
Subject: STATCompiler for pooled prevalences
Posted by [zah123](#) on Wed, 29 May 2024 12:34:26 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hello,

I am trying to analyze the trend of the prevalence of diarrhea in each country over time using the dataset provided by STATcompiler. My goal is to pool the prevalences of the DHS surveys of countries every 5 years (for example, pooling the prevalence of diarrhea from 1985-1989, 1990-1994, etc.) and then plot it over time.

I used a meta-analysis approach with a random effect to pool the weighted prevalences per 5-year period for the countries. However, I have doubts regarding the pooled weighted prevalences: some countries need to be more represented since they have a larger population. Is it acceptable to leave the analysis as it is, or should I adjust it?

I've heard about weight denormalization, but this cannot be applied in my case since I am only working with STATcompiler data and do not have access to each survey. I have attached below the data I'm working with, extracted from STATcompiler. Below is the code I used for the analysis (in R) :

```
meta_analysis <- metaprop(event = data$Events, n = data$Total,  
                          sm = "PFT", title="Meta analysis per Year category", subgroup =  
data$`Year_Category`)
```

Any help would be greatly appreciated !

File Attachments

1) [data STATcompiler.PNG](#), downloaded 57 times

Subject: Re: STATCompiler for pooled prevalences
Posted by [Bridgette-DHS](#) on Thu, 30 May 2024 14:28:29 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS staff member, Tom Pullum:

Whenever you pool survey results, you have to choose from 3 possible weighting schemes. #1, you just take the unweighted arithmetic mean of the separate prevalences. #2, you use weights which are proportional to the sample sizes for the estimates. I believe STATcompiler includes the n's for most estimates. #3, you use weights that are proportional to the estimated size of the relevant subpopulation (e.g. children age 0-4) at the time of the survey. You can get these from the UN Population Division's WPP (World Population Prospects) 2022 spreadsheets that give annual age distributions for every country.

#1 is simplest but seems intuitively to be a bad idea. For #2, the problem is that the sample sizes

are largely arbitrary. It only makes sense for some statistical purposes, such as hypothesis testing. #3 is probably best, although it has the problem that large countries will swamp small countries.

It depends on what you are trying to estimate. If, for example, you are trying to estimate the probability that a child from any randomly selected household in the geographic region of South Asia had diarrhea in the past two weeks, then #3 is definitely the best option, and it will be ok that most South Asian children live in India.

One issue you will face in any region is that DHS does not provide data from every country in the region. Moreover, the coverage is not consistent over time. I don't know how you fill in these holes in the data; all you can do is to list the countries that contribute to each pooled estimate. For the South Asia example, there are 5-year intervals of time during which there was no survey in India. Including India for the intervals when it had a survey, and omitting it when there was no survey, will produce an uninterpretable trend line.

I personally have always avoided this kind of pooling. Apart from the technical issues of weights and missing observations, it masks important differences between the countries in the same region.

Subject: Re: STATCompiler for pooled prevalences
Posted by [zah123](#) on Sun, 02 Jun 2024 06:27:55 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hello,

Thank you very much for your response.

I have decided to take option number 2, pooling the data together to create a prevalence of diarrhea for the available DHS countries at that time just to show that the trend is decreasing over time. The samples were equally represented (with the same weights) in each 5-year category, and I used a random effect to account for the unexplained variability and heterogeneity between countries within a specific 5-year period.