

2023 554 R Notes on Prevalence Mapping

Jon Wakefield
Departments of Biostatistics and Statistics
University of Washington

2023-01-15

Overview and Preliminaries

Overview

In these notes we analyze Demographic Health Survey (DHS) data on HIV prevalence in women age 15-29 in Malawi.

The raw data can be requested from the DHS. Here we use data summaries.

We will use the **SUMMER** package for the computation and plotting.

We will obtain:

- Direct estimates (area-level)
- Smoothed direct estimates (area-level)
- Betabinomial cluster-level estimates (unit-level)

Libraries

```
# Install necessary packages
library(tidyverse)
# install.packages('INLA', repos=c(getOption('repos'),
# INLA='https://inla.r-inla-download.org/R/testing'), dep=TRUE)
library(INLA)
# devtools::install_github('richardli/SUMMER', build_vignettes = F,
# force = T)
library(SUMMER)
library(spdep)
library(mapproj)
library(ggpubr)
library(sf)
```

Read in the data

```
# read in direct estimates and shape files must have a folder named
# R-examples in your current working directory
if (!("R-examples" %in% list.files())) {
  dir.create("R-examples")
}
```

```

download.file("http://faculty.washington.edu/jonno/SISMIDmaterial/direct.csv",
  destfile = "R-examples/direct.csv")
direct_df <- read.csv("R-examples/direct.csv")
download.file("http://faculty.washington.edu/jonno/SISMIDmaterial/gadm36_MWI_1.shx",
  destfile = "R-examples/gadm36_MWI_1.shx")
download.file("http://faculty.washington.edu/jonno/SISMIDmaterial/gadm36_MWI_1.shp",
  destfile = "R-examples/gadm36_MWI_1.shp")
download.file("http://faculty.washington.edu/jonno/SISMIDmaterial/gadm36_MWI_1.dbf",
  destfile = "R-examples/gadm36_MWI_1.dbf")
geo <- st_read("R-examples", layer = "gadm36_MWI_1")
## Reading layer `gadm36_MWI_1' from data source
##   `/Users/kpaulson/Repos/STAB/2023-554/R-examples' using driver `ESRI Shapefile'
## Simple feature collection with 28 features and 10 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: 32.67162 ymin: -17.12628 xmax: 35.91505 ymax: -9.364357
## CRS:            NA

```

Remove a small island

```

# filter out Likoma
geo <- geo[geo$NAME_1 != "Likoma", ]

# Make lowercase to match DHS data
geo$NAME_1 <- geo$NAME_1 %>%
  str_to_lower() %>%
  factor()
geo$id <- rownames(geo)
rownames(geo) <- geo$NAME_1

# create neighbor matrix
nb.r <- spdep::poly2nb(geo, queen = F)
mat <- spdep::nb2mat(nb.r, style = "B", zero.policy = TRUE)
colnames(mat) <- rownames(mat)

```

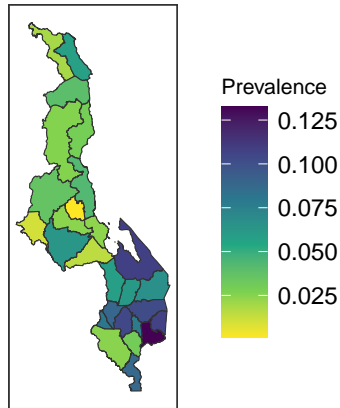
Direct Estimates

Map of direct estimates

```

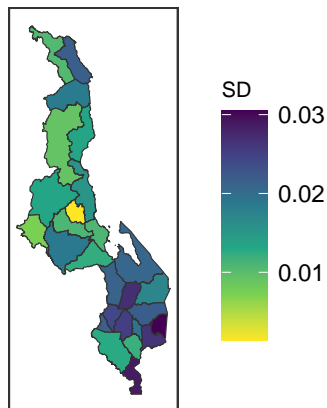
mapPlot(direct_df, variables = "p", geo = geo, by.data = "region", by.geo = "NAME_1",
  direction = -1, legend.label = "Prevalence", removetab = TRUE)

```



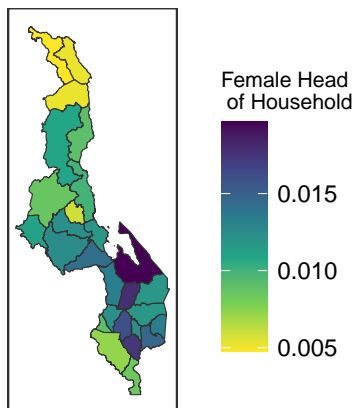
Map of uncertainty of direct estimates

```
mapPlot(direct_df, variables = "p_sd", geo = geo, by.data = "region", by.geo = "NAME_1",
        direction = -1, legend.label = "SD", removetab = TRUE)
```



Map of covariate, which is percent of households in that region that have a female head of household

```
mapPlot(direct_df, variables = "hoh", geo = geo, by.data = "region", by.geo = "NAME_1",
        direction = -1, legend.label = "Female Head \n of Household", removetab = TRUE)
```



Smoothed Direct

Smoothed direct estimates using SUMMER

Obtain smoothed direct estimates using SUMMER - BYM2 model for space.

```
# add variance columns to dataframe
dat <- direct_df %>%
  mutate(logit_p_var = logit_p_sd^2,
         p_var = p_sd^2)

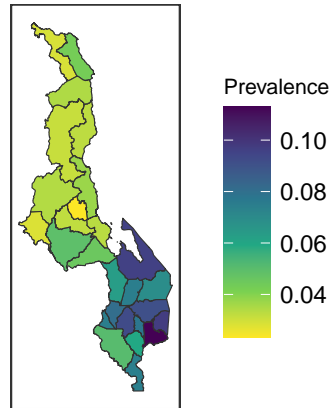
# get smoothed direct estimates from INLA
smoothed <- smoothSurvey(data = NULL, direct.est = dat,
                        X = dat %>% select(region, hoh), # include covariate
                        Amat = mat, regionVar = "region",
                        responseVar = "p", direct.est.var = "p_var",
                        responseType = "binary")
```

Summaries from the smoothed direct model

```
head(smoothed$HT, 3)
##      region      HT.est      HT.var HT.logit.est HT.logit.var HT.logit.prec
## 1  balaka 0.06185091 0.0007240332 -2.719182 0.21504163 4.650262
## 2 blantyre 0.10403239 0.0006399496 -2.153202 0.07365869 13.576130
## 3 chikwawa 0.02508902 0.0001642104 -3.659916 0.27447522 3.643316
##      hoh
## 1 0.017845042
## 2 0.016037949
## 3 0.007228371
smoothed$fit$summary.fixed[, 1:5]
##      mean      sd 0.025quant 0.5quant 0.975quant
## (Intercept) -3.159197 0.2798126 -3.707993 -3.159637 -2.60816
## hoh          23.862507 22.9587898 -21.894862 24.103983 68.27413
smoothed$fit$summary.hyperpar[, 1:5]
##      mean      sd 0.025quant 0.5quant 0.975quant
## Precision for region.struct 7.0260870 4.2205989 2.1427453 5.9887683 18.0554369
## Phi for region.struct      0.5486851 0.2357308 0.1174165 0.5557728 0.9412361
```

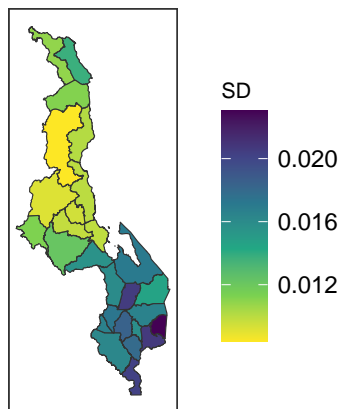
Map of median of smoothed direct estimates

```
mapPlot(smoothed$smooth, variables = "median", geo = geo, by.data = "region",
        by.geo = "NAME_1", direction = -1, legend.label = "Prevalence", removetab = TRUE)
```



Map of uncertainty of smoothed direct estimates

```
smoothed$smooth$sd <- smoothed$smooth$var %>%
  sqrt()
mapPlot(smoothed$smooth, variables = "sd", geo = geo, by.data = "region",
  by.geo = "NAME_1", direction = -1, legend.label = "SD", removetab = TRUE)
```



Betabinomial Model

Betabinomial model: load the data

```
# read in cluster-level data and urban/rural proportions at area
# level
download.file("http://faculty.washington.edu/jonno/SISMIDmaterial/urban_prop_sample_df.csv",
  destfile = "R-examples/urban_prop_sample_df.csv")
download.file("http://faculty.washington.edu/jonno/SISMIDmaterial/binom.csv",
  destfile = "R-examples/binom.csv")
binom_df <- read.csv("R-examples/binom.csv")
urb_prop_df <- read.csv("R-examples/urban_prop_sample_df.csv", stringsAsFactors = FALSE)
binom_df$strata_within_region <- ifelse(binom_df$urban == 1, "urban", "rural")
weights <- urb_prop_df %>%
  mutate(urban = prop, rural = 1 - prop) %>%
  select(region, urban, rural)
```

Sort out the covariate and fit the model

```
# create covariate data frame for input to smoothSurvey() covariate
# is at the area level
X <- binom_df %>%
  distinct(region, .keep_all = TRUE) %>%
  dplyr::select(region, head_house)

bb <- smoothSurvey(data = binom_df, Amat = mat, X = X, is.unit.level = TRUE,
  is.agg = TRUE, responseType = "binary", responseVar = "y", regionVar = "region",
  weightVar = NULL, strataVar.within = "strata_within_region", clusterVar = "clustid",
  totalVar = "n", weight.strata = weights, nsim = 1000)

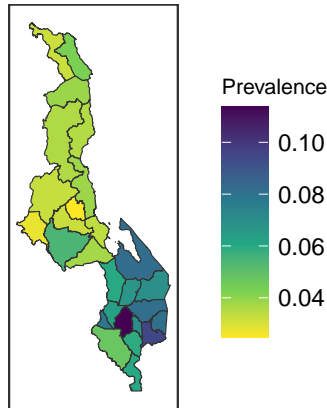
bb$smooth.overall$sd <- bb$smooth.overall$var %>%
  sqrt()
```

Summaries from the betabinomial model

```
bb$fit$summary.fixed[, 1:5]
##               mean               sd 0.025quant  0.5quant  0.975quant
## strata0rural -3.250947  0.2633965  -3.767107  -3.251407  -2.732262
## strata0urban -2.431319  0.2782089  -2.975874  -2.431956  -1.883251
## head_house   19.204451 21.2787013 -23.078562  19.378474  60.516784
bb$fit$summary.hyperpar[, 1:5]
##               mean               sd
## overdispersion for the betabinomial observations 0.0196624 0.008996178
## Precision for region.struct                      9.7131734 5.516713206
## Phi for region.struct                          0.6422870 0.231923448
##               0.025quant  0.5quant
## overdispersion for the betabinomial observations 0.006421162 0.01826948
## Precision for region.struct                      3.145952294 8.39839626
## Phi for region.struct                          0.165913679 0.67502358
##               0.975quant
## overdispersion for the betabinomial observations 0.04097395
## Precision for region.struct                      24.04824307
## Phi for region.struct                          0.97309179
```

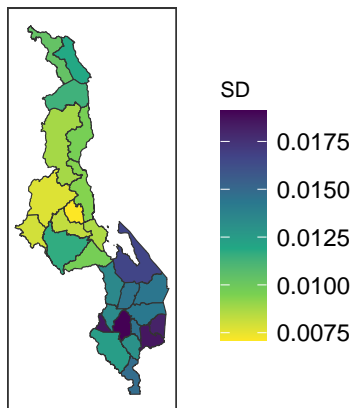
Betabinomial estimates map

```
mapPlot(bb$smooth.overall, variables = "median", geo = geo, by.data = "region",
  by.geo = "NAME_1", direction = -1, legend.label = "Prevalence", removetab = TRUE)
```



Betabinomial uncertainty map

```
mapPlot(bb$smooth.overall, variables = "sd", geo = geo, by.data = "region",
        by.geo = "NAME_1", direction = -1, legend.label = "SD", removetab = TRUE)
```



Combine point estimates

```
# define prevalence plots
direct_prev <- mapPlot(direct_df, variables = "p", geo = geo, by.data = "region",
                      by.geo = "NAME_1", direction = -1, legend.label = "Prevalence", removetab = TRUE)
smooth_prev <- mapPlot(smoothed$smooth, variables = "median", geo = geo,
                      by.data = "region", by.geo = "NAME_1", direction = -1, legend.label = "Prevalence",
                      removetab = TRUE)
bb_prev <- mapPlot(bb$smooth.overall, variables = "median", geo = geo,
                  by.data = "region", by.geo = "NAME_1", direction = -1, legend.label = "Prevalence",
                  removetab = TRUE)
```

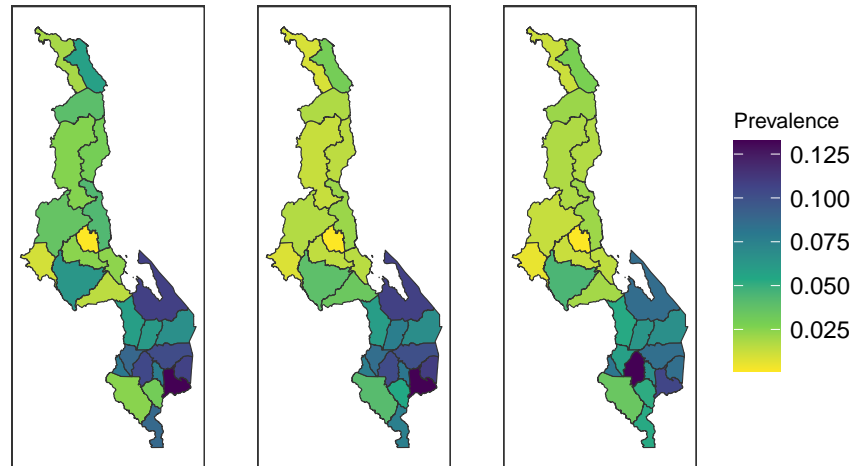
Combine uncertainty estimates

```
# define sd plots
direct_sd <- mapPlot(direct_df, variables = "p_sd", geo = geo, by.data = "region",
                    by.geo = "NAME_1", direction = -1, legend.label = "SD", removetab = TRUE)
smooth_sd <- mapPlot(smoothed$smooth, variables = "sd", geo = geo, by.data = "region",
                    by.geo = "NAME_1", direction = -1, legend.label = "SD", removetab = TRUE)
bb_sd <- mapPlot(bb$smooth.overall, variables = "sd", geo = geo, by.data = "region",
```

```
by.geo = "NAME_1", direction = -1, legend.label = "SD", removetab = TRUE)
```

Comparison of point estimates across methods

```
ggarrange(direct_prev, smooth_prev, bb_prev, nrow = 1, common.legend = TRUE,  
  legend = "right")
```



Comparison of uncertainty across methods

```
ggarrange(direct_sd, smooth_sd, bb_sd, nrow = 1, common.legend = TRUE,  
  legend = "right")
```

