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 ................................. 1
  Dependencies ............................ 1

Main analyses ............................ 2
  Approach ............................... 2
  Meta-analyses .......................... 3
  Creation of combined forest plot ...... 6

Meta-regressions .......................... 8
  TB prevalence .......................... 8
  HIV prevalence ........................ 9
  Calendar time .......................... 11

Sensitivity analyses ..................... 13
  Dorman et al. by country only ....... 13
  Regional groupings .................... 14

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",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:

```r
pkgs.needed <- c("ggplot2","scales","cowplot","ggpubr",
                   "data.table","here",
                   "metafor")
# graphs
# data mgt
# metaanalysis
```
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$platform,  #platform
             sI$running,  #OS
             sI$otherPkg$metafor$Version  #metafor version
)
)
khitr::kable(dI)

<table>
<thead>
<tr>
<th>item</th>
<th>version</th>
</tr>
</thead>
<tbody>
<tr>
<td>R version</td>
<td>R version 4.1.0 (2021-05-18)</td>
</tr>
<tr>
<td>platform</td>
<td>x86_64-pc-linux-gnu (64-bit)</td>
</tr>
<tr>
<td>OS</td>
<td>Pop!_OS 21.04</td>
</tr>
<tr>
<td>metafor version</td>
<td>3.0-2</td>
</tr>
</tbody>
</table>

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:

```r
DD <- fread(file=here("SRMdata.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
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`:~N[NotTB/N]
for(i in 1:nrow/DD)){ DD[i,c('lo','hi'):=ciz(N[NotTB,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 = N[NotTB], # 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
forest(maPU, transf = transf.ilogit, refline = NA)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boehme et al.*.1</td>
<td>0.34</td>
<td>[0.32, 0.36]</td>
</tr>
<tr>
<td>Boehme et al.*.2</td>
<td>0.47</td>
<td>[0.41, 0.53]</td>
</tr>
<tr>
<td>Bruchfield et al.#</td>
<td>0.43</td>
<td>[0.39, 0.47]</td>
</tr>
<tr>
<td>Jayasooriya et al.#</td>
<td>0.44</td>
<td>[0.38, 0.50]</td>
</tr>
<tr>
<td>Munyati et al.</td>
<td>0.57</td>
<td>[0.53, 0.61]</td>
</tr>
<tr>
<td>Nliwasa et al.</td>
<td>0.76</td>
<td>[0.70, 0.81]</td>
</tr>
<tr>
<td>Reither et al.</td>
<td>0.54</td>
<td>[0.47, 0.62]</td>
</tr>
<tr>
<td>Theron et al.</td>
<td>0.33</td>
<td>[0.29, 0.37]</td>
</tr>
</tbody>
</table>

RE Model

<table>
<thead>
<tr>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
</tr>
<tr>
<td>0.4</td>
</tr>
<tr>
<td>0.6</td>
</tr>
<tr>
<td>0.8</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8757</td>
<td>0.2078</td>
<td>4.2139</td>
<td>&lt;0.0001</td>
<td>0.4684</td>
<td>1.2830 ***</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
forest(maPW, transf = transf.ilogit, reline=NA)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuevas et al.*.1</td>
<td>0.67</td>
<td>[0.65, 0.69]</td>
</tr>
<tr>
<td>Cuevas et al.*.2</td>
<td>0.81</td>
<td>[0.78, 0.83]</td>
</tr>
<tr>
<td>Dorman et al.*.1</td>
<td>0.83</td>
<td>[0.76, 0.88]</td>
</tr>
<tr>
<td>Dorman et al.*.2</td>
<td>0.68</td>
<td>[0.62, 0.74]</td>
</tr>
<tr>
<td>Dorman et al.*.3</td>
<td>0.79</td>
<td>[0.72, 0.85]</td>
</tr>
<tr>
<td>Dorman et al.*.4</td>
<td>0.63</td>
<td>[0.56, 0.70]</td>
</tr>
<tr>
<td>Hanrahan et al.</td>
<td>0.81</td>
<td>[0.79, 0.82]</td>
</tr>
<tr>
<td>Lawson et al.</td>
<td>0.38</td>
<td>[0.36, 0.41]</td>
</tr>
<tr>
<td>Ling et al.</td>
<td>0.65</td>
<td>[0.60, 0.70]</td>
</tr>
</tbody>
</table>

### RE Model

<table>
<thead>
<tr>
<th>Proportion</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>0.71</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

Meta-analysis for actively found TB patients:

```r
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)

<table>
<thead>
<tr>
<th>logLik</th>
<th>deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10.4692</td>
<td>0.2060</td>
<td>24.9385</td>
<td>23.7111</td>
<td>36.9385</td>
</tr>
</tbody>
</table>

### 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:

<table>
<thead>
<tr>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5537</td>
<td>0.4199</td>
<td>6.0817</td>
<td>&lt;.0001</td>
<td>1.7307</td>
<td>3.3767</td>
</tr>
</tbody>
</table>

### Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

forest(maA, transf = transf.ilogit, reline=NA)
Deribrew et al. 0.96 [0.94, 0.98]
Hamusse et al. 0.96 [0.94, 0.97]
Merid et al. 0.94 [0.91, 0.96]
Sekandi et al. 0.76 [0.68, 0.82]

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

Combined plot data:
B <- rbind(
    DD[,.(lab,'Not TB Proportion',lo,hi,SE,mode,clinical,
        qty,bac,CI,wt,w)],
    predz[,.(lab,'Not TB 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')]#or 3
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:
SA <- ggplot(B,aes(lab,y='Not TB 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')+ #or 3
    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))+
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.

```r
DD[,tb:='WHO TB estimate (per 100 000 year of study)']
a <- 0.3

plt <- 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')+
  ggttitle('Influence of population TB burden')
```

---

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

\[
\text{tbmr} \leftarrow \text{rma.glmm(measure = "PLO", \# binomial w/ logit link}
\xi = \text{NnotTB}, \quad \# \text{numerator}
ni = \text{N}, \quad \# \text{denominator}
data = \text{DD}, \quad \# \text{what data to use}
\text{mods} = \sim \text{mode*clinical + tb)}
\]

## Warning: Studies with NAs omitted from model fitting.
## Warning: Some yi/vi values are NA.
## Warning: Redundant predictors dropped from the model.

\text{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  ci.ub
## intrcpt  2.5877  0.3453  7.4931 <.0001  1.9109  3.2646***
## modePassive -1.6174  0.4233 -3.8210  0.0001 -2.4471 -0.7878***
## clinical(Unconfirmed TB included) -0.8999  0.3286 -2.7386  0.0062 -1.5439 -0.2559**
## tb -0.0002  0.0004 -0.4084  0.6830 -0.0009
##
## 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 influencing 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/le2, '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)) +
ggeom_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.glm(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:

<table>
<thead>
<tr>
<th></th>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>intrcpt</td>
<td>2.5888</td>
<td>0.3279</td>
<td>7.8949</td>
<td>&lt;.0001</td>
<td>1.9461</td>
<td>3.2315***</td>
</tr>
<tr>
<td>modePassive</td>
<td>-1.5920</td>
<td>0.4019</td>
<td>-3.9609</td>
<td>&lt;.0001</td>
<td>-2.3798</td>
<td></td>
</tr>
<tr>
<td>clinical(Unconfirmed TB included)</td>
<td>-0.8801</td>
<td>0.3153</td>
<td>-2.7914</td>
<td>0.0052</td>
<td>-1.4981</td>
<td></td>
</tr>
<tr>
<td>hiv</td>
<td>-0.0325</td>
<td>0.0420</td>
<td>-0.7730</td>
<td>0.4395</td>
<td>-0.1149</td>
<td></td>
</tr>
</tbody>
</table>

## 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

```r
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')
```

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

eyearmr <- rma.glmm(measure = "PL0", 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):
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
```

```r
tmp <- tmp[,.(NnotTB=sum(NnotTB),N=sum(N)),by=authorcountry]
kntir::kable(tmp) #check
```

<table>
<thead>
<tr>
<th>authorcountry</th>
<th>NnotTB</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuevas, Ethiopia</td>
<td>1184</td>
<td>1770</td>
</tr>
<tr>
<td>Cuevas, Nigeria</td>
<td>963</td>
<td>1196</td>
</tr>
<tr>
<td>Dorman, South Africa</td>
<td>285</td>
<td>384</td>
</tr>
<tr>
<td>Dorman, Kenya</td>
<td>107</td>
<td>135</td>
</tr>
<tr>
<td>Dorman, Uganda</td>
<td>114</td>
<td>181</td>
</tr>
<tr>
<td>Hanrahan, South Africa</td>
<td>1685</td>
<td>2091</td>
</tr>
<tr>
<td>Lawson, Nigeria</td>
<td>455</td>
<td>1186</td>
</tr>
<tr>
<td>Ling, South Africa</td>
<td>257</td>
<td>395</td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>logLik</th>
<th>deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>-26.5760</td>
<td>0.1654</td>
<td>57.1519</td>
<td>57.3108</td>
<td>59.5519</td>
</tr>
</tbody>
</table>

# 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:

<table>
<thead>
<tr>
<th>estimate</th>
<th>se</th>
<th>zval</th>
<th>pval</th>
<th>ci.lb</th>
<th>ci.ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8252</td>
<td>0.2149</td>
<td>3.8406</td>
<td>0.0001</td>
<td>0.4041</td>
<td>1.2463</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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

<table>
<thead>
<tr>
<th>Country, South Africa</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuevas, Ethiopia</td>
<td>0.67 [0.65, 0.69]</td>
</tr>
<tr>
<td>Cuevas, Nigeria</td>
<td>0.81 [0.78, 0.83]</td>
</tr>
<tr>
<td>Dorman, South Africa</td>
<td>0.74 [0.70, 0.78]</td>
</tr>
<tr>
<td>Dorman, Kenya</td>
<td>0.79 [0.72, 0.85]</td>
</tr>
<tr>
<td>Dorman, Uganda</td>
<td>0.63 [0.56, 0.70]</td>
</tr>
<tr>
<td>Hanrahan, South Africa</td>
<td>0.81 [0.79, 0.82]</td>
</tr>
<tr>
<td>Lawson, Nigeria</td>
<td>0.38 [0.36, 0.41]</td>
</tr>
<tr>
<td>Ling, South Africa</td>
<td>0.65 [0.60, 0.70]</td>
</tr>
</tbody>
</table>

### RE Model

0.70 [0.60, 0.78]

Proportion

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
```
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=.5)+
  xlab('Country or country-group')+ 
  ylab('Proportion not TB in study')+ 
  ggttitle('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
```
## 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 | ci.ub |
## |----------|-----|-------|------|-------|-------|
## | intercept | 2.1567 | 0.5023 | 4.2940 | <.0001 | 1.1723 |
## | mode Passive | -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.Group Other | 0.2021 | 0.3565 | 0.5669 | 0.5708 | -0.4966 |
## | Country.Group Ethiopia | 0.4592 | 0.4521 | 1.0158 | 0.3097 | -0.4269 |
## | Country.Group Nigeria | -0.4006 | 0.5052 | -0.7931 | 0.4277 | -1.3908 |
## | ci.lb | ci.ub |
## | ----- | ----- |
## | intercept | 3.1412 *** |
## | mode Passive | -0.3597 ** |
## | clinical(Unconfirmed TB included) | -0.4544 *** |
## | Country.Group Other | 0.9008 |
## | Country.Group Ethiopia | 1.3453 |
## | Country.Group Nigeria | 0.5895 |
##
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1