ITS Analysis

5/31/23

About this script

This script was crafted using Ellen's example codes. The script has three sections:

1. Data info
2. Exploration
3. Model fit

Install and/or load packages if required

```
knitr::opts_chunk$set(eval = TRUE, include = TRUE)
# for first use only
#install.packages("INLA", repos = "https://inla.r-inla-download.org/R/stable", dep = TRUE)
library(tidyverse) # data management
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.1     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.2     v tibble    3.2.1
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.1
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```
library(readxl) # read xls file
library(hablar) # convert variable type
```

```
Attaching package: 'hablar'

The following object is masked from 'package:forcats':

    fct

The following object is masked from 'package:dplyr':

    na_if

The following object is masked from 'package:tibble':

    num
```

```r
library(janitor) # clean names
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```r
library(INLA) # modelling
```

```
Warning: package 'INLA' was built under R version 4.2.3

Loading required package: Matrix

Attaching package: 'Matrix'

The following objects are masked from 'package:tidyr':

    expand, pack, unpack

Loading required package: foreach

Attaching package: 'foreach'
```

```
The following objects are masked from 'package:purrr':

    accumulate, when

Loading required package: parallel
Loading required package: sp
This is INLA_23.04.24 built 2023-04-24 19:15:35 UTC.
 - See www.r-inla.org/contact-us for how to get help.
 - To enable PARDISO sparse library; see inla.pardiso()
```

```
library(lubridate) # date
library(zoo) # irregular time
```

```
Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric
```

Standard ITS

The data was aggregated monthly from 2013 to 2018 at the village level in Amhara Region, Ethiopia. The below code will read the data from the `xlsx` file. The intervention took a 2014 mass test-and-treat campaign for malaria control, followed by focal test-and-treat through 2017. Intervention impact is estimated using the INLA package.

Case data: There are 6 villages, all intervention groups. Villages are observed for 61 months and incidence cases were captured each month.

Intervention (MTAT): the intervention rolled out on month 12 (Sept 2014) in all intervention villages. Intervention villages are filtered for standard ITS

Inspect the data structure and variable lists

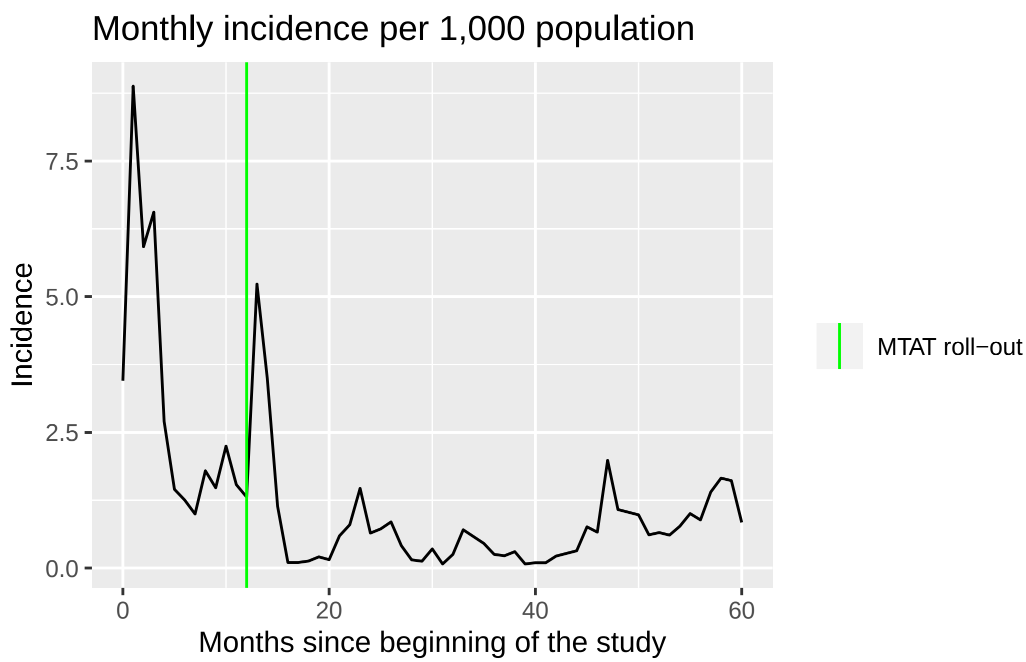
```
s_data %>%
  glimpse()
```

```
Rows: 366
Columns: 20
```

```
$ village         <chr> "Kumer Aftit", "Kumer Aftit", "Kumer Aftit", "Kumer Af~
$ period          <chr> "September 2013", "October 2013", "November 2013", "De~
$ stratum         <chr> "3", "3", "3", "3", "3", "3", "3", "3", "3", "3", "3",~
$ population      <dbl> 3199, 3199, 3199, 3199, 3263, 3263, 3263, 3263, 3263, ~
$ cases           <dbl> 47, 65, 29, 17, 7, 12, 11, 10, 27, 23, 23, 15, 26, 53,~
$ ir              <chr> "14.69209127852454", "20.31884964051266", "9.065332916~
$ altitude        <dbl> 735, 735, 735, 735, 735, 735, 735, 735, 735, 735, 735,~
$ longitude       <dbl> 36.36133, 36.36133, 36.36133, 36.36133, 36.36133, 36.3~
$ latitude        <dbl> 12.80153, 12.80153, 12.80153, 12.80153, 12.80153, 12.8~
$ evi             <dbl> 0.4484723, 0.3960933, 0.2563009, 0.1690017, 0.1485518,~
$ lst_day         <dbl> 30.32500, 32.55667, 36.82500, 38.05000, 40.28500, 43.1~
$ lst_night       <dbl> 19.47119, 20.75667, 19.16000, 16.62387, 18.32500, 19.7~
$ rainfall        <dbl> 1.281828e+02, 5.595180e+01, 1.619233e+00, 2.295833e-01~
$ lli_ns_coverage <dbl> 100, 100, 100, 100, 100, 100, 100, 100, 100, 100, 100,~
$ irs_coverage    <dbl> 80, 80, 80, 80, 80, 80, 80, 80, 80, 80, 80, 80, 82, 82~
$ id              <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
$ time            <int> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ~
$ time_post       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 3, 4, 5, ~
$ post            <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, ~
$ date            <date> 2013-09-01, 2013-10-01, 2013-11-01, 2013-12-01, 2014-~
```

Standard ITS - Exploratory

```
ggplot(s_data %>%
         group_by(time) %>%
         summarise(incidence = sum(cases)/sum(population)*1000)) +
  geom_line(aes(x = time, y = incidence)) +
  geom_vline(aes(xintercept = 12, linetype = "MTAT roll-out"), color = "green") +
  labs(title = "Monthly incidence per 1,000 population", linetype = "",
       y = "Incidence", x = "Months since beginning of the study")
```



Standard ITS - Modelling

Monthly incidence rate estimates assuming a negative binomial distribution, using the log population as an offset, with stratum as fixed effect, random intercepts for each village (IID), and the month of observation (AR-1).

```
# build a model
models = inla(cases ~ time + time_post + post + stratum + I(lag(evi)) +
               I(lag(rainfall)) +
               I(lag(lst_night)) +
               I(lag(lst_day)) +
               I(as.factor(month(date))) +
               offset(log(population)) +
               f(time2, model = "ar1") + f(id, model = "iid"),
               # model fitness & comparison
               control.compute=list(dic=TRUE,cpo=TRUE,  waic = TRUE),
               family = "nbinomial", data = s_data%>%
               mutate(time2 = time))

# save the model results in data frame
results = exp(summary(models)[[3]])[,c(1,3,5)] %>%
```

```
  round(3) %>%
  as.data.frame()

colnames(results) = c("Incidence rate ratio", "2.5th quantile, IRR", "97.5th quantile, IRR
rownames(results) = c("Intercept", "Months since study beginning", "Months since MTAT", "S
                      "Post period", "1-month lagged EVI", "1-month lagged rainfall",
                      "1-month lagged night temp", "1-month lagged day temp",
                      "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov"

kableExtra::kable(results, booktabs = T)
```

	Incidence rate ratio	2.5th quantile, IRR	97.5th quantile, IRR
Intercept	0.000	0.000	0.017
Months since study beginning	0.861	0.698	1.060
Months since MTAT	1.138	0.910	1.420
Stratum2b	0.563	0.106	3.050
Stratum3	0.960	0.115	8.645
Post period	19.708	2.282	207.680
1-month lagged EVI	0.013	0.000	1.027
1-month lagged rainfall	1.003	0.999	1.007
1-month lagged night temp	0.935	0.809	1.081
1-month lagged day temp	1.064	0.972	1.162
Feb	0.742	0.289	1.917
Mar	0.785	0.273	2.270
Apr	0.613	0.183	2.069
May	0.899	0.261	3.149
Jun	2.404	0.696	8.466
Jul	3.951	1.179	13.437
Aug	6.733	1.489	31.000
Sep	7.207	1.436	36.892
Oct	17.099	3.888	76.096
Nov	12.528	3.554	44.567
Dec	3.920	1.484	10.340

Next, we conduct a counter factual analysis to estimate how many additional cases would have been observed if not for the intervention. To accomplish this, we convert the predicted case counts, which are labelled `mean` in `models$summary.fitted.values`, back to the log scale so that we can remove the effects associated with the intervention. Specifically, the regression coefficient for the post-intervention period (level change) and time since intervention roll-out (trend change), multiplied by time since roll-out, are subtracted off. Now, we have counterfactual case counts in log-scale, which we can exponeniate to get counterfactual cases. In our data, cases declined with 532, or 32% decrease, during the MTAT/FTAT period.

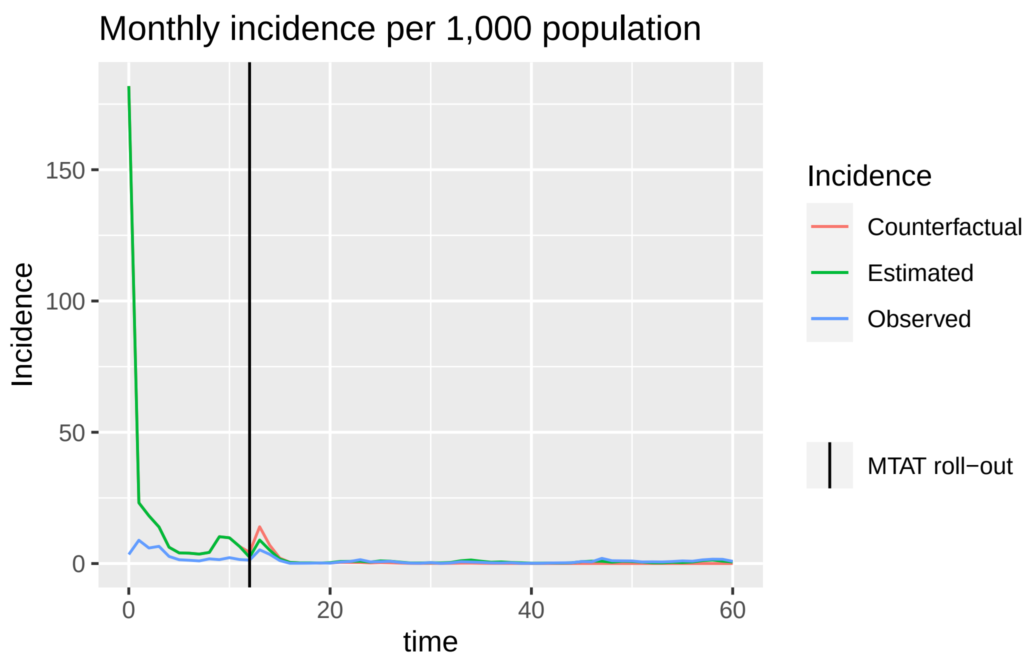
```
counterfactual = s_data %>%
  bind_cols(models$summary.fitted.values) %>% #Join predicted case counts to data
  filter(post==1) %>% #Restrict to the post period
  mutate(log_mean = log(mean), #Convert estimates back to log scale
         counterfactual_cases = exp(log_mean -
                                      summary(models)[[3]]["time_post","mean"] * time_post
                                      summary(models)[[3]]["post","mean"] * post)) %>%
  summarise(counterfactual_cases = sum(counterfactual_cases),
            estimated_cases = sum(mean))

counterfactual$counterfactual_cases - counterfactual$estimated_cases #Cases averted
```

```
[1] -550.2246
```

```
(counterfactual$counterfactual_cases - counterfactual$estimated_cases)/counterfactual$esti
```

```
[1] -32.71774
```

```
ggplot(s_data %>%
         bind_cols(models$summary.fitted.values) %>% #Join predicted case counts to data
         mutate(log_mean = log(mean), #Convert estimates back to log scale
                counterfactual_cases = exp(log_mean -
                                      summary(models)[[3]]["time_post","mean"] * time_post
                                      summary(models)[[3]]["post","mean"] * post)) %>%
         group_by(time) %>%
         summarise(incidence = sum(cases)/sum(population)*1000,
                   estimated_incidence = sum(mean)/sum(population)*1000,
                   counterfactual_incidence = sum(counterfactual_cases)/sum(population)*10
  geom_line(aes(x = time, y = counterfactual_incidence, color = "Counterfactual")) +
  geom_line(aes(x = time, y = estimated_incidence, color = "Estimated")) +
  geom_line(aes(x = time, y = incidence, color = "Observed")) +
  geom_vline(aes(xintercept = 12, linetype = "MTAT roll-out")) +
  labs(title = "Monthly incidence per 1,000 population", linetype = "", color = "Incidence
       y = "Incidence")
```



Controlled ITS

The data was aggregated monthly from 2013 to 2018 at the village level in Amhara Region, Ethiopia. The code provided below will read the data from the specified xlsx file (e.g., "filename.xlsx") and the corresponding sheet name (if applicable). The purpose of this analysis is to conduct a controlled interrupted time series analysis in R, focusing on a mass test-and-treat campaign in 2014 (MTAT), followed by focal test-and-treat interventions until 2017 for malaria control. The intervention impact will be estimated using the Integrated Nested Laplace Approximation (INLA) package.

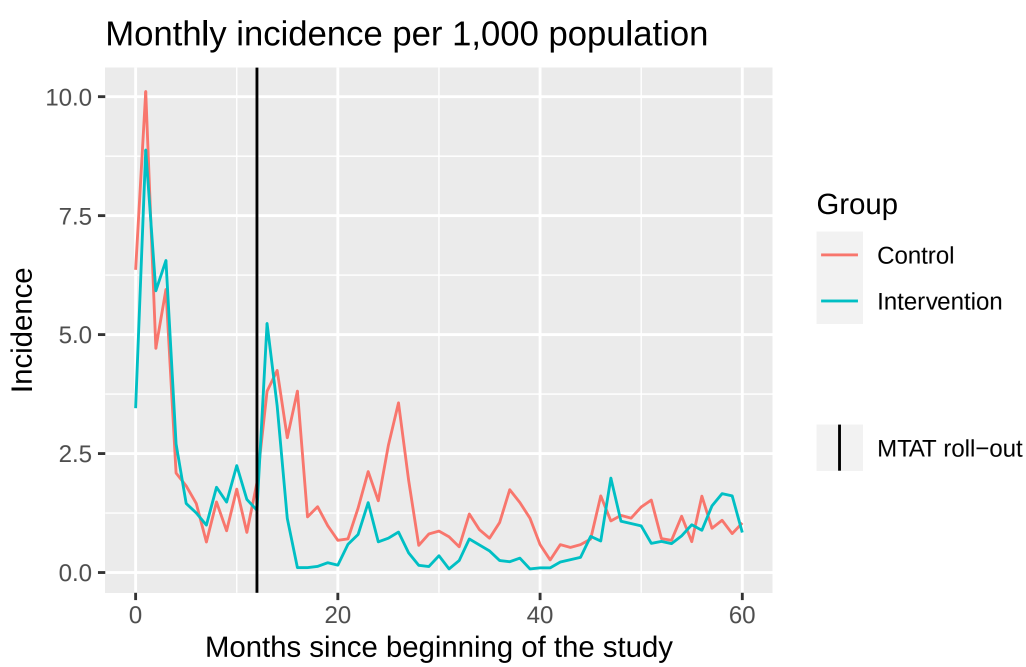
Case data: The study includes a total of 12 villages, with 6 villages assigned to the intervention group and 6 to the control group. These villages were observed for 61 months, and the incidence cases were captured monthly.

Intervention (MTAT): The intervention was implemented in all intervention villages in September 2014 (month 12). The villages were divided into two groups: control and intervention.

Time: Time is defined in months. In the dataset, there are three time variables: (1) time since the beginning of the study, (2) a spline term representing months since the beginning of the intervention, (3) time_post spline representing the post-intervention period, and (4) a post indicator to determine observations following the intervention roll-out. Our exposures of interest are the interactions of the group with (1) the post-intervention period and (2) time since the intervention.

Covariates: The analysis includes environmental data extracted from the Malaria Atlas Project database, such as evi, daylight temp, night light temp, and rainfall. Additionally, intervention data related to vector control measures, including Long-Lasting Insecticidal Nets (LLINs) and Indoor Residual Spraying (IRS), are considered as covariates.

```
ggplot(c_data %>%
         group_by(group, time) %>%
         summarise(incidence = sum(cases)/sum(population)*1000)) +
  geom_line(aes(x = time, y = incidence, color = group)) +
  geom_vline(aes(xintercept = 12, linetype = "MTAT roll-out")) +
  labs(title = "Monthly incidence per 1,000 population", linetype = "",
       y = "Incidence", x = "Months since beginning of the study", color = "Group")
```



We estimate monthly incidence rates, assuming a negative binomial distribution, using log population as an offset with, `fixed effects of stratum`, random intercepts for each village (IID) and the month of observation (AR-1). Interaction terms are used to estimate the relative change in incidence rates in intervention areas compared to control areas 1) immediately following MTAT (level change) and 2) each additional month thereafter (trend change).

In the intervention group, we estimate a –% greater reduction in incidence following roll-out compared to the control group. A –% decrease per month over baseline is estimated in the intervention group above and beyond the decrease estimated in the control group (--%). Prior distribution is set to fix the prcesion.

Random intercept for each village (IID), the month of observation (AR-1), and stratum (iid)

Model comparison: DIC, CPO and/or PIT

The precision of stratum has been fixed with the prior information.

Deviance information criterion (DIC) - the smallest the DIC indicates the best fit model.

```
c(modelc$dic$dic, modelc2$dic$dic )
```

```
[1] 3342.594 3342.221
```

Conditional predictive ordinate - the smallest the CPO indicates the best fit model.

```
# stratum as fixed effect
cpo.ctis <- modelc$cpo$cpo
cITS <- sum(log(cpo.ctis))
cITS
```

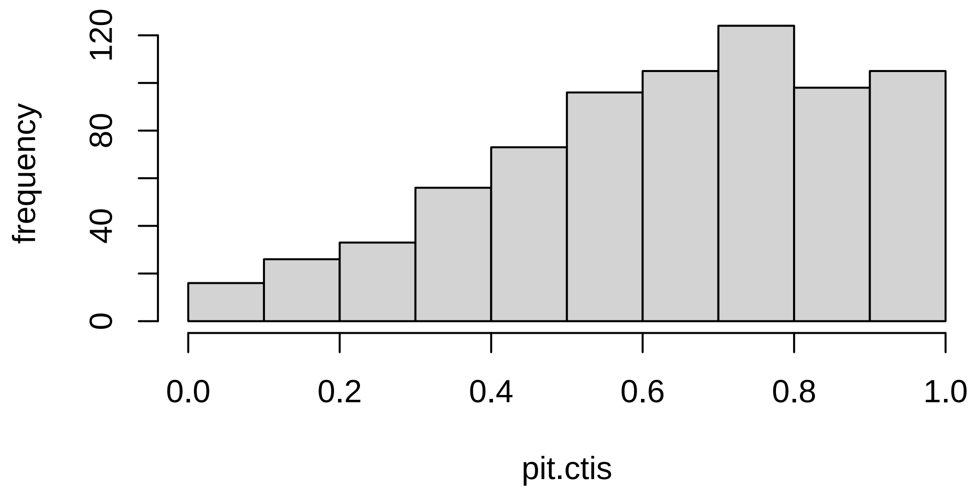
```
[1] -1697.864
```

```
# stratum as random effect
cpo.c2tis <- modelc2$cpo$cpo

controlITS <- sum(log(cpo.c2tis))

controlITS #
```

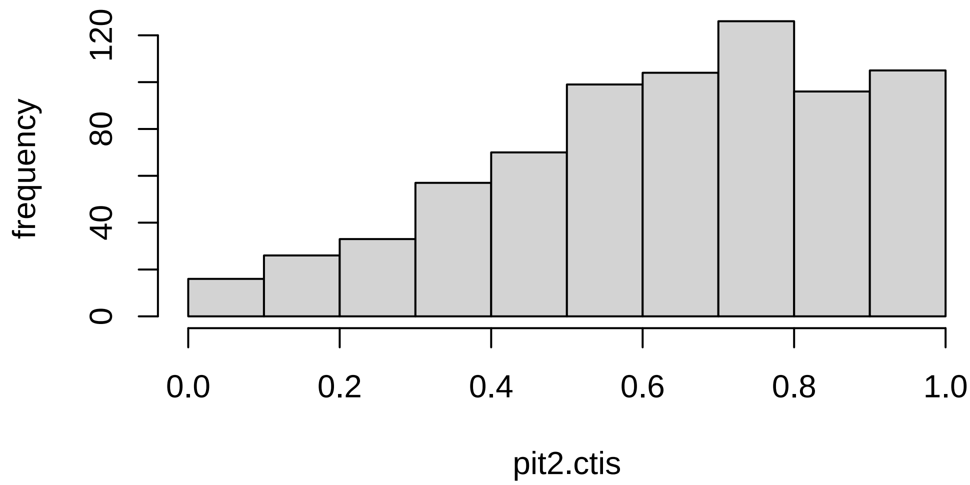
```
[1] -1700.385
```

A histogram of PIT must resemble a uniform distribution;extreme values indicate outlying observations. Both models outputs doesn't show a noral distributions - skewed to the right

```
# probability integral transform (PIT)
pit.ctis <- modelc$cpo$pit

hist(pit.ctis,main =" ", ylab="frequency")
```




```
pit2.ctis <- modelc2$cpo$pit

hist(pit2.ctis,main =" ", ylab="frequency")
```



Final Results - Modelc2

```
resultsc = exp(summary(modelc2)[[3]])[,c(1,3,5)] %>%
  round(3) %>%
  as.data.frame()

colnames(resultsc) = c("Incidence rate ratio", "2.5th quantile, IRR", "97.5th quantile, IR
rownames(resultsc) = c("Intercept",
                      "Months since study beginning",
                      "Intervention" ,
                      "Months since MTAT",
                      "Post period",
                      "1-month lagged EVI",
                      "1-month lagged rainfall",
                      "1-month lagged night temp",
                      "1-month lagged day temp",
                      "Feb",
                      "Mar",
                      "Apr",
                      "May",
                      "Jun",
                      "Jul",
```

```
                    "Aug",
                    "Sep",
                    "Oct",
                    "Nov",
                    "Dec",
                    "time:groupIntervention",
                    "groupIntervention:time_post",
                    "groupIntervention:post")

kableExtra::kable(resultsc, booktabs = T)
```

	Incidence rate ratio	2.5th quantile, IRR	97.5th quantile, IRR
Intercept	0.001	0.000	0.031
Months since study beginning	0.818	0.717	0.933
Intervention	0.929	0.224	3.857
Months since MTAT	1.194	1.040	1.370
Post period	4.233	1.419	12.897
1-month lagged EVI	0.090	0.007	1.077
1-month lagged rainfall	1.001	0.999	1.004
1-month lagged night temp	0.961	0.888	1.040
1-month lagged day temp	1.049	0.996	1.104
Feb	0.868	0.517	1.455
Mar	0.786	0.437	1.413
Apr	0.678	0.341	1.349
May	0.896	0.440	1.822
Jun	1.806	0.898	3.644
Jul	2.812	1.426	5.568
Aug	4.071	1.664	10.034
Sep	4.874	1.889	12.718
Oct	11.670	4.963	27.605
Nov	6.966	3.408	14.282
Dec	3.287	1.931	5.596
time:groupIntervention	1.114	1.006	1.232
groupIntervention:time_post	0.903	0.815	1.000
groupIntervention:post	0.081	0.033	0.198

```
counterfactualc = c_data %>%
  bind_cols(modelc2$summary.fitted.values) %>% #Join predicted case counts to data
  filter(post==1) %>% #Restrict to the post period
```

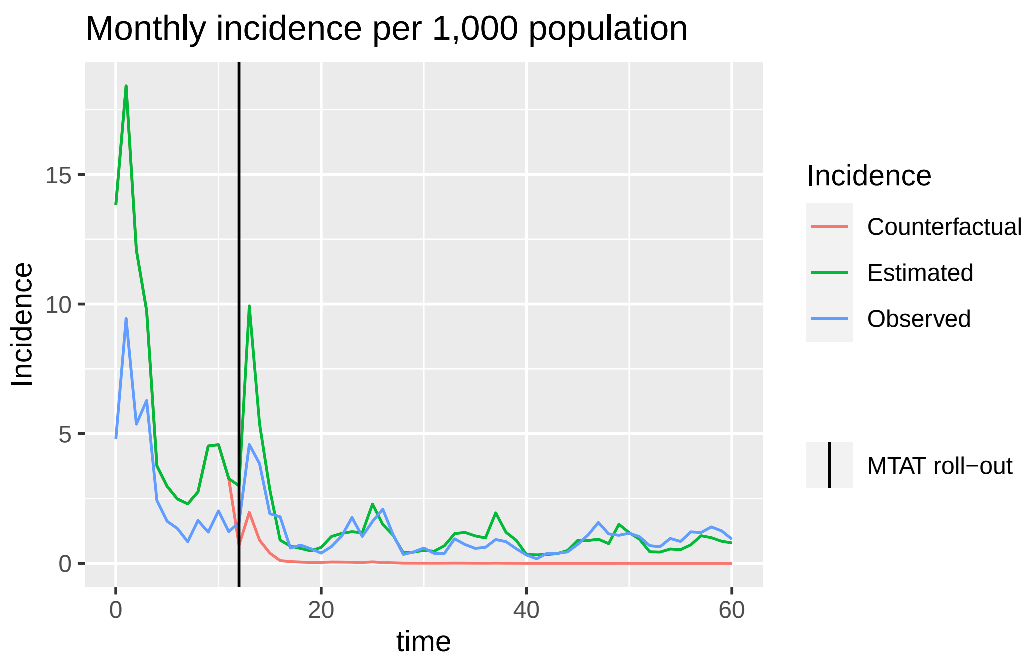
```r
  mutate(log_mean = log(mean), #Convert estimates back to log scale
         counterfactual_cases = exp(log_mean -
                                      summary(modelc2)[[3]]["time_post","mean"] * time_pos
                                      summary(modelc2)[[3]]["post","mean"] * post)) %>%
  summarise(counterfactual_cases = sum(counterfactual_cases),
            estimated_cases = sum(mean))

counterfactualc$counterfactual_cases - counterfactualc$estimated_cases #Cases decreased
```

```
[1] -3934.226
```

```r
(counterfactualc$counterfactual_cases - counterfactualc$estimated_cases)/
  counterfactualc$estimated_cases * 100 #Percent reduction (92%)
```

```
[1] -92.9039
```

```r
ggplot(c_data %>%
         bind_cols(modelc2$summary.fitted.values) %>% #Join predicted case counts to data
         mutate(log_mean = log(mean), #Convert estimates back to log scale
                counterfactual_cases = exp(log_mean -
                                      summary(modelc2)[[3]]["time_post","mean"] * time_pos
                                      summary(modelc2)[[3]]["post","mean"] * post)) %>%
         group_by(time) %>%
         summarise(incidence = sum(cases)/sum(population)*1000,
                   estimated_incidence = sum(mean)/sum(population)*1000,
                   counterfactual_incidence = sum(counterfactual_cases)/sum(population)*10
  geom_line(aes(x = time, y = counterfactual_incidence, color = "Counterfactual")) +
  geom_line(aes(x = time, y = estimated_incidence, color = "Estimated")) +
  geom_line(aes(x = time, y = incidence, color = "Observed")) +
  geom_vline(aes(xintercept = 12, linetype = "MTAT roll-out")) +
  labs(title = "Monthly incidence per 1,000 population", linetype = "", color = "Incidence
       y = "Incidence")
```



Monthly incidence per 1,000 population
Incidence
15
10
5
0
0
20
40
60
time
Incidence
Counterfactual
Estimated
Observed
MTAT roll−out