Patients with presumed tuberculosis in sub-Saharan Africa that are not diagnosed with tuberculosis: a systematic review and meta-analysis (statistical appendix)

S Jayasooriya,   F Dimambro-Denson,   C Beecroft,   J Balen,   B Awokola,
C Mitchell,   B Kampmann,   F Campbell,   PJ Dodd,   K Mortimer

August, 2021

Contents

Pre-amble

This document is generated from an R script in literate programming fashion. All R code is quoted in this document, together with output (preceeded by '##') and figures. The article forest plot is saved to the `output` folder but not included in the document since it is too cramped. The script and data are publicly available on GitHub at https://github.com/petedodd/NotTB and once the repository is downloaded, it should be possible to generate this document using R with the command `rmarkdown::render('NotTBmeta.R')` within R, or from a unix-like command line with `R -q -e "rmarkdown::render(\"NotTBmeta.R\",output_dir=\"./output\")"`. Alternatively, the R script can be run in whole or part as a conventional R script.

Dependencies

To compile this document, the rmarkdown & knitr packages must be installed. The other R packages required to run this analysis should be installed if necessary, and loaded, with:

```
pkgs.needed <- c("ggplot2","scales","cowplot","ggpubr", #graphs
                 "data.table","here",                   #data mgt
                 "metafor")                              #metaanalysis
```

1

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

Thorax

```
install.packages(setdiff(pkgs.needed, rownames(installed.packages())))
suppressMessages(
    devnull <- lapply(pkgs.needed, require, character.only = TRUE) #load for use
)
```

This analysis was run using:

```
sI <- sessionInfo()
dI <- data.frame(
    item=c('R version','platform','OS','metafor version'),
    version=c(
        sI$R.version$version.string, #R version
        sI$platform,                 #platofm
        sI$running,                  #OS
        sI$otherPkgs$metafor$Version #metafor version
    )
)
knitr::kable(dI)
```

item	version
R version	R version 4.1.0 (2021-05-18)
platform	x86_64-pc-linux-gnu (64-bit)
OS	Pop!_OS 21.04
metafor version	3.0-2

Main analyses

Approach

We use a generalized linear mixed effects (GLMM) approach to meta-analysis assuming a binomial response and logit link[1]. This means we assume

$$k_i \sim \text{Binomial}(N_i, p_i)$$
$$\text{logit}(p_i) = \mu + \varepsilon_i$$
$$\varepsilon_i \sim \mathcal{N}(0, \tau^2)$$

where $i = 1, \ldots, S$ indexes the numbers of studies.

Use of arcsine or double arcsine transformations has been criticized in this context, with the GLMM approach recommended instead.[2]

Read in the data and ensure that factors behave as intended:

```
DD <- fread(file=here('SRMAdata.csv'))
DD[,lab:=factor(lab,levels=rev(DD[order(bac)]$lab),ordered = TRUE)]
```

Create exact binomial confidence intervals:

---

[1] Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data.

[2] Schwarzer G, Chemaitelly H, Abu-Raddad LJ, Rücker G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

Thorax

```
ciz <- function(x,y){
    x <- as.integer(x); y <- as.integer(y)
    list(binom.test(x,y)$conf.int[1],binom.test(x,y)$conf.int[2])
}
DD[,`NotTB Proportion`:=NnotTB/N]
for(i in 1:nrow(DD)){ DD[i,c('lo','hi'):=ciz(NnotTB,N)]; }
DD[,SE:=(hi-lo)/3.92]
```

Meta-analyses

Meta-analysis for passively found TB patients with bacteriologically unconfirmed TB included:

```
maPU <- rma.glmm(measure = "PLO", #  binomial w/ logit link
            xi = NnotTB,      # numerator
            ni = N,           # denominator
            data = DD[mode=='Passive' &
                    clinical=='(Unconfirmed TB included)'],
            slab = Author)      # what to use as labels on graphs

## Registered S3 methods overwritten by 'lme4':
##   method                          from
##   cooks.distance.influence.merMod car
##   influence.merMod                car
##   dfbeta.influence.merMod         car
##   dfbetas.influence.merMod        car

summary(maPU)

##
## Random-Effects Model (k = 8; tau^2 estimator: ML)
##
##   logLik  deviance      AIC       BIC      AICc
## -25.7259    0.4121  55.4518   55.6107   57.8518
##
## tau^2 (estimated amount of total heterogeneity): 0.2977
## tau (square root of estimated tau^2 value):      0.5457
## I^2 (total heterogeneity / total variability):   97.0524%
## H^2 (total variability / sampling variability):  33.9255
##
## Tests for Heterogeneity:
## Wld(df = 7) = 221.8886, p-val < .0001
## LRT(df = 7) = 243.5648, p-val < .0001
##
## Model Results:
##
## estimate      se     zval     pval    ci.lb    ci.ub
##  -0.0619  0.1971  -0.3140   0.7535  -0.4482   0.3244
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
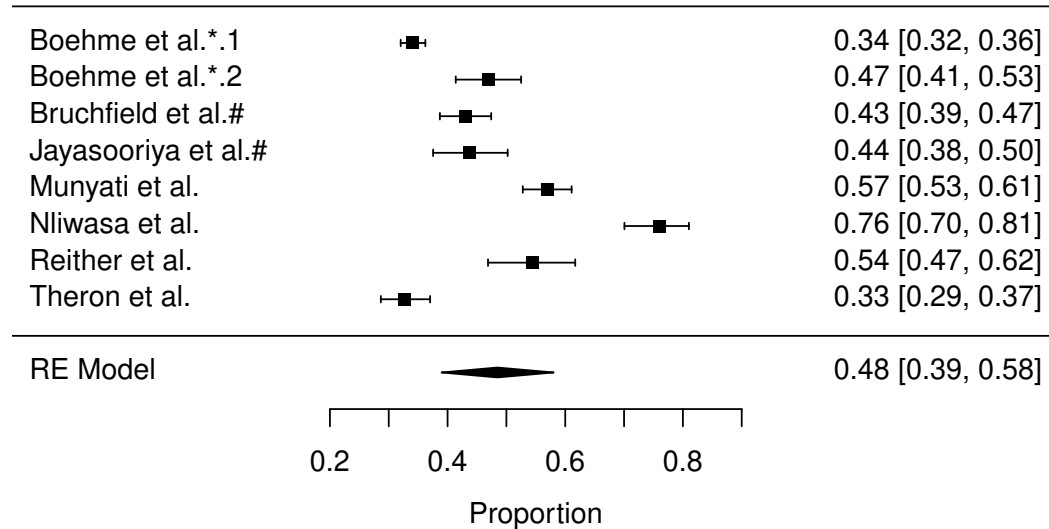
Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

Thorax

```
forest(maPU,transf = transf.ilogit,refline=NA)
```


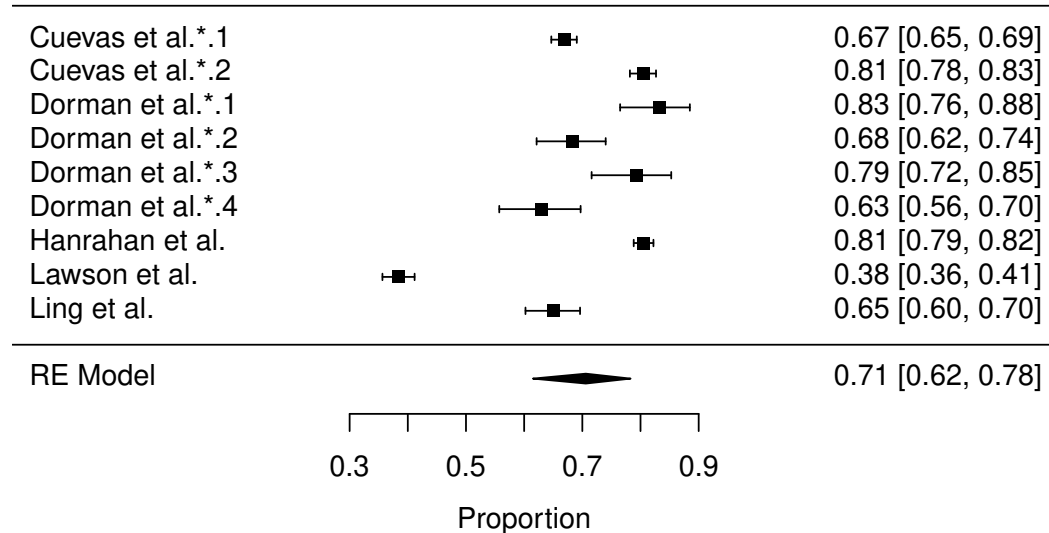|	
Boehme et al.*.1	0.34 [0.32, 0.36]
Boehme et al.*.2	0.47 [0.41, 0.53]
Bruchfield et al.#	0.43 [0.39, 0.47]
Jayasooriya et al.#	0.44 [0.38, 0.50]
Munyati et al.	0.57 [0.53, 0.61]
Nliwasa et al.	0.76 [0.70, 0.81]
Reither et al.	0.54 [0.47, 0.62]
Theron et al.	0.33 [0.29, 0.37]
RE Model	0.48 [0.39, 0.58]

Proportion

Meta-analysis for passively found TB patients with bacteriologically unconfirmed TB excluded:

```
maPN <- rma.glmm(measure = "PLO", #  binomial w/ logit link
          xi = NnotTB,     # numerator
          ni = N,          # denominator
          data = DD[mode=='Passive' &
                    clinical=='(No unconfirmed TB)'],
          slab = Author)        # what to use as labels on graphs
summary(maPN)
```

```
##
## Random-Effects Model (k = 9; tau^2 estimator: ML)
##
##   logLik  deviance       AIC       BIC      AICc
## -28.8910    0.2865   61.7821   62.1765   63.7821
##
## tau^2 (estimated amount of total heterogeneity): 0.3714
## tau (square root of estimated tau^2 value):      0.6094
## I^2 (total heterogeneity / total variability):   98.1427%
## H^2 (total variability / sampling variability):  53.8403
##
## Tests for Heterogeneity:
## Wld(df = 8) = 679.9414, p-val < .0001
## LRT(df = 8) = 727.2051, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.8757  0.2078  4.2139  <.0001  0.4684  1.2830   ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
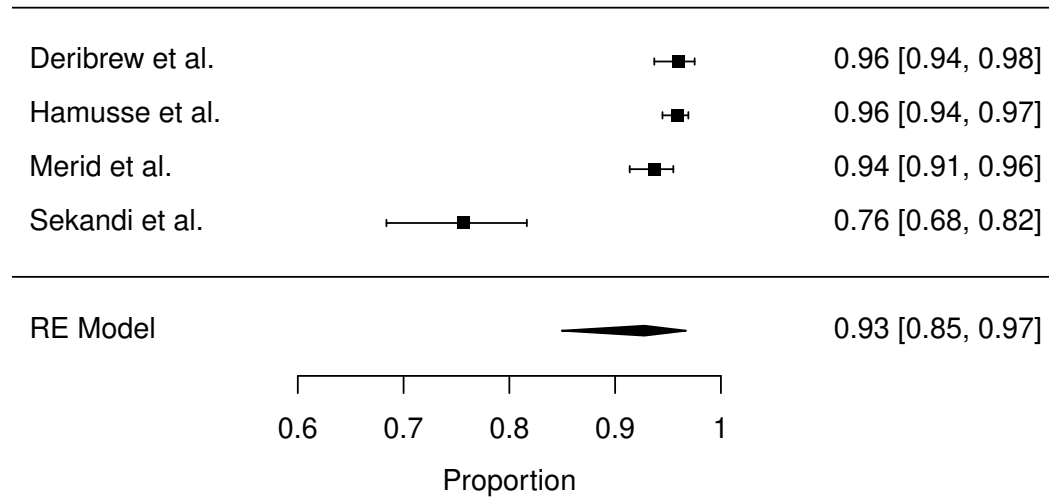
```
forest(maPN,transf = transf.ilogit,refline=NA)
```


Cuevas et al.*.1	0.67 [0.65, 0.69]
Cuevas et al.*.2	0.81 [0.78, 0.83]
Dorman et al.*.1	0.83 [0.76, 0.88]
Dorman et al.*.2	0.68 [0.62, 0.74]
Dorman et al.*.3	0.79 [0.72, 0.85]
Dorman et al.*.4	0.63 [0.56, 0.70]
Hanrahan et al.	0.81 [0.79, 0.82]
Lawson et al.	0.38 [0.36, 0.41]
Ling et al.	0.65 [0.60, 0.70]
RE Model	0.71 [0.62, 0.78]

Meta-analysis for actively found TB patients:

```
maA <- rma.glmm(measure = "PLO", #  binomial w/ logit link
           xi = NnotTB,     # numerator
           ni = N,          # denominator
           data = DD[mode=='Active'],
           slab = Author)        # what to use as labels on graphs
summary(maA)


##
## Random-Effects Model (k = 4; tau^2 estimator: ML)
##
##   logLik  deviance      AIC       BIC      AICc
## -10.4692    0.2060   24.9385   23.7111   36.9385
##
## tau^2 (estimated amount of total heterogeneity): 0.6678
## tau (square root of estimated tau^2 value):      0.8172
## I^2 (total heterogeneity / total variability):   95.0642%
## H^2 (total variability / sampling variability):  20.2600
##
## Tests for Heterogeneity:
## Wld(df = 3) = 81.2135, p-val < .0001
## LRT(df = 3) = 67.4266, p-val < .0001
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   2.5537  0.4199  6.0817  <.0001  1.7307  3.3767   ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

forest(maA,transf = transf.ilogit,refline=NA)
```

5

Make predictions for plot data:

```
map <- predict(maA,transf = transf.ilogit)
mup <- predict(maPU,transf = transf.ilogit)
mnp <- predict(maPN,transf = transf.ilogit)
```

Creation of combined forest plot

Summary data for combined forest plot:

```
f1 <- function(x)format(round(x,1),nsmall=1)
cnz <- c('(Unconfirmed TB included)',
         '(No unconfirmed TB)',
         '(No unconfirmed TB)')
predz <- data.table(mode=c('Passive','Passive','Active'),
                    clinical=cnz,
                    `NotTB Proportion` = c(mup$pred,mnp$pred,map$pred),
                    lo = c(mup$ci.lb,mnp$ci.lb,map$ci.lb),
                    hi = c(mup$ci.ub,mnp$ci.ub,map$ci.ub),
                    lab=paste0('SUMMARY (',expression(I^2),'=',
                            f1(c(maA$I2,maPN$I2,maPU$I2)),'%)')
                    )
predz[,SE:=(hi-lo)/3.92]
predz[,qty:='summary']
predz[,bac:=0]
predz[,mid:=`NotTB Proportion`]
predz[,CI:=paste0(f1(1e2*mid),' (',f1(1e2*lo),' - ',f1(1e2*hi),')')]
predz[,wt:='100.0%']
predz[,w:=1]
```

Appending plot data to inputs:

```
DD[,qty:='study']
DD[,mid:=`NotTB Proportion`]
DD[,CI:=paste0(f1(1e2*mid),' (',f1(1e2*lo),' - ',f1(1e2*hi),')')]
DD[,wt:=1/SE^2]
DD[,wtt:=sum(wt),by=.(mode,clinical)]
DD[,wt:=1e2*wt/wtt]
```

```r
DD[,wt:=paste0(f1(wt),'%')]
DD[,w:=0]
```

Combined plot data:

```r
B <- rbind(
    DD[,.(lab,`NotTB Proportion`,lo,hi,SE,mode,clinical,
          qty,bac,CI,wt,w)],
    predz[,.(lab,`NotTB Proportion`,lo,hi,SE,mode,clinical,
             qty,bac,CI,wt,w)]
)
lbz <- as.character(B[order(bac)]$lab)
lbz2 <- c(lbz[1:3],rev(lbz[-c(1:3)]))
B[,lab:=factor(lab,levels=lbz2,ordered = TRUE)]
B[,clinical.g:='Clinically diagnosed tuberculosis included']
B[clinical=='(No unconfirmed TB)',
  clinical.g:='No clinically diagnosed tuberculosis included']
B[mode=='Active',clinical.g:='']
B[,mode:=paste0(mode,' case-finding')]
B[,mode:=factor(mode,levels=c('Passive case-finding',
                              'Active case-finding'),
                ordered = TRUE)]
B[,clinical.g:=factor(clinical.g,levels=unique(clinical.g))]
labdat <- B[1]
labdat[,txt:=' weight (%)']
labdat2 <- B[1]
labdat2[,txt:='prevalence (95% confidence interval)']
```

Create publication forest plot figure:

```r
SA <- ggplot(B,aes(lab,y=`NotTB Proportion`,
                   ymin=lo,
                   ymax=hi,
                   col=qty)) +
    geom_point(aes(size=1/SE^2,shape=qty)) +
    geom_errorbar(aes(width=w/2)) +
    scale_y_continuous(label=percent,limits = c(0,NA))+
    scale_color_manual(values=c('study'="black",'summary'="blue"))+
    scale_shape_manual(values=c('study'=22,'summary'=23))+
    xlab('') +
    ylab('Proportion of patients with presumptive tuberculosis not diagnosed as tuberculosis')+
    facet_grid(mode + clinical.g ~ .,
               scales = 'free',space='free',
               switch='x'
               )+
    coord_flip() +
    guides(size='none',color='none',shape='none')+
    theme_classic() +
    theme(panel.spacing = unit(2, "lines"), #or 3
          strip.background = element_blank(),
          strip.placement = "outside") +
    geom_text(aes(x=lab,y=1.2,label=CI,hjust='right')) +
    geom_text(aes(x=lab,y=0.0,label=wt))+
    geom_text(data=labdat,aes(x=9.5,y=0,label=txt))+
    geom_text(data=labdat2,aes(x=9.5,y=1,label=txt))+
```

7

```
        ggpubr::grids()

ggsave(SA,file=here('output/ForestPlot.pdf'),h=13,w=12)
ggsave(SA,file=here('output/ForestPlot.eps'),h=13,w=12)
```

Meta-regressions

In this section we consider various potential sources of heterogeneity through scatter plots and meta-regression.
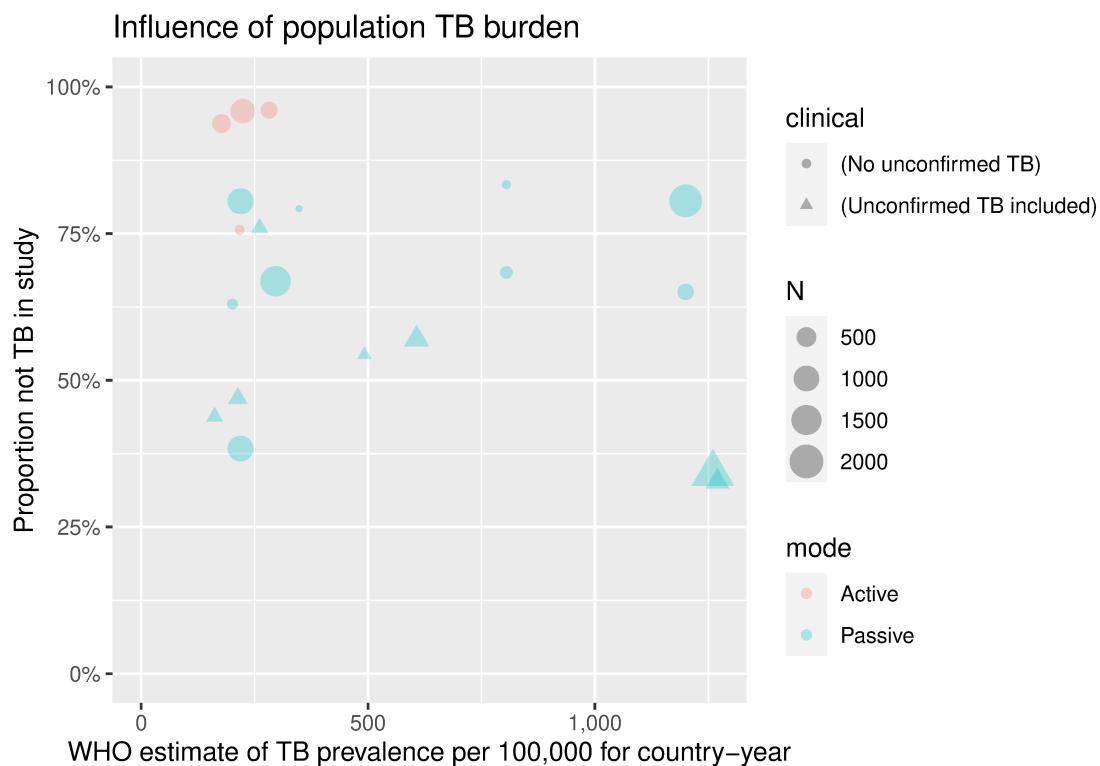
TB prevalence

The burden of TB in a population might reasonably be expected to influence the proportion of presumptive TB that is not TB.

```
DD[,tb:=`WHO TB estimate (per 100 000 year of study)`]
a <- 0.3
ggplot(DD,aes(tb,`NotTB Proportion`,
            size=N,col=mode,shape=clinical))+
    scale_x_continuous(label=comma,limits=c(0,NA))+
    scale_y_continuous(label=percent,limits=c(0,1))+
    geom_point(alpha=a)+
    xlab('WHO estimate of TB prevalence per 100,000 for country-year')+
    ylab('Proportion not TB in study')+
    ggtitle('Influence of population TB burden')
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

We can formally investigating the influence of TB burden in explaining heterogeneity with a meta-regression:

```
tbmr <-  rma.glmm(measure = "PLO",  #binomial w/ logit link
                  xi = NnotTB,     # numerator
                  ni = N,          # denominator
                  data = DD,        # what data to use
                  mods = ~mode*clinical + tb)
```
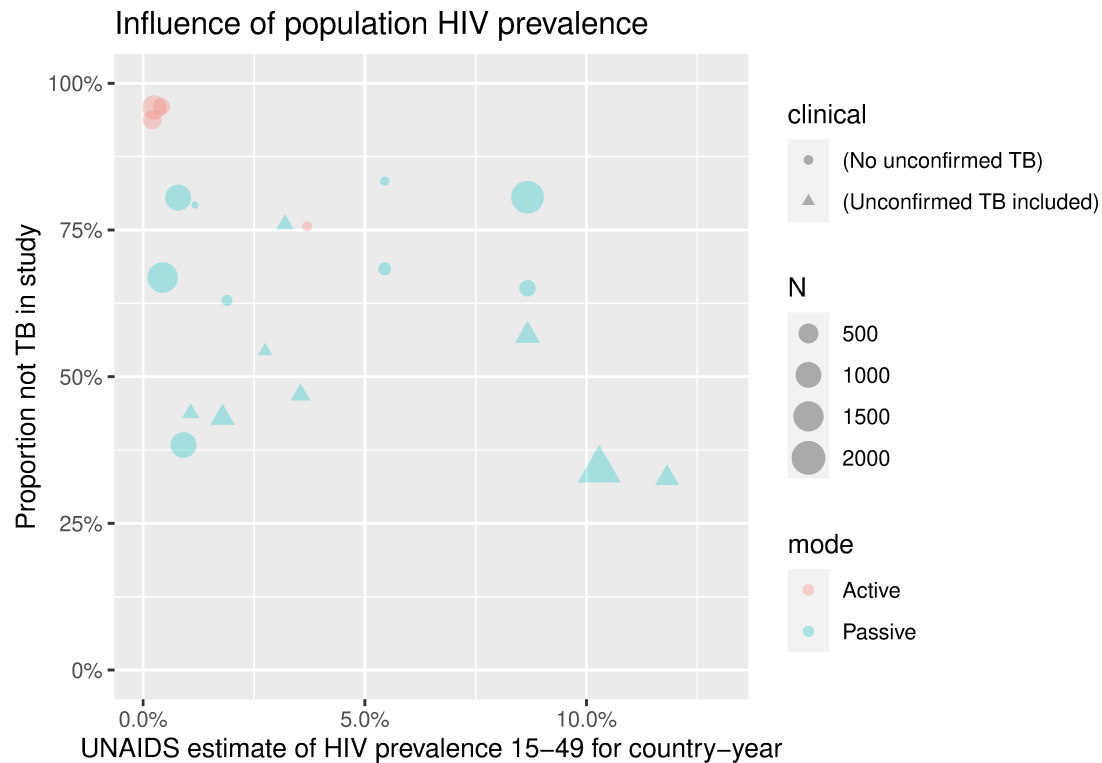
```
## Warning: Studies with NAs omitted from model fitting.

## Warning: Some yi/vi values are NA.

## Warning: Redundant predictors dropped from the model.
```

```
summary(tbmr)
```

```
##
## Mixed-Effects Model (k = 20; tau^2 estimator: ML)
##
##   logLik  deviance      AIC       BIC      AICc
## -61.7991    0.9638  133.5982  138.5769  137.8839
##
## tau^2 (estimated amount of residual heterogeneity):     0.4095
## tau (square root of estimated tau^2 value):             0.6399
## I^2 (residual heterogeneity / unaccounted variability): 97.6536%
## H^2 (unaccounted variability / sampling variability):   42.6180
##
## Tests for Residual Heterogeneity:
## Wld(df = 16) =  973.5088, p-val < .0001
## LRT(df = 16) = 1028.1407, p-val < .0001
##
## Test of Moderators (coefficients 2:4):
## QM(df = 3) = 38.8326, p-val < .0001
##
## Model Results:
##
##                                   estimate      se     zval    pval     ci.lb
## intrcpt                             2.5877  0.3453   7.4931  <.0001    1.9109
## modePassive                        -1.6174  0.4233  -3.8210  0.0001   -2.4471
## clinical(Unconfirmed TB included)  -0.8999  0.3286  -2.7386  0.0062   -1.5439
## tb                                 -0.0002  0.0004  -0.4084  0.6830   -0.0009
##
##                                      ci.ub
## intrcpt                             3.2646   ***
## modePassive                        -0.7878   ***
## clinical(Unconfirmed TB included)  -0.2559    **
## tb                                  0.0006
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

HIV prevalence

Population HIV prevalence may plausibly influence the proportion of presumptives not diagnosed with TB both by inluencing TB burden, but also by changing the typical clinical characteristics of TB and most importantly, the burden of other illness that could be designated presumptive TB.

```
ggplot(DD,aes(hiv/1e2,`NotTB Proportion`,
              size=N,col=mode,shape=clinical))+
    scale_x_continuous(label=percent,limits=c(0,0.13))+
    scale_y_continuous(label=percent,limits=c(0,1))+
    geom_point(alpha=a)+
    xlab('UNAIDS estimate of HIV prevalence 15-49 for country-year')+
    ylab('Proportion not TB in study')+
    ggtitle('Influence of population HIV prevalence')
```


Influence of population HIV prevalence

We can formally investigating the influence of HIV in explaining heterogeneity with a meta-regression:

```
hivmr <-  rma.glmm(measure = "PLO",   #binomial w/ logit link
                   xi = NnotTB,       # numerator
                   ni = N,            # denominator
                   data = DD,         # what data to use
                   mods = ~mode*clinical + hiv)
```

```
## Warning: Redundant predictors dropped from the model.
```

```
summary(hivmr)
```

```
##
## Mixed-Effects Model (k = 21; tau^2 estimator: ML)
##
##   logLik  deviance       AIC       BIC      AICc
## -65.1479    1.0280  140.2958  145.5184  144.2958
##
## tau^2 (estimated amount of residual heterogeneity):    0.3839
```

10

```
## tau (square root of estimated tau^2 value):            0.6196
## I^2 (residual heterogeneity / unaccounted variability): 97.5586%
## H^2 (unaccounted variability / sampling variability):   40.9604
##
## Tests for Residual Heterogeneity:
## Wld(df = 17) =  973.1809, p-val < .0001
## LRT(df = 17) = 1025.3297, p-val < .0001
##
## Test of Moderators (coefficients 2:4):
## QM(df = 3) = 44.7803, p-val < .0001
##
## Model Results:
##
##                                   estimate      se     zval    pval    ci.lb
## intrcpt                             2.5888  0.3279   7.8949  <.0001   1.9461
## modePassive                        -1.5920  0.4019  -3.9609  <.0001  -2.3798
## clinical(Unconfirmed TB included)  -0.8801  0.3153  -2.7914  0.0052  -1.4981
## hiv                                -0.0325  0.0420  -0.7730  0.4395  -0.1149
##                                      ci.ub
## intrcpt                             3.2315  ***
## modePassive                        -0.8042  ***
## clinical(Unconfirmed TB included)  -0.2622   **
## hiv                                 0.0499
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
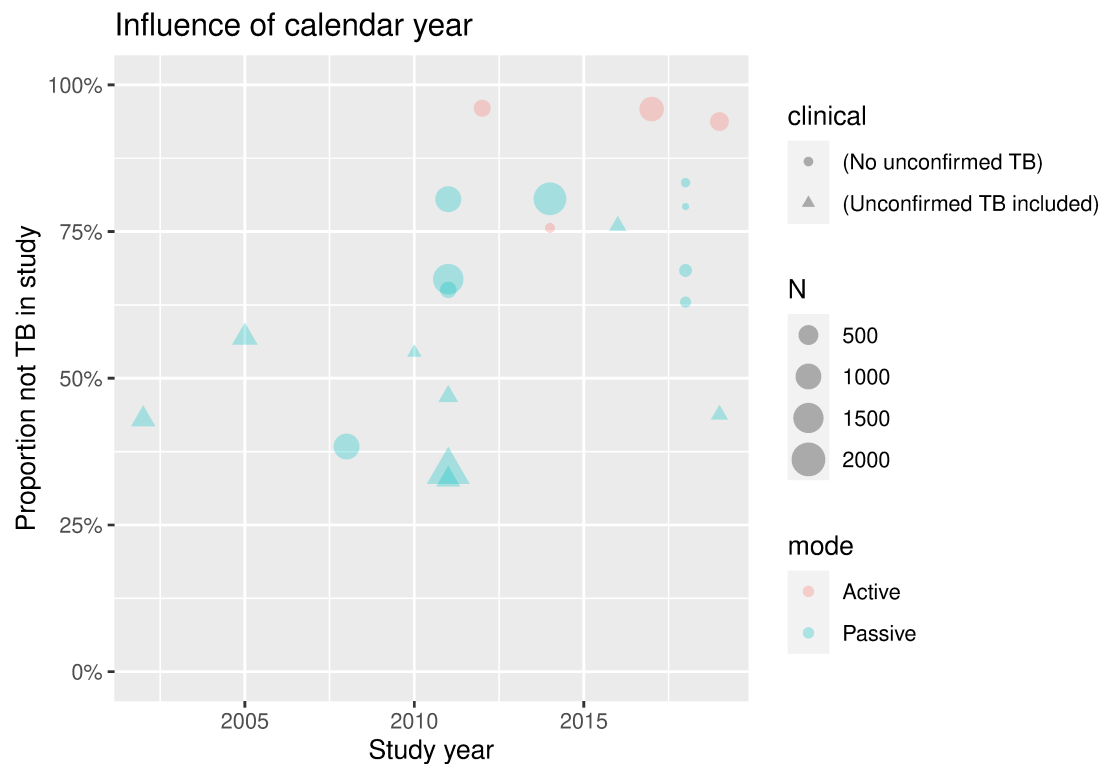
Calendar time

To explore whether there has been any change over time, we consider calendar year

```
ggplot(DD,aes(Year,`NotTB Proportion`,
              size=N,col=mode,shape=clinical))+
    scale_y_continuous(label=percent,limits=c(0,1))+
    geom_point(alpha=a)+
    xlab('Study year')+
    ylab('Proportion not TB in study')+
    ggtitle('Influence of calendar year')
```

11

Influence of calendar year

We can formally investigating the influence of year in explaining heterogeneity with a meta-regression:

```
yearmr <-  rma.glmm(measure = "PLO",  #binomial w/ logit link
              xi = NnotTB,     # numerator
              ni = N,          # denominator
              data = DD,       # what data to use
              mods = ~mode*clinical + Year)
```

```
## Warning: Redundant predictors dropped from the model.
```

```
summary(yearmr)
```

```
##
## Mixed-Effects Model (k = 21; tau^2 estimator: ML)
##
##   logLik  deviance       AIC       BIC      AICc
## -65.2094    1.1510  140.4188  145.6414  144.4188
##
## tau^2 (estimated amount of residual heterogeneity):    0.3586
## tau (square root of estimated tau^2 value):            0.5989
## I^2 (residual heterogeneity / unaccounted variability): 97.5232%
## H^2 (unaccounted variability / sampling variability):   40.3748
##
## Tests for Residual Heterogeneity:
## Wld(df = 17) = 882.4776, p-val < .0001
## LRT(df = 17) = 919.1171, p-val < .0001
##
## Test of Moderators (coefficients 2:4):
```

12

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

Thorax

```
## QM(df = 3) = 49.0787, p-val < .0001
##
## Model Results:
##
##                                   estimate        se      zval     pval
## intrcpt                           -88.8442   64.9689   -1.3675   0.1715
## modePassive                        -1.6045    0.3784   -4.2400   <.0001
## clinical(Unconfirmed TB included)  -0.7813    0.3167   -2.4673   0.0136
## Year                                0.0453    0.0322    1.4068   0.1595
##                                       ci.lb     ci.ub
## intrcpt                           -216.1809   38.4926
## modePassive                         -2.3462   -0.8628   ***
## clinical(Unconfirmed TB included)   -1.4019   -0.1606     *
## Year                                -0.0178    0.1085
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Sensitivity analyses

Dorman et al. by country only

In the main analysis, we considered the different sites in the 2018 study by Dorman et al to be separate data. This included considering the two sites in South Africa - Cape Town and Johannesburg - as different, which was motivated by the very distinct TB epidemiology in the Western Cape. Here we investigate the impact of aggregating the two South African sites in Dorman et al on the meta-analysis for studies with passive case finding excluding clinically diagnosed TB.

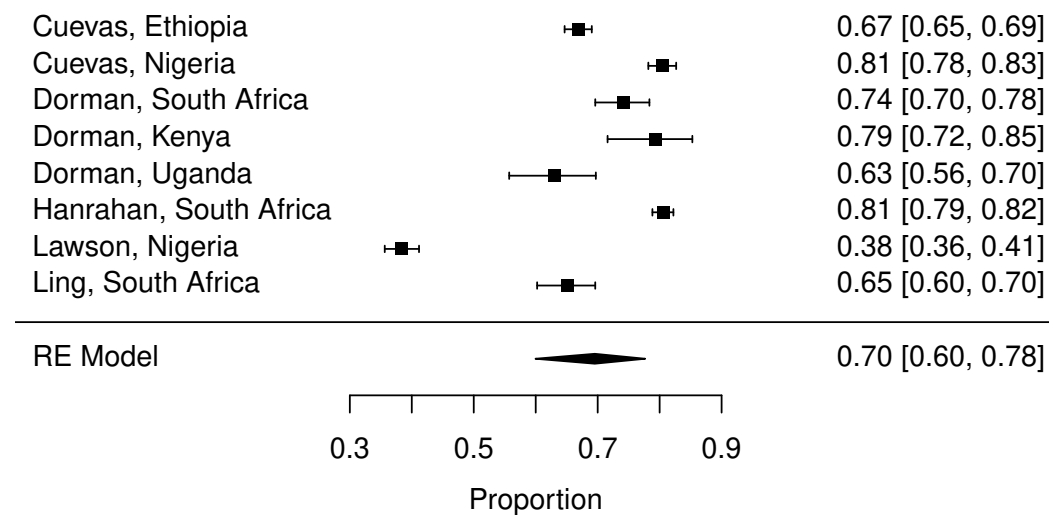Restrict to relevant data & aggregate over Dorman in South Africa:

```r
tmp <- DD[mode=='Passive' & clinical=='(No unconfirmed TB)']
tmp[,Country.Simple:=gsub(" \\-.+$","",Country)]                        #remove cities
tmp[,authorcountry:=paste(gsub("^([A-Za-z]+).*","\\1",Author),Country.Simple,sep = ", ")] #new label
tmp <- tmp[,.(NnotTB=sum(NnotTB),N=sum(N)),by=authorcountry]
knitr::kable(tmp) #check
```

authorcountry	NnotTB	N
Cuevas, Ethiopia	1184	1770
Cuevas, Nigeria	963	1196
Dorman, South Africa	285	384
Dorman, Kenya	107	135
Dorman, Uganda	114	181
Hanrahan, South Africa	1685	2091
Lawson, Nigeria	455	1186
Ling, South Africa	257	395

Rerun this meta-analysis with the new data:

```r
maPNsa <- rma.glmm(measure = "PLO", # binomial w/ logit link
          xi = NnotTB,     # numerator
          ni = N,          # denominator
          data =tmp,       # new data
          slab = authorcountry)    # what to use as labels on graphs
summary(maPNsa)
```

13

```
## 
## Random-Effects Model (k = 8; tau^2 estimator: ML)
## 
##    logLik  deviance       AIC       BIC      AICc
## -26.5760    0.1654   57.1519   57.3108   59.5519
## 
## tau^2 (estimated amount of total heterogeneity): 0.3563
## tau (square root of estimated tau^2 value):      0.5969
## I^2 (total heterogeneity / total variability):   98.3044%
## H^2 (total variability / sampling variability):  58.9761
## 
## Tests for Heterogeneity:
## Wld(df = 7) = 671.4861, p-val < .0001
## LRT(df = 7) = 716.0656, p-val < .0001
## 
## Model Results:
## 
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.8252  0.2149  3.8406  0.0001  0.4041  1.2463   ***
## 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
forest(maPNsa,transf = transf.ilogit,refline=NA)
```


Cuevas, Ethiopia		0.67 [0.65, 0.69]
Cuevas, Nigeria		0.81 [0.78, 0.83]
Dorman, South Africa		0.74 [0.70, 0.78]
Dorman, Kenya		0.79 [0.72, 0.85]
Dorman, Uganda		0.63 [0.56, 0.70]
Hanrahan, South Africa		0.81 [0.79, 0.82]
Lawson, Nigeria		0.38 [0.36, 0.41]
Ling, South Africa		0.65 [0.60, 0.70]
RE Model		0.70 [0.60, 0.78]

Proportion: 0.3 0.5 0.7 0.9

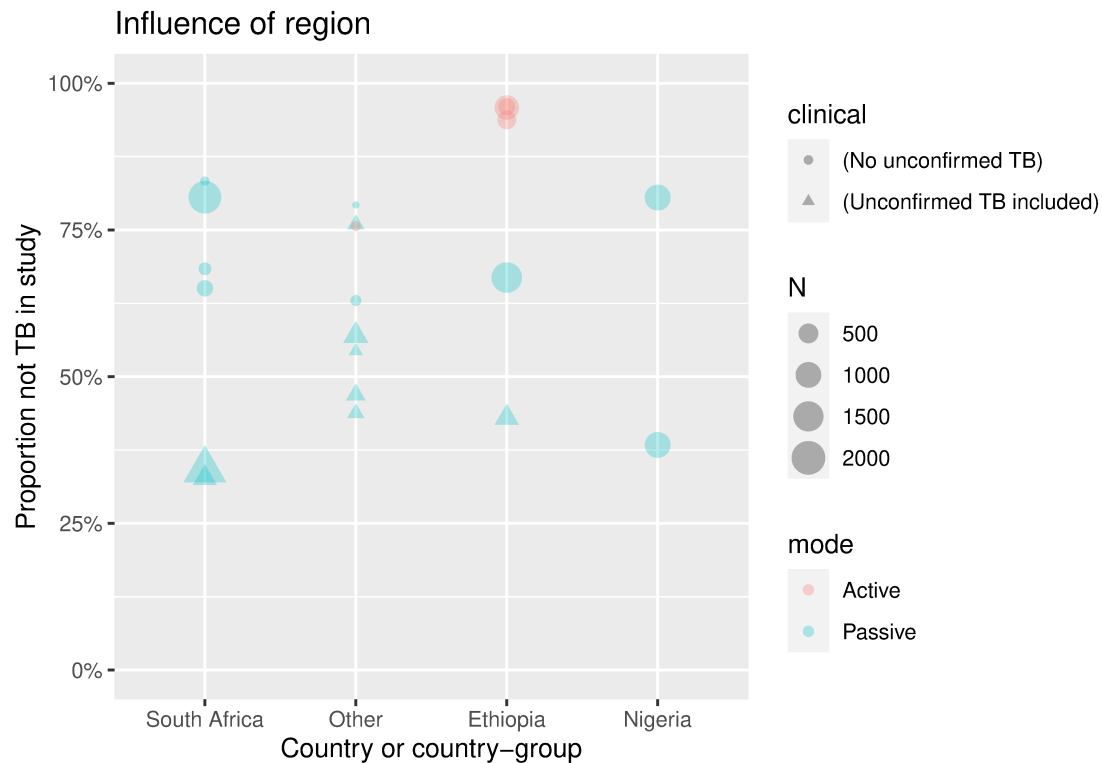This is very similar to the main analysis above.

Regional groupings

Here we investigate whether country can explain some heterogeneity. Since when countries have occur only once, it is not possible to identify a country coefficient, we these countries into an "Other" category.

```
DD[,Country.Group:=gsub(" \\-.+$","",Country)]                       #remove cities
DD[!Country.Group %in% c("South Africa","Ethiopia","Nigeria"),Country.Group:="Other"] #group
DD[,Country.Group:=factor(Country.Group,levels=unique(Country.Group))] #make factor
```

14

Plot this data:

```
ggplot(DD,aes(Country.Group,`NotTB Proportion`,
              size=N,col=mode,shape=clinical))+
   scale_y_continuous(label=percent,limits=c(0,1))+
   geom_point(alpha=a)+
   xlab('Country or country-group')+
   ylab('Proportion not TB in study')+
   ggtitle('Influence of region')
```


Perform meta-regression on country-group:

```
cgmr <-  rma.glmm(measure = "PLO",  #binomial w/ logit link
              xi = NnotTB,      # numerator
              ni = N,           # denominator
              data = DD,        # what data to use
              mods = ~mode*clinical + Country.Group)
```

```
## Warning: Redundant predictors dropped from the model.
```

```
summary(cgmr)
```

```
##
## Mixed-Effects Model (k = 21; tau^2 estimator: ML)
##
##   logLik  deviance       AIC       BIC      AICc
## -65.1801    1.0924  144.3602  151.6718  152.9755
##
## tau^2 (estimated amount of residual heterogeneity):     0.3559
```

15

```
## tau (square root of estimated tau^2 value):            0.5966
## I^2 (residual heterogeneity / unaccounted variability): 96.9246%
## H^2 (unaccounted variability / sampling variability):   32.5156
##
## Tests for Residual Heterogeneity:
## Wld(df = 15) = 776.0219, p-val < .0001
## LRT(df = 15) = 809.5261, p-val < .0001
##
## Test of Moderators (coefficients 2:6):
## QM(df = 5) = 49.6317, p-val < .0001
##
## Model Results:
##
##                                   estimate     se    zval    pval    ci.lb
## intrcpt                             2.1567  0.5023  4.2940  <.0001   1.1723
## modePassive                        -1.2854  0.4723 -2.7217  0.0065  -2.2110
## clinical(Unconfirmed TB included)  -1.1151  0.3371 -3.3082  0.0009  -1.7757
## Country.GroupOther                  0.2021  0.3565  0.5669  0.5708  -0.4966
## Country.GroupEthiopia               0.4592  0.4521  1.0158  0.3097  -0.4269
## Country.GroupNigeria               -0.4006  0.5052 -0.7931  0.4277  -1.3908
##                                     ci.ub
## intrcpt                            3.1412  ***
## modePassive                       -0.3597   **
## clinical(Unconfirmed TB included) -0.4544  ***
## Country.GroupOther                 0.9008
## Country.GroupEthiopia              1.3453
## Country.GroupNigeria               0.5895
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```