is shown below with

- `id` child identification number
- `enum` the event number (1 = first malaria infection, 2 = second event/first recurrence)
- `start` the time the subject enters the risk set for the event number `enum`
- `time` the time at which subject experiences an event or is censored for event number `enum`
- `status` (0 = censored, 1 = event)
- `dobday` day of year at which child was born (1-365)
- `treatment` SP or DP intermittent preventive treatment given to mothers during pregnancy

- `preterm` yes or no for whether gestational age is under 37 weeks
- `sex` female or male
- `gravidity` (1-9) mother's number of prior pregnancies

```
malaria <- readRDS("malaria.Rds")
malaria %>% filter(id %in% c("31161", "32260", "31088"))
```

```
## # A tibble: 8 x 10
##      id  enum start  time status dobday treatment preterm sex
##   <dbl> <int> <dbl> <dbl>  <dbl>  <dbl> <fct>     <fct>   <fct>
## 1 31088     1     0    32      1     64 SP        no      Female
## 2 31088     2    32   101      1     64 SP        no      Female
## 3 31088     3   101   192      0     64 SP        no      Female
## 4 31161     1     0    28      1     94 DP        no      Male
## 5 31161     2    28   238      1     94 DP        no      Male
## 6 31161     3   238   308      1     94 DP        no      Male
## 7 31161     4   308   364      0     94 DP        no      Male
## 8 32260     1     0     3      0     51 SP        yes     Male
## # ... with 1 more variable: gravidity <dbl>
```

Here we aim to analyze the data using PAMMs on the calendar time scale, additionally estimating the seasonality of occurrences and checking potential interactions between child age and seasonality (day of year).

Piece-wise exponential data

In order to apply PAMMs, we first transform the data to the *piece-wise exponential data* (PED) format (see here for details). The transformation is performed by the `as_ped` function from **`pammtools`**. For recurrent events on the calendar time scale,

- we have to provide the name of the column that stores the information about the event number via the `transition` argument
- we also have to specify `timescale = "calendar"`, as `"gap"` is the default
- the `Surv` object must be in start-stop format because we want to work on the calendar time scale,
- we also use the day of year a child was born to create a variable `doy` (day of year that the child entered each risk interval) that will be used to model seasonality.

```
ped <- malaria %>%
  as_ped(
    formula = Surv(start, time, status)~treatment + sex + preterm + gravidity + dobday,
    id = "id",
    transition = "enum",
    timescale = "calendar") %>%
  mutate(id=as.factor(id))
# create day of year variable for seasonality effects
ped$doy <- ped$dobday + ped$tstart
ped$doy = ifelse(
  ped$doy>730,
  ped$doy-730,
  ifelse(ped$doy>365,ped$doy-365,ped$doy))
```

An excerpt of the data for the same 3 subjects as before is shown below (we show the first and last row per subject and event number). Here,

- `ped_status` will be the target variable for the Poisson regression (these are the δ_{ijk} from the manuscript) and `offset` will enter as an offset (these are the o_{ijk}). The `t_j` that are used to estimate the baseline hazard are given as `tend`. Note, here `tend` is equivalent to the age of the child (at the end of the interval), thus we rename the variable in this case for better interpretation.

```
ped %>%
  filter(id %in% c("31161", "32260", "31088")) %>%
  group_by(id, enum) %>%
  mutate(offset=round(offset,2)) %>%
  slice(1, n()) %>%
  knitr::kable()
```

id	tstart	tend	interval	offset	ped_status	treatment	sex	preterm	gravidity	dobday	enum	doy
31088	0	0.5	(0,0.5]	-0.69	0	SP	Female	no	1	64	1	64
31088	29	32.0	(29,32]	1.10	1	SP	Female	no	1	64	1	93
31088	32	33.0	(32,33]	0.00	0	SP	Female	no	1	64	2	96
31088	100	101.0	(100,101]	0.00	1	SP	Female	no	1	64	2	164
31088	101	102.0	(101,102]	0.00	0	SP	Female	no	1	64	3	165
31088	191	192.0	(191,192]	0.00	0	SP	Female	no	1	64	3	255
31161	0	0.5	(0,0.5]	-0.69	0	DP	Male	no	3	94	1	94
31161	26	28.0	(26,28]	0.69	1	DP	Male	no	3	94	1	120
31161	28	29.0	(28,29]	0.00	0	DP	Male	no	3	94	2	122
31161	237	238.0	(237,238]	0.00	1	DP	Male	no	3	94	2	331
31161	238	239.0	(238,239]	0.00	0	DP	Male	no	3	94	3	332
31161	307	308.0	(307,308]	0.00	1	DP	Male	no	3	94	3	36
31161	308	309.0	(308,309]	0.00	0	DP	Male	no	3	94	4	37
31161	363	364.0	(363,364]	0.00	0	DP	Male	no	3	94	4	92
32260	0	0.5	(0,0.5]	-0.69	0	SP	Male	yes	1	51	1	51
32260	2	3.0	(2,3]	0.00	0	SP	Male	yes	1	51	1	53

Model 1

Below, we fit the model with an interaction effect between sex and treatment (in the paper where the data was analyzed there appeared to possibly be a `treatment*sex` interaction). Additionally, malaria incidence in Uganda is known to have seasonal increase and decrease due to rainy and dry seasons, therefore, we additionally introduce a non-linear smooth function that models such seasonality as a continuous effect. We also include preterm birth and gravidity in the model.

Fitting the model

In the code chunk below we use the `pamm` function to fit a Piecewise exponential Additive Mixed Model (PAMM), which is a wrapper around `mgcv::gam` or `mgcv::bam`, depending of the specification of the `engine` argument. The other arguments of the functions are directly passed to these functions (with `family = poisson()` and `offset` set to the `offset` variable in the `ped` data set for convenience).

In the formula specification of the model

- `s(tend)` indicates the smooth effect that estimates the deviation of the log baseline hazard over time (`tend`) from the estimated intercept,
- `s(doy, bs = "cc")`, where `cc` indicates a cyclical spline (see `?mgcv::smooth.terms` for details), and
- `s(id, bs = "re")` indicates a random effect (frailty) for each child where random effects basis are specified for id `bs="re"`, to allow child-specific Gaussian distributed random effects.

```
pamm_malaria <- pamm(
  formula  = ped_status ~ s(tend) + sex *treatment + preterm + gravidity +
    s(doy, bs = "cc") + s(id, bs = "re"),
  data     = ped,
  engine   = "bam",
```

```
  method   = "fREML",
  discrete = TRUE)
summary(pamm_malaria)
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## ped_status ~ s(tend) + sex * treatment + preterm + gravidity +
##     s(doy, bs = "cc") + s(id, bs = "re")
##
## Parametric coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -5.81466    0.11540 -50.388   <2e-16 ***
## sexFemale              0.17079    0.12222   1.397   0.1623
## treatmentSP            0.29687    0.12286   2.416   0.0157 *
## pretermyes            -0.11159    0.18909  -0.590   0.5551
## gravidity              0.04580    0.02104   2.177   0.0295 *
## sexFemale:treatmentSP -0.27909    0.16946  -1.647   0.0996 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##             edf  Ref.df Chi.sq p-value
## s(tend)   7.171   8.169 167.19  <2e-16 ***
## s(doy)    5.225   8.000  70.68  <2e-16 ***
## s(id)   267.725 672.000 559.66  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.005   Deviance explained = 9.08%
## fREML = 1.8779e+05  Scale est. = 1         n = 198172
```

Firstly, we see that the frailty term `s(id)` is statistically significant $p < 2e-16$. The estimated standard deviation for the frailty distribution can be found using the following code chunk.

```
gam.vcomp(pamm_malaria)
```

```
##
## Standard deviations and 0.95 confidence intervals:
##
##             std.dev        lower       upper
## s(tend) 0.002897123 0.0014049249 0.005974214
## s(doy)  0.001721390 0.0008398855 0.003528080
## s(id)   0.691774576 0.6144940932 0.778774068
##
## Rank: 3/3
```

Further, the summary indicates that main effect of treatment is statistically significant. This coincides with the baseline sex in the model, boys. The interaction effect however was not statistically significant. Instead of doing manual calculations for the treatment effect for girls, the hazard ratio for DP treatment vs. SP treatment for each sex can be calculated using the chunk below, where

- `make_newdata` creates a data frame with unique values of `sex` and treatment variable set to `"DP"` value (all other covariates set to mean/modus)

- `add_hazard` calculates the hazard, given model `pamm_malaria`
- note that the `reference` argument is specified in `add_hazard`, which means the function calculates the hazard ratio of covariate values in the data set created by `make_newdata` and a data set where the covariates specified in the `reference` argument are replaced by the respective values

Note that even though the column header states "hazard", this is indeed a hazard ratio (of `DP` compared with `SP`) because we specified `SP` as a reference in the chunk below. Here we see that while we have a significant effect for boys (`DP` is associated with a lower hazard than `SP`), this is not necessarily true for girls.

```
ped %>%
 make_newdata(sex = unique(sex), treatment = c("DP")) %>%
  add_hazard(
    pamm_malaria,
    reference = list(treatment = c("SP"))) %>%
  select(sex, hazard, ci_lower, ci_upper)
```

```
##      sex    hazard  ci_lower  ci_upper
## 1   Male 0.7431386 0.5812411 0.9501307
## 2 Female 0.9823767 0.7779921 1.2404547
```

We also find a non-linear baseline hazard, which in this case is equivalent to dependence of the hazard on the child's age and also support for the seasonality hypothesis, as indicated by the excerpt from the summary output shown below (effective degrees of freedom (edf) for `s(tend)` (baseline) and `s(doy)` (seasonality) are both substantially larger than 1 and significant):

```
summary(pamm_malaria)
```

```
...
##
## Approximate significance of smooth terms:
##             edf  Ref.df Chi.sq p-value
## s(tend)   7.171   8.169 167.19  <2e-16 ***
## s(doy)    5.225   8.000  70.68  <2e-16 ***
## s(id)   267.725 672.000 559.66  <2e-16 ***
## ---
...
```

The respective estimates are better understood through visualization.

Estimates over time and visualization

The baseline hazards

The baseline hazard (when all covariates are set to zero, or the reference category) coincides with the baseline incidence rate in Poisson models, where the incidence rate is the number of events per child unit of time. In the analysis of the malaria data, time was set in days. So the incidence is the number of malaria events per child day over time in the first year of life. The incidence can then simply be turned into a number of events per child-year by a simple multiplication of 365.25 (or any other timescale that is more clinically interpretable).

From 's(tend)' in the output, we saw that there is a statistically significant effect of time/age on malaria incidence that appears to be non-linear. For clearer interpretation, it is better to visualize the baseline hazards and 95% confidence intervals over time. Since the minimum gravidity is 1, we estimate the baseline hazard at gravidity equal to 1, and for seasonality, day of year cannot be zero, so we set this equal to 1 too. All other variables are set to the reference categories. Note, that the coefficients of the output indicate whether the curve increases or decreases for values of other covariates (i.e. the proportional hazards assumption). For the visualization, we multiply the estimated hazards by 365.25 for clearer epidemiological interpretation, and similarly transform the Age in days to months.

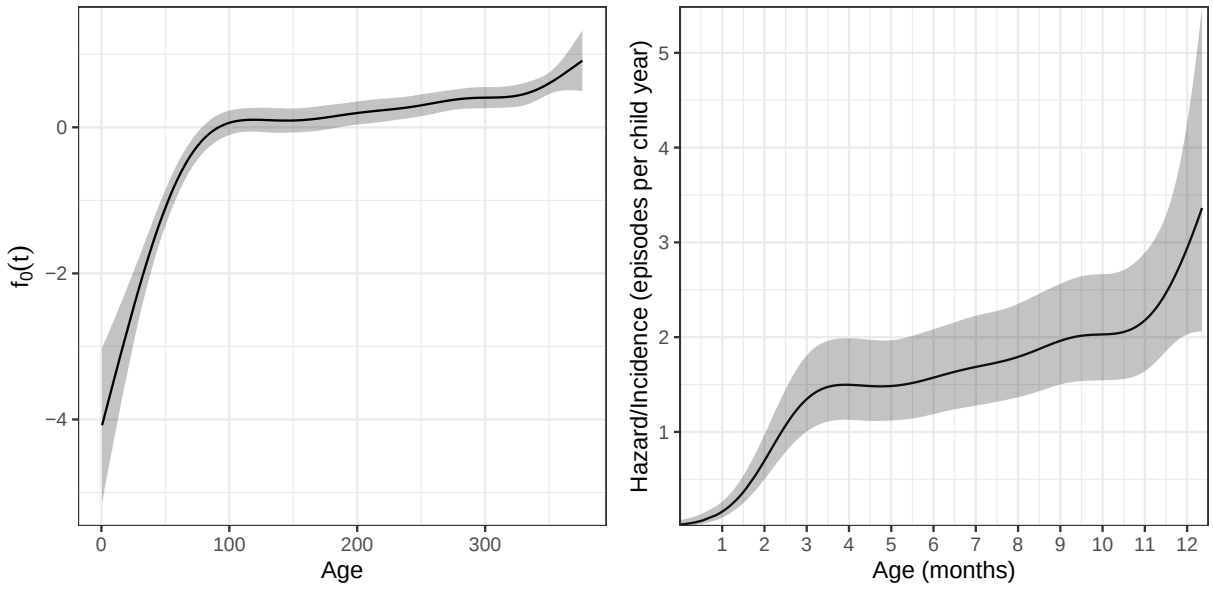
The most general way in **pammtools** to create such plots is to create a new data frame with the desired values of the covariates (`make_newdata`) and then use one of the `add_*` functions (here `add_hazard`). The `make_newdata` function sets all other numeric covariates to their mean and categorical covariates to their mode, thus to obtain only the non-linear part of the baseline-hazard, we use the `exclude` argument to set the contributions of these terms to 0. Obtaining specific term contributions on the log-scale is easier, as we can use the `gg_slice` convenience function. For more control, we also could have used `make_newdata` in combination with `add_term` (see the Seasonality example further below).

```
ndf_base <- ped %>%
  make_newdata(tend = unique(tend),gravidity=c(1),doy=c(1)) %>%
  add_hazard(pamm_malaria, exclude = c("s(id)"))

p_term <- gg_slice(ped, pamm_malaria, term = "tend", tend = unique(tend)) +
 ylab(expression(f[0](t))) + xlab("Age")

p_hazard <- ggplot(ndf_base, aes(x = tend/(365.25/12), y = hazard*365.25)) +
  geom_line() +
  geom_ribbon(aes(ymin = ci_lower*365.25, ymax = ci_upper*365.25), alpha = .3) +
  ylab("Hazard/Incidence (episodes per child year)") + xlab("Age (months)")+
  scale_x_continuous(breaks=seq(1,12,1),limits=c(0,12.5), expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))

p_term + p_hazard
```



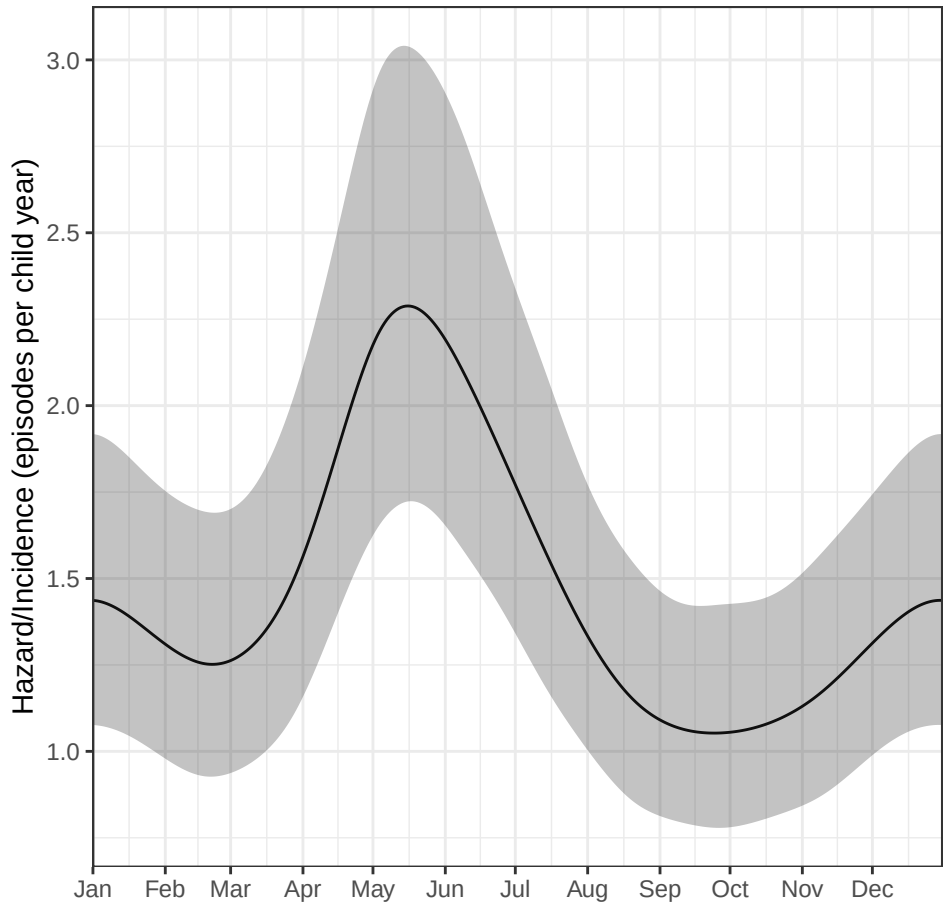
We can clearly see, that the hazard rapidly increases with increasing age until an age of ~ 3.5 months and only a slow increase afterwards. For this particular set of covariates; we see that the incidence reaches 1.5 malaria episodes per child year by 3.5 months. For a child with all covariates the same but 1 unit higher gravidity, we expect the incidence to be higher.

Seasonality

We will use the same set of covariates as when we looked at visualizing the baseline incidence/hazard rates over time. However now we will choose a set value for time, say 100 days (roughly 3.5 months), and we will allow season (`doy`) values to vary from 1 to 365 days. We do this in the same way we created the earlier visualization over time.

```r
ndf <- ped %>%
  make_newdata(doy = 1:365,gravidity=c(1),tend=c(100)) %>%
  mutate(date = as.Date("2017-01-01") + (doy-1)) %>%
  #converts doy into dates for easier interpretation
  add_hazard(
    pamm_malaria,
    exclude = c("s(id)"))

ggplot(ndf, aes(x = date, y = hazard*365.25)) +
  geom_line() +
  geom_ribbon(aes(ymin = ci_lower*365.25, ymax = ci_upper*365.25), alpha = .3) +
  scale_x_date(expand = c(0, 0),date_breaks = "1 month",date_labels = "%b") +
  ylab("Hazard/Incidence (episodes per child year)") + xlab("")
```



We can clearly see the seasonality of the effect with higher according to the wet and dry seasons.We see a small peak in mid-January and a larger peak at the end of May/early June - towards the end of the rainy season.

Model 2

The addition of time-varying effects of sex to model 1

Even though the treatment and sex interaction in model 1 was statistically not significant, there appeared to be relative treatment differences in boys and not girls. There are currently other studies investigating sex differences in malaria incidence amongst children, and we thought that it may be interesting to consider a time-varying effect of sex added to model 1.

Fitting the model

The smooth function for the time-varying effect can be fit using either stratified smooth functions by specifying `s(tend, by=sex)` (where sex must be defined as a factor variable) or by using difference smooth functions which uses the `as.ordered()` function as we've done in the code chunk below. Stratified smooth functions will fit a separate smooth function for boys and girls and estimate separate dofs. Difference smooth functions will fit a smooth function of the deviation of the difference in log-hazards between boys and girls over time, from the average difference in log-hazards between boys and girls. We have a slight preference for the difference smooth function because the p-value from the difference smooth function will inform us of whether there is significant evidence for a time-varying effect.

```
pamm_malaria_tv <- pamm(
  formula  = ped_status ~ s(tend) + sex*treatment + preterm + s(tend,by=as.ordered(sex)) +
    gravidity + s(doy, bs = "cc") + s(id, bs = "re"),
  data     = ped,
  engine   = "bam",
  method   = "fREML",
  discrete = TRUE)
summary(pamm_malaria_tv)
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## ped_status ~ s(tend) + sex * treatment + preterm + s(tend, by = as.ordered(sex)) +
##     gravidity + s(doy, bs = "cc") + s(id, bs = "re")
##
## Parametric coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -5.85081    0.11709 -49.968   <2e-16 ***
## sexFemale              0.22117    0.12498   1.770   0.0768 .
## treatmentSP            0.29854    0.12286   2.430   0.0151 *
## pretermyes            -0.11161    0.18901  -0.591   0.5548
## gravidity              0.04584    0.02103   2.180   0.0293 *
## sexFemale:treatmentSP -0.27973    0.16939  -1.651   0.0987 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                   edf  Ref.df  Chi.sq p-value
## s(tend)                         7.163   8.145 110.324  <2e-16
## s(tend):as.ordered(sex)Female   2.765   3.446   7.372  0.0703
## s(doy)                          5.211   8.000  70.309  <2e-16
## s(id)                         267.397 672.000 558.550  <2e-16
##
```

```
## s(tend)                         ***
## s(tend):as.ordered(sex)Female .
## s(doy)                          ***
## s(id)                           ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.00502   Deviance explained = 9.16%
## fREML = 1.8779e+05  Scale est. = 1         n = 198172
```

Here we see that the estimate dof for the difference smooth function is 2.7 suggesting some "wiggliness" of the log hazard ratio over time. While this is not statistically significant ($p = 0.070$), this p-value is not far from 0.05 and warrants visualization.

Visualization of the model

To visualize and understand the impact of a time-varying hazard ratio, it is important to visualize how both the hazards stratified for sex and the hazard ratio evolve over time. We again do this using the `make_newdata()` and `add_hazard()` convenience functions.

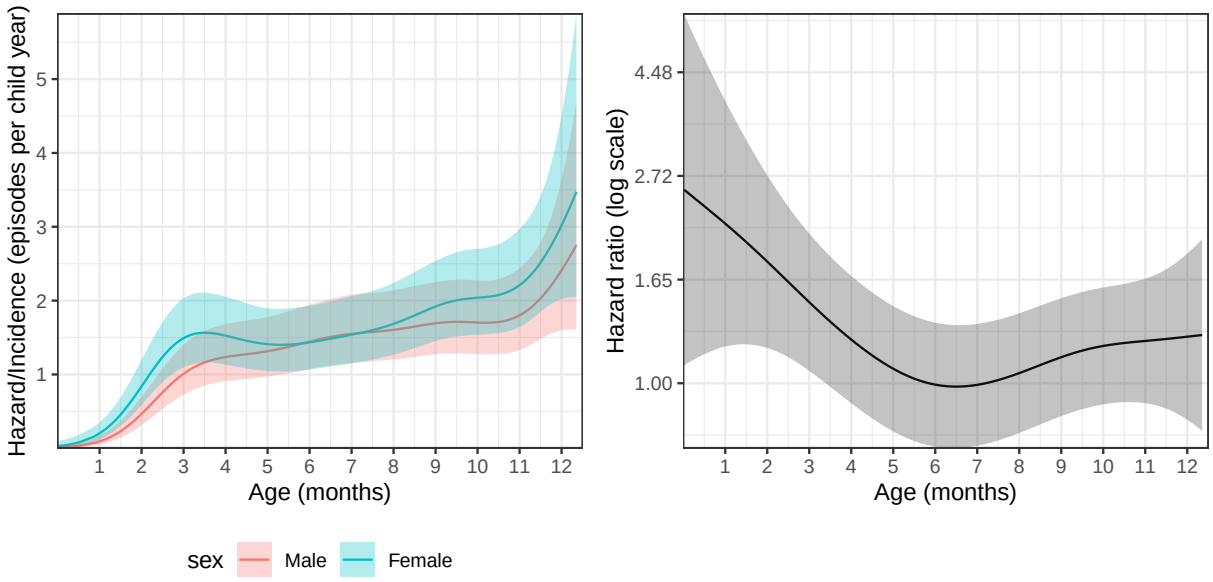
```
#first the data for the stratified hazards
ndf_tv <- ped %>%
  make_newdata(tend = unique(tend),sex=unique(sex),gravidity=c(1),doy=c(1)) %>%
  add_hazard(pamm_malaria_tv, exclude = c("s(id)"))

#then the data for the time-varying HRs
ndf_tvhr <- ped %>%
  make_newdata(tend = unique(tend), sex=c("Female"), treatment=c("DP"), gravidity=c(1),
        doy=c(1)) %>%
  add_hazard(
    pamm_malaria_tv,
    reference = list(sex = c("Male")))

p_haz <- ggplot(ndf_tv, aes(x = tend/(365.25/12), y = hazard*365.25)) +
  geom_line(aes(col=sex)) +
  geom_ribbon(aes(ymin = ci_lower*365.25, ymax = ci_upper*365.25,fill=sex), alpha = .3) +
  ylab("Hazard/Incidence (episodes per child year)") + xlab("Age (months)")+
  scale_x_continuous(breaks=seq(1,12,1),limits=c(0,12.5), expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))

p_hr <- ggplot(ndf_tvhr, aes(x = tend/(365.25/12), y = hazard)) +
  geom_line() +
  geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper), alpha = .3) +
  ylab("Hazard ratio (log scale)") + xlab("Age (months)")+
  scale_x_continuous(breaks=seq(1,12,1),limits=c(0,12.5), expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0), trans='log',
    breaks=scales::trans_breaks('log', function(x) exp(x)),
    labels = function(x) sprintf("%0.2f",round(x,2)))

p_haz+p_hr
```



From the right panel in the figure above, we see that the hazards of girls relative to the hazards of boys are much higher in the first three months and thereafter approaches a null effect. It can also be seen from the left panel that the absolute risk (incidence rates) are very low during this period.

Model 3

The addition of an interaction of child age and seasonality to model 1

As children get older, they might be left unattended for longer periods of time and potentially spent more time outside or otherwise unprotected (e.g. by nets).

It could therefore be assumed that there might be an interaction between child age and seasonality when it comes to the hazard of malaria infection, e.g. such that seasonality doesn't affect the hazard for very young children.

Fitting the model

To investigate such a relationship, we replace the baseline hazard by a bivariate tensor product spline specified by `te` terms. Note that because the basis functions of the tensor product spline are comprised of the marginal basis function for each of the variables involved, here day of year (`doy`) and child age (`tend`) we can also mix and match different basis functions for the respective dimensions (here cubic spline bases (`cr`) for the marginal basis functions w.r.t. `tend` and cyclic cubic spline bases (`cc`) for the marginal basis functions of the `doy` variable).

```
pamm_malaria_te <- pamm(
    formula = ped_status ~  te(tend, doy, bs = c("cr", "cc")) +
      sex*treatment + preterm + gravidity +
      s(id, bs = "re"),
    data     = ped,
    engine   = "bam",
    method   = "fREML",
    discrete = TRUE)
summary(pamm_malaria_te)
```

```
##
```

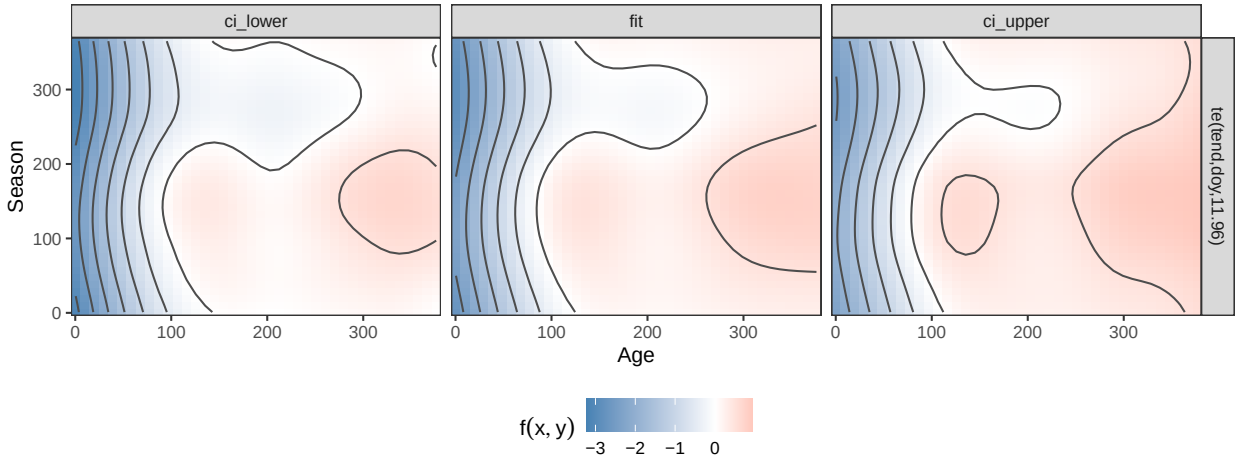
```
## Family: poisson
## Link function: log
##
## Formula:
## ped_status ~ te(tend, doy, bs = c("cr", "cc")) + sex * treatment +
##     preterm + gravidity + s(id, bs = "re")
##
## Parametric coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -5.77036    0.11535 -50.025   <2e-16 ***
## sexFemale              0.17085    0.12219   1.398   0.1620
## treatmentSP            0.30059    0.12287   2.446   0.0144 *
## pretermyes            -0.11165    0.18889  -0.591   0.5544
## gravidity              0.04575    0.02103   2.175   0.0296 *
## sexFemale:treatmentSP -0.27913    0.16942  -1.648   0.0995 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                 edf Ref.df Chi.sq p-value
## te(tend,doy)  11.96  14.28  249.4  <2e-16 ***
## s(id)        266.77 672.00  556.5  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.00465   Deviance explained = 8.66%
## fREML = 1.8781e+05  Scale est. = 1         n = 198172
```

The summary output, indicates that the `te` term is significant, however, this doesn't necessarily mean, that there is an interaction. To investigate, we compare the models with and without the interaction.

Visualization of the model

For a quick overview of bivariate functions estimated by the `te` terms, **`pammtools`** provides the convenience function `gg_tensor`, which shows the term contribution on the log-hazard scale (i.e. the deviations from the intercept given that particular set of covariates).

```
p_tensor <- gg_tensor(pamm_malaria_te, ci = TRUE) +
   xlab("Age") + ylab("Season")
p_tensor
```



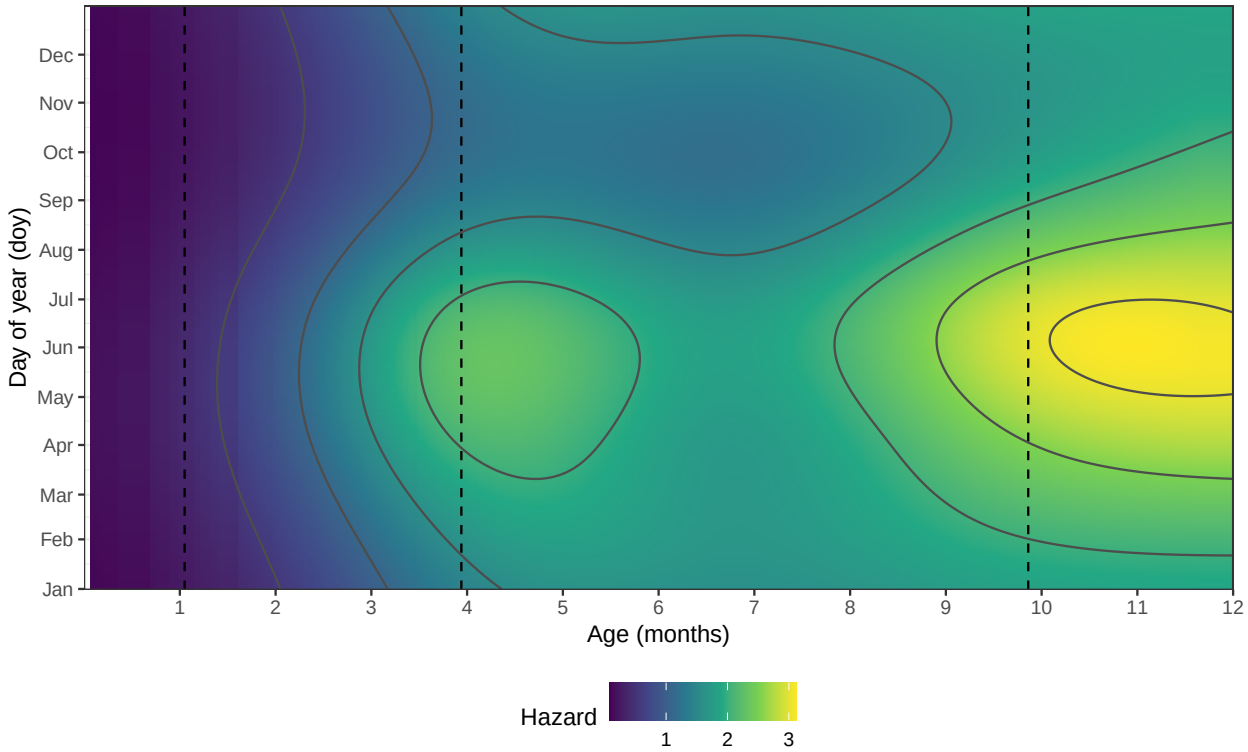
Note, that to visualize the hazard (given a set of covariates) we could also use the `make_newdata` and `add_term` or `add_hazard` combination and then use `ggplot2::geom_tile` for visualization. Here we see that the yellow regions are where we expect the highest incidence rates of more than 3 episodes per child year for children with this particular set of covariates (`treat`, `sex` and `preterm` are set at their reference categories `male`, `DP` and `no` respectively).

```
ndf_te <- ped %>%
  make_newdata(doy = 1:365,gravidity=c(1),tend=1:365) %>%
  mutate(date = as.Date("2017-01-01") + (doy-1)) %>%
  #converts doy into dates for easier interpretation.
  add_hazard(
    pamm_malaria_te,
    exclude = c("s(id)"))
```

```
## Some values of 'tend' have been set to the respecitve interval end-points
```

```
p_tensor <- ggplot(ndf_te, aes(x = tend/(365.25/12), y = date, z = hazard*365.25)) +
  geom_tile(aes(fill = hazard*365.25), width=0.33) +
  scale_y_date(expand = c(0, 0),date_breaks = "1 month",date_labels = "%b")+
  scale_x_continuous(expand = c(0, 0),limits=c(0,12),breaks=seq(1,12,1)) +
  scale_fill_viridis_c(
    name = "Hazard",
    breaks = seq(0, 11, by = 1)) +
  geom_hline(yintercept = c(54, 151, 300), lty = 2) +
  geom_vline(xintercept = c(1.05, 3.94, 9.86), lty = 2) +
  geom_contour(col = "grey30") +
  ylab("Day of year (doy)") + xlab("Age (months)")

p_tensor
```

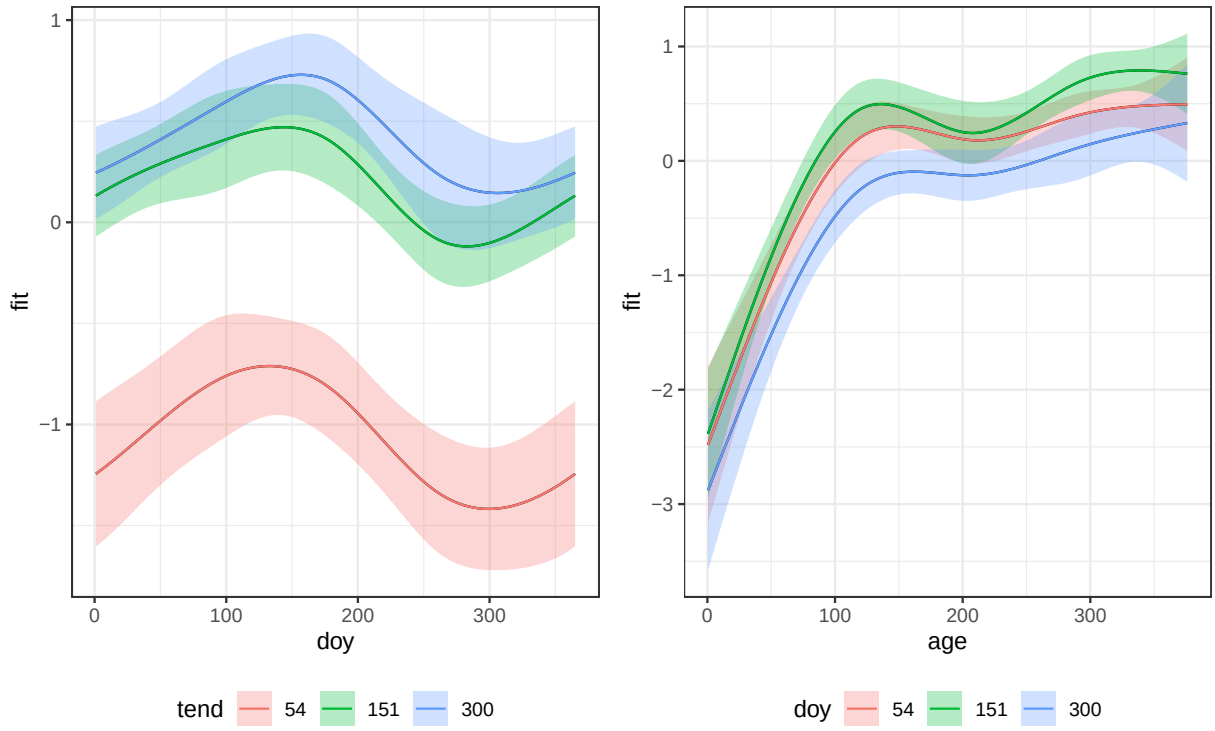


While the increase in hazard with increasing age is clearly visible from this figure, as well as a visible seasonality effect, we cannot however determine whether there is a large interaction effect. We therefore also look at slices trough this surface.

Slices through the surface w.r.t. to day of year (`doy`) and child age (`tend`), for first event are given below. The functional shapes of age and season do not change substantially for different values of the other covariate, which indicates no or only weak interaction.

```
slice_age<- gg_slice(ped, pamm_malaria_te, term = "doy",
  tend  = c(54, 151,  300), doy=1:365)
slice_doy <- gg_slice(ped, pamm_malaria_te, term = "doy", tend  = unique(tend),
  doy=c(54, 151, 300)) + xlab("age")

slice_age + slice_doy
```



ANOVA type decomposition of the model

Instead of a visual inspection, we could also fit the model using ANOVA type decomposition of the `te` term into main effects and an interaction effect (`ti(tend, doy)`). The summary indicates that this interaction effect is not significant and the effective degrees of freedom for the interaction term are almost 0. The rest of the summary output is almost identical to the output of the non-interaction model (`pamm_malaria`).

```
pamm_malaria_ti <- pamm(
    formula = ped_status ~  ti(tend, bs = "cr", k = 10) +
      ti(doy, bs = "cc", k = 10) + ti(tend, doy, bs = c("cr", "cc")) +
      sex*treatment + preterm + gravidity +
      s(id, bs = "re"),
    data     = ped,
    engine   = "bam",
    method   = "fREML",
    discrete = TRUE)
summary(pamm_malaria_ti)
```

```
##
## Family: poisson
## Link function: log
##
## Formula:
## ped_status ~ ti(tend, bs = "cr", k = 10) + ti(doy, bs = "cc",
##     k = 10) + ti(tend, doy, bs = c("cr", "cc")) + sex * treatment +
##     preterm + gravidity + s(id, bs = "re")
##
## Parametric coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)           -5.81448    0.11533 -50.417   <2e-16 ***
## sexFemale              0.17080    0.12221   1.398   0.1622
## treatmentSP            0.29689    0.12286   2.417   0.0157 *
## pretermyes            -0.11164    0.18909  -0.590   0.5549
## gravidity              0.04579    0.02104   2.177   0.0295 *
## sexFemale:treatmentSP -0.27909    0.16946  -1.647   0.0996 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                     edf  Ref.df  Chi.sq p-value
## ti(tend)     7.175e+00   8.216 166.835  <2e-16 ***
## ti(doy)      5.224e+00   8.000  70.649  <2e-16 ***
## ti(tend,doy) 8.275e-04  12.000   0.001   0.436
## s(id)        2.677e+02 672.000 559.621  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.005   Deviance explained = 9.08%
## fREML = 1.8779e+05  Scale est. = 1         n = 198172
```