

results.R

joebrew

Wed May 17 14:00:11 2017

```
source('make_wide.R')
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
##
## Parsed with column specification:
## cols(
##   iso3 = col_character(),
##   `COUNTRY NAME` = col_character(),
##   country_number = col_integer()
## )
##
## Attaching package: 'tidyr'
##
## The following objects are masked _by_ 'GlobalEnv':
##
##   population, who
##
## Warning in read_fun(path = path, sheet = sheet, limits = limits, shim =
## shim, : Coercing text to numeric in BV2291 / R2291C74: '4.5'
##
## Warning in read_fun(path = path, sheet = sheet, limits = limits, shim =
## shim, : Coercing text to numeric in BV2292 / R2292C74: '8.2'
##
## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   country = col_character(),
##   iso2 = col_character(),
##   iso3 = col_character(),
##   g_whoregion = col_character(),
##   new_sn_sexunk04 = col_character(),
##   new_sn_sexunk514 = col_character(),
##   new_sn_sexunk014 = col_character(),
##   new_sn_sexunk15plus = col_character(),
##   new_ep_m04 = col_character(),
##   new_ep_sexunkageunk = col_character(),
##   newrel_sexunk04 = col_character(),
##   newrel_sexunk514 = col_character(),
##   newrel_sexunk014 = col_character(),
##   newrel_sexunk15plus = col_character(),
##   newrel_sexunkageunk = col_character(),
```

```

##   rdxsurvey_newinc = col_character(),
##   rdxsurvey_newinc_rdx = col_character()
## )

## See spec(...) for full column specifications.

## Parsed with column specification:
## cols(
##   .default = col_integer(),
##   country = col_character(),
##   iso2 = col_character(),
##   iso3 = col_character(),
##   iso_numeric = col_character(),
##   g_whoregion = col_character()
## )

## See spec(...) for full column specifications.

## Parsed with column specification:
## cols(
##   .default = col_character(),
##   year = col_integer(),
##   e_rr_pct_new = col_double(),
##   e_rr_pct_new_lo = col_double(),
##   e_rr_pct_new_hi = col_double(),
##   e_mdr_pct_rr_new = col_integer(),
##   e_rr_pct_ret = col_double(),
##   e_rr_pct_ret_lo = col_double(),
##   e_rr_pct_ret_hi = col_double(),
##   e_mdr_pct_rr_ret = col_double(),
##   e_inc_rr_num = col_integer(),
##   e_inc_rr_num_lo = col_integer(),
##   e_inc_rr_num_hi = col_integer(),
##   e_rr_in_notified_pulm = col_integer(),
##   e_rr_in_notified_pulm_lo = col_integer(),
##   e_rr_in_notified_pulm_hi = col_integer()
## )

## See spec(...) for full column specifications.

## Warning in log(stand_dif): NaNs produced

## Warning in log(ab): NaNs produced

# A total of 195 countries had TB mortality estimates for 2015 by both WHO and IHME.
table(df$have_both)

##
## TRUE
## 195

# Among those countries, WHO estimated 1,768,482 total number of deaths attributable to tuberculosis, w
sum(df$i_both_all_tbttotal_nd)

## [1] 1322916

sum(df$w_both_all_tbttotal_nd)

## [1] 1768482

```

```

sum(df$i_both_all_tbttotal_nd) -
sum(df$w_both_all_tbttotal_nd)

## [1] -445566.6

(sum(df$i_both_all_tbttotal_nd) -
 sum(df$w_both_all_tbttotal_nd)) /
sum(df$i_both_all_tbttotal_nd)

## [1] -0.3368064

(sum(df$i_both_all_tbttotal_nd) -
 sum(df$w_both_all_tbttotal_nd)) /
sum(df$w_both_all_tbttotal_nd)

## [1] -0.2519485
#This difference in TB mortality was higher in people living with HIV (211,604 by IHME vs 389,042 by WHO)
sum(df$i_both_all_htb_nd)

## [1] 211603.9
sum(df$w_both_all_htb_nd)

## [1] 389042
# The relative difference in number of deaths was especially higher for the paediatric population, where
sum(df$i_both_014_tb_nd + df$i_both_014_htb_nd)

## [1] 69658.59
sum(df$w_both_014_tb_nd + df$w_both_014_htb_nd)

## [1] 209837.1
# Among adult TB deaths, there were not large differences in the sex-specific mortality estimates (table 1)
# There were 84 countries (43.1%), in which WHO estimated higher number of deaths attributable to TB than IHME
table(df$w_both_all_tbttotal_nd > df$i_both_all_tbttotal_nd)

##
## FALSE TRUE
## 109 86
prop.table(table(df$w_both_all_tbttotal_nd > df$i_both_all_tbttotal_nd))

##
## FALSE TRUE
## 0.5589744 0.4410256
# Those countries with larger absolute differences in total number of TB deaths were (by decreasing magnitude)
df %>%
  mutate(x = w_both_all_tbttotal_nd - i_both_all_tbttotal_nd) %>%
  arrange(desc(x)) %>%
  mutate(x = round(x)) %>%
  dplyr::select(country, x) %>%
  head(10)

##          country      x
## 1      Nigeria 216621

```

```
## 2      Bangladesh 49863
## 3      Tanzania  38272
## 4      South Africa 29108
## 5      Mozambique 28909
## 6      Indonesia 26121
## 7      Dem. Rep. Congo 26010
## 8      India 20696
## 9      North Korea 13218
## 10     Angola 9910
```

```
# The countries in which IHME estimated higher number of deaths than WHO were: Ethiopia (22650), China
df %>%
```

```
  mutate(x = i_both_all_tbttotal_nd - w_both_all_tbttotal_nd) %>%
  arrange(desc(x)) %>%
  mutate(x = round(x)) %>%
  dplyr::select(country, x) %>%
  head(10)
```

```
##      country      x
## 1     Ethiopia 22650
## 2       China 13538
## 3    Zimbabwe 11082
## 4  Philippines  9436
## 5       Nepal  5477
## 6       Uganda  5081
## 7  Burkina Faso  4837
## 8        Niger  3758
## 9     Viet Nam  3252
## 10    Senegal  3147
```

```
# Map 1 shows how the largest differences in terms of absolute number of deaths are concentrated in few
cor(df$i_both_all_tbttotal_nd, df$w_both_all_tbttotal_nd)
```

```
## [1] 0.9209976
```

```
t.test(df$i_both_all_tbttotal_nd,
       df$w_both_all_tbttotal_nd,
       paired=TRUE)
```

```
##
## Paired t-test
##
## data: df$i_both_all_tbttotal_nd and df$w_both_all_tbttotal_nd
## t = -1.89, df = 194, p-value = 0.06025
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -4669.36003 99.44636
## sample estimates:
## mean of the differences
## -2284.957
```

```
#For the African region, for those countries with lower mortality burden, IHME estimates higher mortality
x <- df %>% filter(who_region == 'AFR')
plot(x$w_both_all_tbttotal_nd,
     x$i_both_all_tbttotal_nd)
```

```
# After standardizing the absolute difference by the total number of reported deaths, thus taking into
```

```
df %>%
  mutate(x = adjusted_stand_dif) %>%
  arrange(desc(x)) %>%
  mutate(x = round(x, 1)) %>%
  dplyr::select(country, x) %>%
  head(10)
```

```
##           country      x
## 1      Nigeria 100.0
## 2 Marshall Islands 92.1
## 3      Timor-Leste 91.9
## 4 Papua New Guinea 90.6
## 5         Libya 90.1
## 6      North Korea 88.7
## 7      Greenland 78.8
## 8         Iceland 77.1
## 9         Sudan 70.2
## 10        Laos 68.9
```

#Likewise, the most important differences in which IHME estimated higher mortality are Eritrea (-13.4),

```
df %>%
  mutate(x = adjusted_stand_dif) %>%
  arrange(x) %>%
  mutate(x = round(x, 1)) %>%
  dplyr::select(country, x) %>%
  head(10)
```

```
##           country      x
## 1      Bermuda -119.4
## 2 Virgin Islands U.S. -119.4
## 3      Seychelles -119.0
## 4      Azerbaijan -101.1
## 5 Macedonia the former Yugoslav Republic of -85.3
## 6         Rwanda -80.1
## 7      Saint Lucia -73.4
## 8         Egypt -69.7
## 9      Burkina Faso -66.8
## 10      Barbados -66.3
```

Twenty-three countries did not have an adjusted standardized difference since reported number of deaths was zero

```
table(is.na(df$adjusted_stand_dif))
```

```
##
## FALSE
##    195
```

Standardization of the absolute difference in TB number of deaths by incident number of TB cases (as a proportion of the total number of TB cases)

```
x <- df %>% filter(who_region == 'SEA')
table(x$w_both_all_tbttotal_nd > x$i_both_all_tbttotal_nd)
```

```
##
## FALSE  TRUE
##      2    9
```

```

# As for the absolute differences we can see that the most dramatic differences in TB mortality estimat

# When looking at the association between the adjusted standardized difference between IHME and WHO mor

#MDR/RR prevalence (r= -0.05, 95%CI)

#and case fatality rate (r= -0.19, 95%CI).

#There is a moderate correlation with case detection rate (as estimated by WHO), (r= -0.32, 95%CI), whi

# 1. Can you tell me the number of countries with WHO estimates without IHME estimates? (and the total

#
# 2. The paragraph of the associations with different factors, you did not include the 95% CI, which
#
# 3. Lastly, i wanted to mention something about children. Could you tell me top 10 countries with high

x <- df %>%
  dplyr::select(country,
                w_both_014_htb_nd,
                w_both_014_tb_nd,
                i_both_014_htb_nd,
                i_both_014_tb_nd,
                newrel_f014,
                newrel_m014) %>%
  mutate(difference = (w_both_014_htb_nd + w_both_014_tb_nd) - (i_both_014_htb_nd + i_both_014_tb_nd))
  mutate(absolute_difference = abs(difference)) %>%
  arrange(desc(absolute_difference)) %>%
  # In order to further appraise the children issue, i had an idea. Could we standardize (divide) the d
  mutate(y = absolute_difference / (newrel_f014 + newrel_m014))

x %>%
  mutate(x = difference) %>%
  arrange(x) %>%
  mutate(x = round(x, 1)) %>%
  dplyr::select(country, x) %>%
  head(10)

```

```

##      country      x
## 1    Zimbabwe -2191.6
## 2 South Africa -1525.0
## 3     Malawi  -1022.4
## 4     Uganda   -504.5
## 5 Burkina Faso  -499.9
## 6     Lesotho  -415.2
## 7      Rwanda  -378.9
## 8     Burundi  -279.3
## 9   Swaziland  -223.3
## 10    Namibia  -170.1

```

```
x %>%
  mutate(x = difference) %>%
  arrange(desc(x)) %>%
  mutate(x = round(x, 1)) %>%
  dplyr::select(country, x) %>%
  head(10)
```

```
##           country      x
## 1           India 49508.8
## 2          Nigeria 32004.0
## 3        Indonesia 12752.4
## 4      Bangladesh  8616.8
## 5          Tanzania  5992.6
## 6 Dem. Rep. Congo  5147.4
## 7             China  3214.4
## 8           Myanmar  3169.5
## 9           Pakistan 2334.5
## 10          Angola  2089.4
```

```
x %>%
  mutate(x = y) %>%
  filter(!is.na(x),
         !is.infinite(y)) %>%
  arrange(x) %>%
  # mutate(x = round(x, 1)) %>%
  dplyr::select(country, x) %>%
  head(10)
```

```
##           country      x
## 1           Andorra 0.0009907814
## 2          Greenland 0.0014613359
## 3             Guam 0.0038623173
## 4        Tajikistan 0.0042242240
## 5             Sweden 0.0042686713
## 6          Luxembourg 0.0043598820
## 7          Switzerland 0.0065882240
## 8          Philippines 0.0067090392
## 9 Northern Mariana Islands 0.0068886413
## 10          Bahamas 0.0071177561
```

```
x %>%
  mutate(x = y) %>%
  filter(!is.na(x),
         !is.infinite(y)) %>%
  arrange(desc(x)) %>%
  # mutate(x = round(x, 1)) %>%
  dplyr::select(country, x) %>%
  head(10)
```

```
##           country      x
## 1          Viet Nam 14.094966
## 2             Laos  8.361368
## 3        Burkina Faso  8.332194
## 4             Nigeria  6.705207
## 5             Algeria  5.501055
```

```

## 6 Papua New Guinea 3.219987
## 7 Ghana 2.197464
## 8 Guinea-Bissau 1.970405
## 9 Libya 1.718705
## 10 Lesotho 1.701528

write_csv(x, '~/Desktop/children_updated.csv')

# How many deaths in those 23 countries with who and not ihme
sum(df$w_both_all_tbttotal_nd[df$have_who & !df$have_ihme], na.rm = T)

## [1] 0

cor(df$adjusted_stand_dif, df$newrel_hivpos/df$newrel_hivtest,
    use = 'complete.obs')

## [1] -0.001091458

# b) CDR by IHME
# adjusted_stand_diff and cdr_ihme
cor(df$adjusted_stand_dif, df$cdr_ihme,
    use = 'complete.obs')

## [1] 0.4283954

cor(df$stand_dif, df$cdr_ihme,
    use = 'complete.obs')

## [1] 0.4283954

# c) CDR by WHO
# adjusted_stand_diff and gb_c_cdr
cor(df$adjusted_stand_dif, as.numeric(df$gb_c_cdr),
    use = 'complete.obs')

## [1] -0.3688292

cor(df$stand_dif, as.numeric(df$gb_c_cdr),
    use = 'complete.obs')

## [1] -0.3688292

# d) Reported case fatality rate
# adjusted_stand_diff and case_fatality_rate_2015_adjusted
cor(df$adjusted_stand_dif, df$case_fatality_rate_2015_adjusted,
    use = 'complete.obs')

## [1] -0.1825428

cor(df$stand_dif, df$case_fatality_rate_2015_adjusted,
    use = 'complete.obs')

## [1] -0.1825428

# e) Reported MDR
# adjusted_stand_diff and reported_mdr
cor(df$adjusted_stand_dif, df$reported_mdr,
    use = 'complete.obs')

```



```
## [1] -0.0131539
```

```
cor(df$stand_dif,df$reported_mdr,  
    use = 'complete.obs')
```

```
## [1] -0.0131539
```

```
library(psychometric)
```

```
## Loading required package: multilevel
```

```
## Loading required package: nlme
```

```
##
```

```
## Attaching package: 'nlme'
```

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

```
##
```

```
## collapse
```

```
## Loading required package: MASS
```

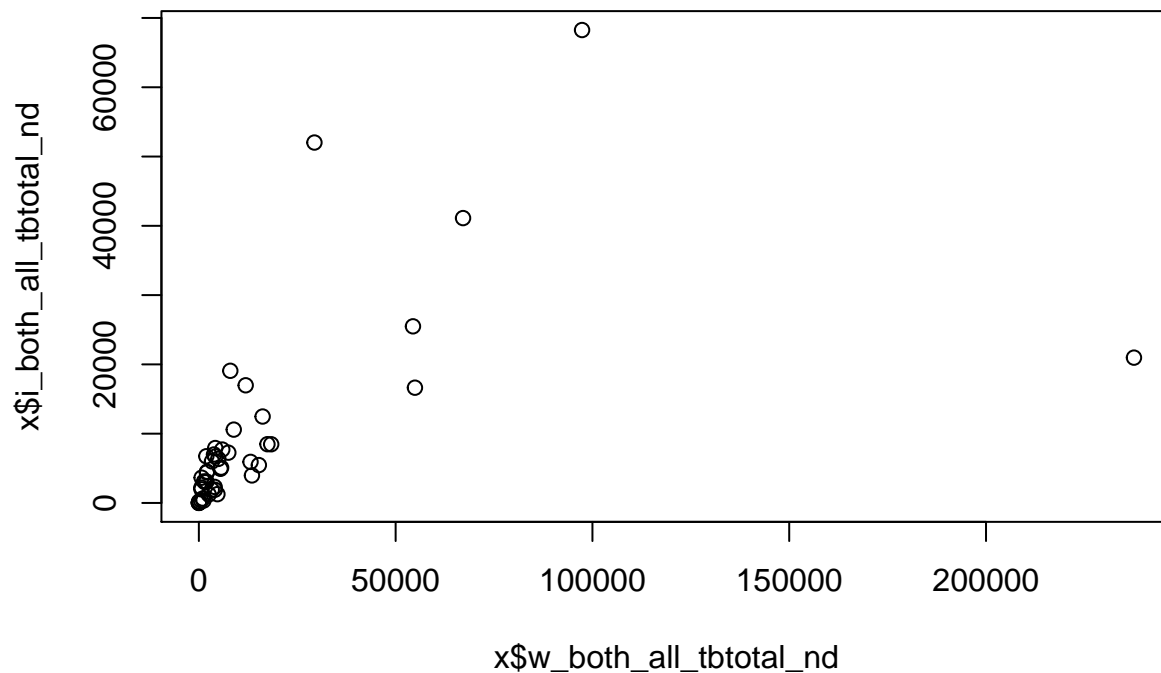
```
##
```

```
## Attaching package: 'MASS'
```

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

```
##
```

```
## select
```



```
# Correlation coefficients
```

```
confy <- function(x){  
  # fit <- lm(df$adjusted_stand_dif ~ x)  
  r <- cor(df$adjusted_stand_dif, x,  
          use = 'complete.obs')  
  CIr(r=r, n = nrow(df), level = .95)  
}
```

```

confy(df$newrel_hivpos/df$newrel_hivtest)

## [1] -0.1415821  0.1394423
confy(df$cdr_ihme)

## [1] 0.3063221 0.5366068
confy(df$gb_c_cdr)

## [1] -0.4842455 -0.2407962
confy(df$case_fatality_rate_2015_adjusted)

## [1] -0.31497616 -0.04313687
# a) was the WHO estimated CDR significantly lower for countries that had prevalence surveys than for c
fit <- lm(gb_c_cdr ~ prevsurvey, data = df)
confint(fit)

##                2.5 %    97.5 %
## (Intercept)  74.43136  79.02103
## prevsurvey  -29.50452 -15.10575
# b) Could be to do a box and whisker plot of CDR estimates for WHO and IHME (e.g. different colours)
x <- df %>%
  dplyr::select(country, prevsurvey,
                gb_c_cdr,
                cdr_ihme) %>%
  rename(IHME = cdr_ihme,
         WHO = gb_c_cdr,
         `Prevalence survey` = prevsurvey) %>%
  mutate(`Prevalence survey` = ifelse(`Prevalence survey` == 1, 'Survey',
                                     'No survey')) %>%

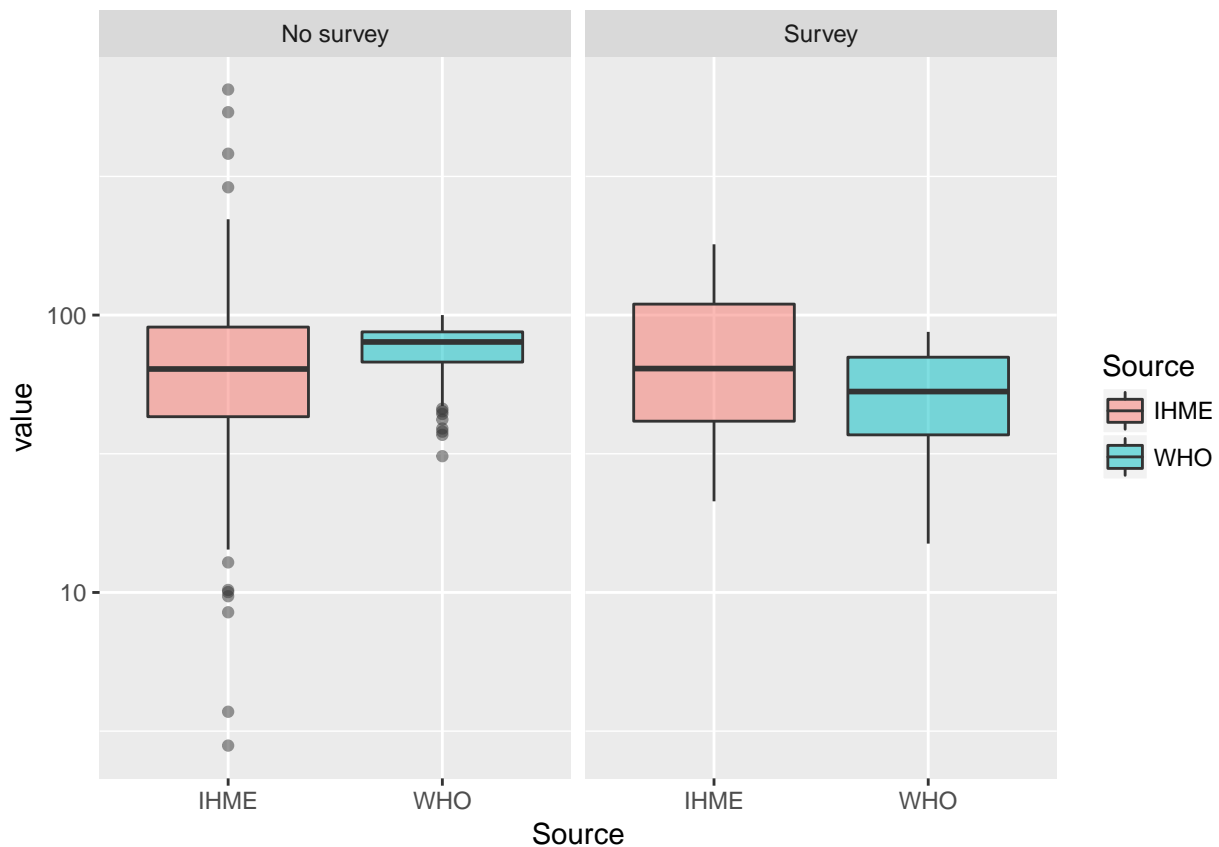
  gather(Source,
         value,
         WHO:IHME)
library(ggplot2)

##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:psychometric':
##
##      alpha
ggplot(data = x,
       aes(x = Source,
           y = value,
           fill = Source,
           group = Source)) +
  geom_boxplot(alpha = 0.5) +
  facet_wrap(~`Prevalence survey`) +
  scale_y_log10()

## Warning: Transformation introduced infinite values in continuous y-axis
## Warning: Removed 21 rows containing non-finite values (stat_boxplot).

```

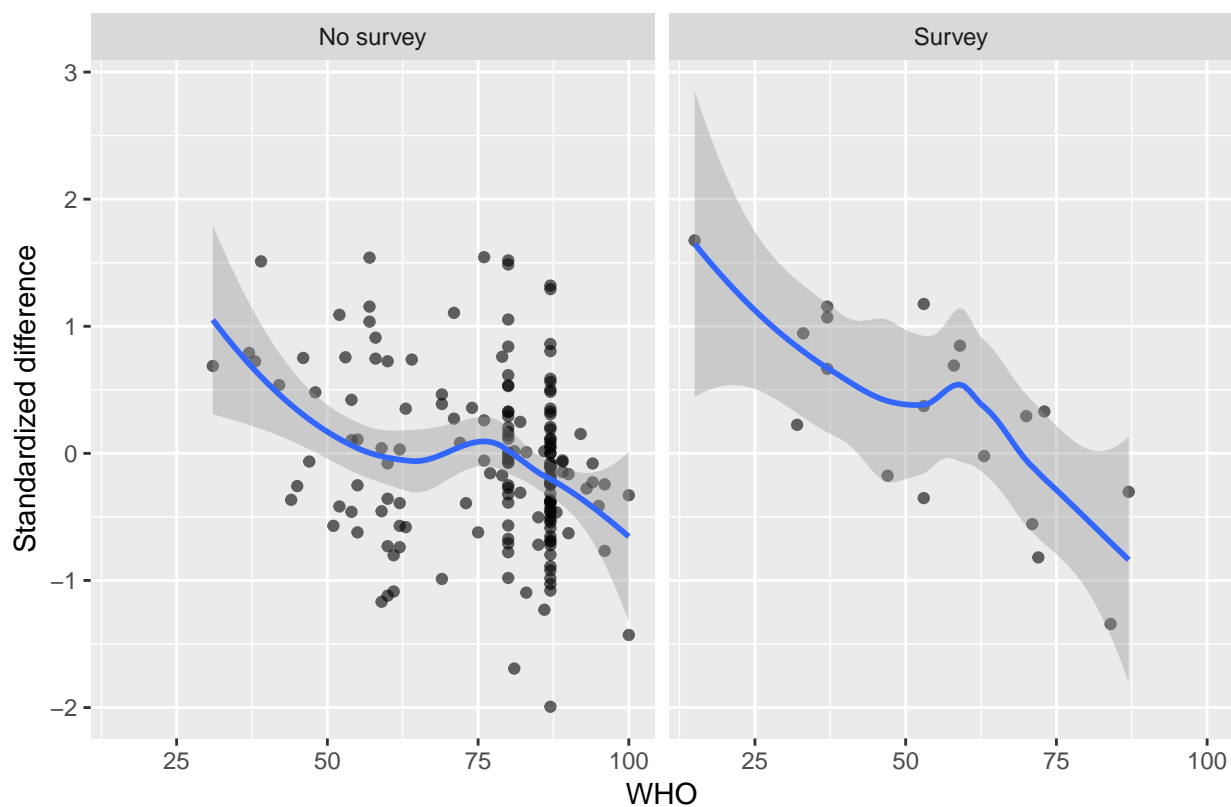


```
# c) to what extent the association of CDR are stand diff changes if we remove those countries with prevalence survey
x <- df %>%
  dplyr::select(country, prevsurvey,
                gb_c_cdr,
                cdr_ihme,
                stand_dif) %>%
  rename(IHME = cdr_ihme,
         WHO = gb_c_cdr,
         `Prevalence survey` = prevsurvey,
         `Standardized difference` = stand_dif) %>%
  mutate(`Prevalence survey` = ifelse(`Prevalence survey` == 1, 'Survey',
                                       'No survey'))

ggplot(data = x,
       aes(x = WHO,
          y = `Standardized difference`)) +
  geom_point(alpha = 0.6) +
  facet_wrap(~`Prevalence survey`) +
  geom_smooth() +
  labs(title = 'WHO CDR and standardized difference')

## `geom_smooth()` using method = 'loess'
## Warning: Removed 8 rows containing non-finite values (stat_smooth).
## Warning: Removed 8 rows containing missing values (geom_point).
```

WHO CDR and standardized difference



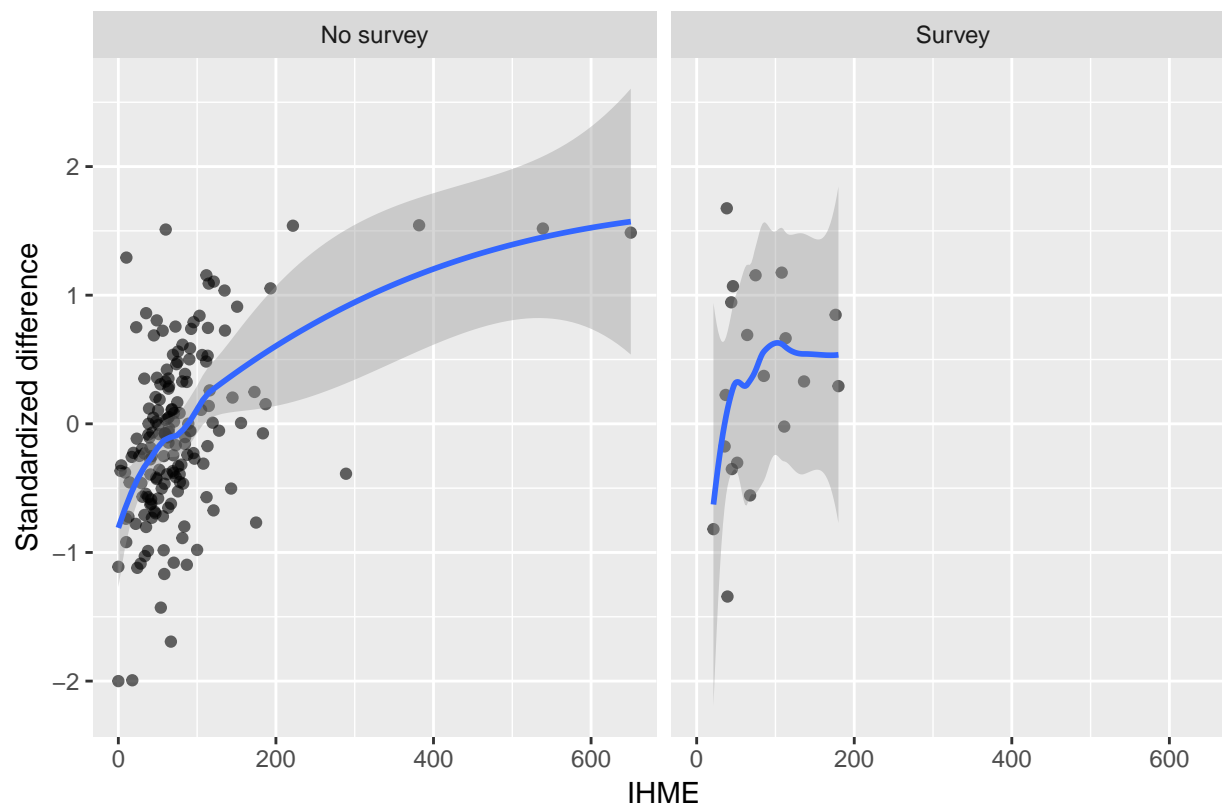
```
ggplot(data = x,
       aes(x = IHME,
           y = `Standardized difference`)) +
  geom_point(alpha = 0.6) +
  facet_wrap(~`Prevalence survey`) +
  geom_smooth() +
  labs(title = 'IHME CDR and standardized difference')
```

```
## `geom_smooth()` using method = 'loess'
```

```
## Warning: Removed 11 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 11 rows containing missing values (geom_point).
```

IHME CDR and standardized difference



d) In the paper of TB by IHME, they report 1,324,342 deaths (195 countries). Our results for IHME are