# Bayesian model for COVID IFR: baggr implementation

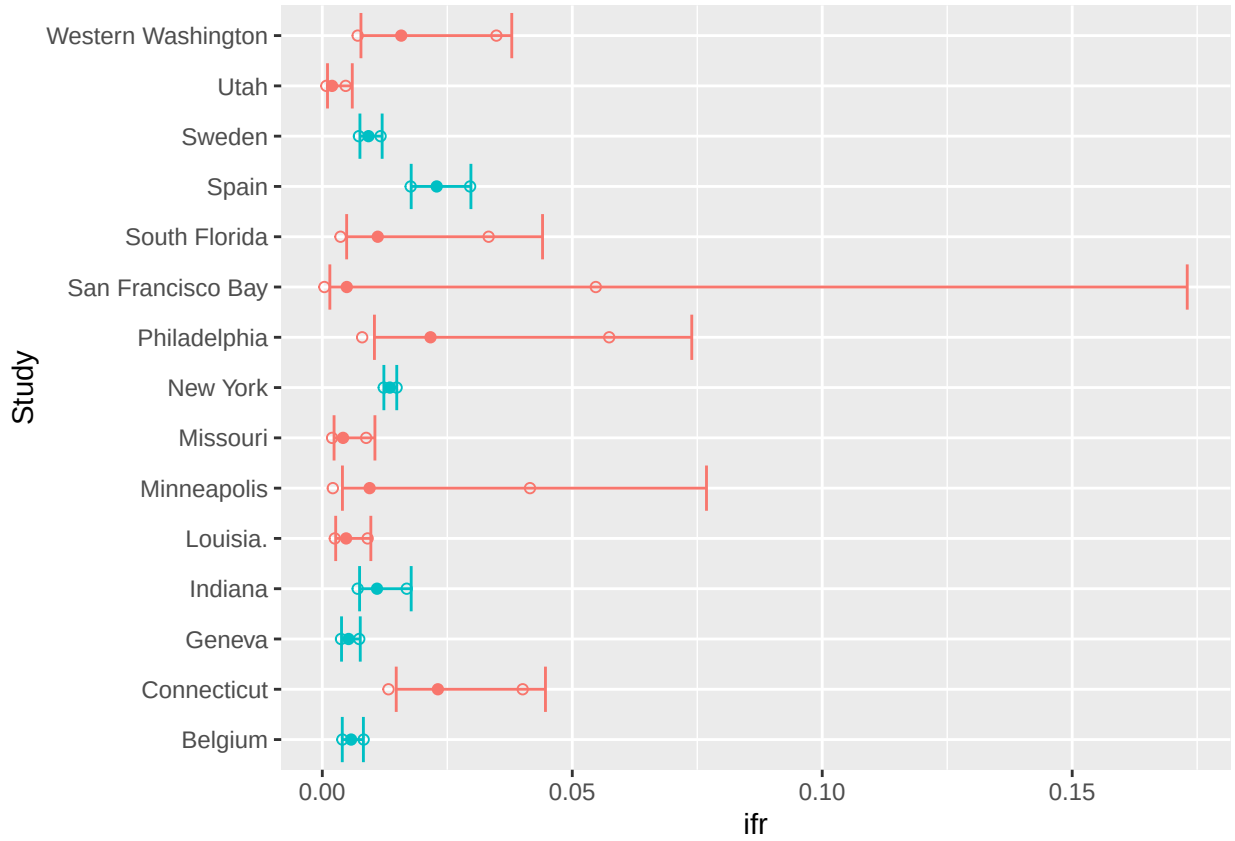
Witold Wiecek for 1 Day Sooner

Last updated 2020-09-03

**This is a simpler implementation of the model that is stored here for posterity, but for the main model please refer to `bayesian_ifr_model.Rmd`.**

## Model data

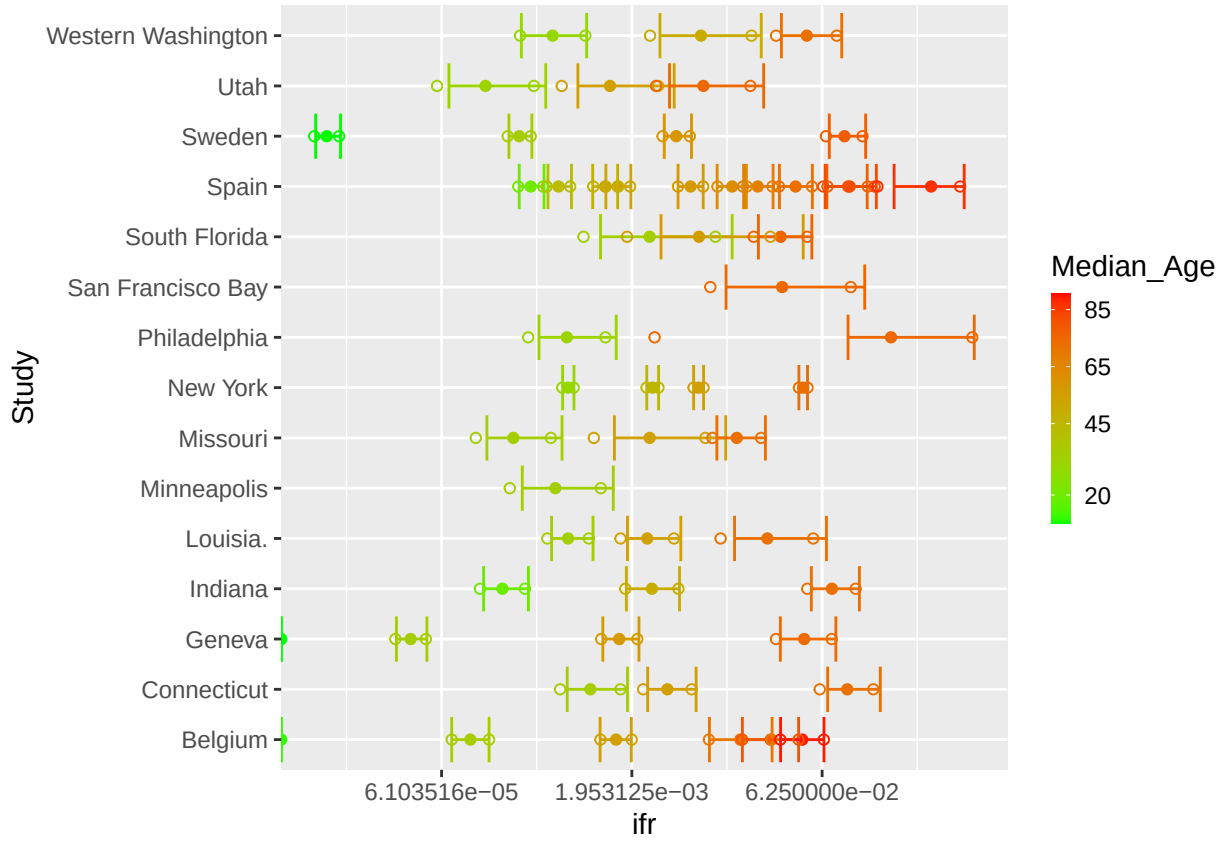
We use estimates originally collected by Levin, Cochran, and Walsh (2020) to construct the first version of analysis dataset. The input data into our model consists of deaths (treated as known) and prevalences (treated as logit-distributed parameter with known mean and SD) in all reported age groups in all studies[1].




```
## Warning: Transformation introduced infinite values in continuous x-axis
```

[1]This basic approach exaggerates uncertainty, as we treat different 95% intervals reported in the study as uncorrelated.

```
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Transformation introduced infinite values in continuous x-axis
## Warning: Removed 2 rows containing missing values (geom_point).
## Warning: Removed 2 rows containing missing values (geom_point).
```



## Results for analysis of overall IFRs

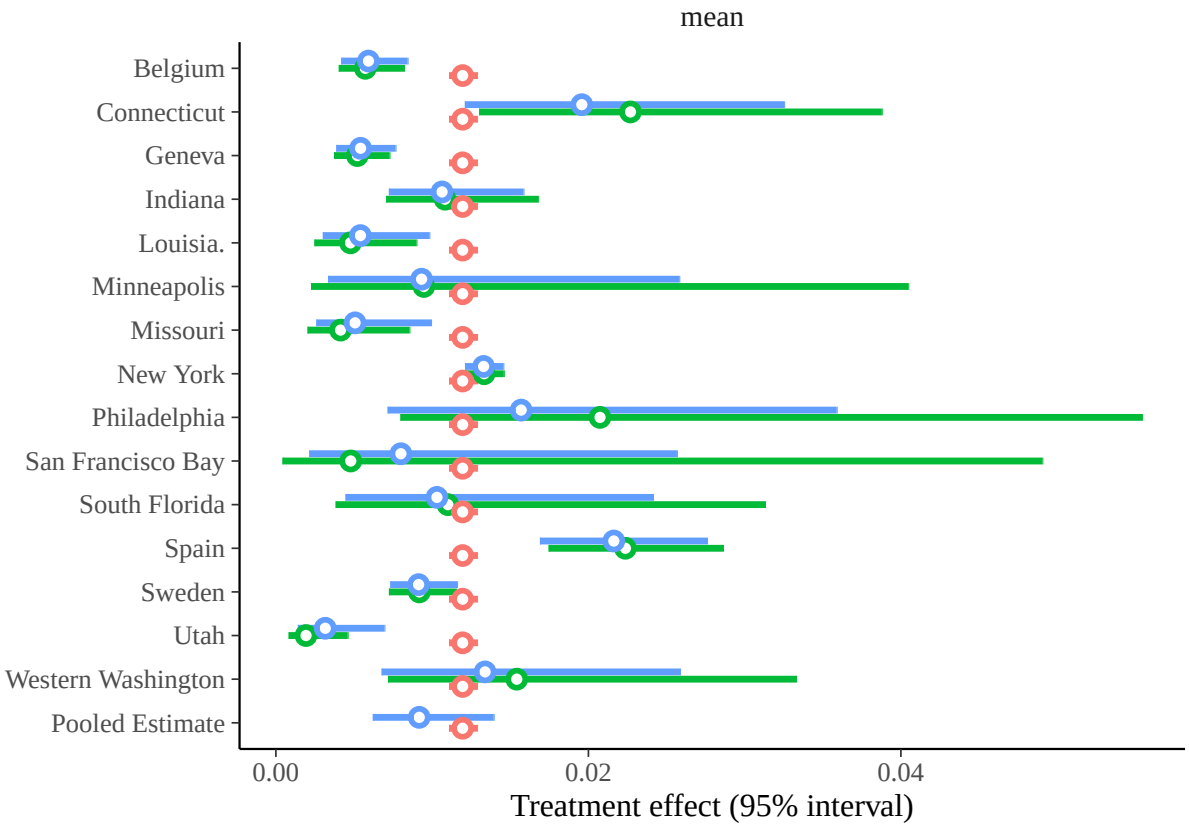
We label the collected estimates as $IFR_i$. The corresponding standard errors (after the logit transform) are $se_i$. The model is

$$\text{logit}(\hat{IFR}_i) \sim \mathcal{N}(\theta_i, se_i) \tag{1}$$

$$\theta_i \sim \mathcal{N}(\tau, \sigma) \tag{2}$$

where $\theta_i$ is the real value of underlying logit of IFR in study $i$. We assume $\tau \sim \mathcal{N}(0, 100)$, $\sigma \sim \mathcal{N}(0, 100)$ and $se_i$'s are treated as known parameters, derived from the assumption of logit-normality of IFR's.

We fit two models, one with partial pooling (assumptions as above) and one with full pooling (fixing $\sigma = 0$). We also show no pooling estimates for comparison.




We can conduct a formal comparison of full vs partial pooling to confirm that there is a considerable heterogeneity and that partial pooling is preferred, but this should be obvious from the plots.

```
## Comparison of cross-validation
##
##                     ELPD ELPD SE
## Model 1 - Model 2 55.3      21
```

A summary of the partially pooled model (we use exp transform rather than inv logit for technical reasons, **will be fixed**)

```
## Model type: Rubin model with aggregate data
## Pooling of effects: partial
##
## Aggregate treatment effect (on mean):
## Exponent of hypermean (exp(tau)) =  0.0094 with 95% interval 0.0058 to 0.0138
##
## Treatment effects on mean (converted to exp scale):
##                       mean    lci    uci pooling
## Connecticut         0.0200 0.0119 0.0342  0.1568
## Louisia.            0.0055 0.0030 0.0096  0.1999
## Minneapolis         0.0093 0.0034 0.0255  0.5445
## Missouri            0.0051 0.0026 0.0100  0.2500
## Philadelphia        0.0159 0.0072 0.0363  0.3565
## San Francisco Bay   0.0078 0.0021 0.0257  0.7456
## South Florida       0.0103 0.0043 0.0248  0.4086
## Utah                0.0032 0.0014 0.0068  0.3094
## Western Washington 0.0136 0.0068 0.0276  0.2719
## Belgium             0.0060 0.0042 0.0085  0.0750
## Geneva              0.0055 0.0039 0.0076  0.0672
## Indiana             0.0107 0.0071 0.0162  0.1044
## New York            0.0135 0.0123 0.0148  0.0057
## Spain               0.0221 0.0172 0.0283  0.0399
## Sweden              0.0092 0.0073 0.0117  0.0328
```

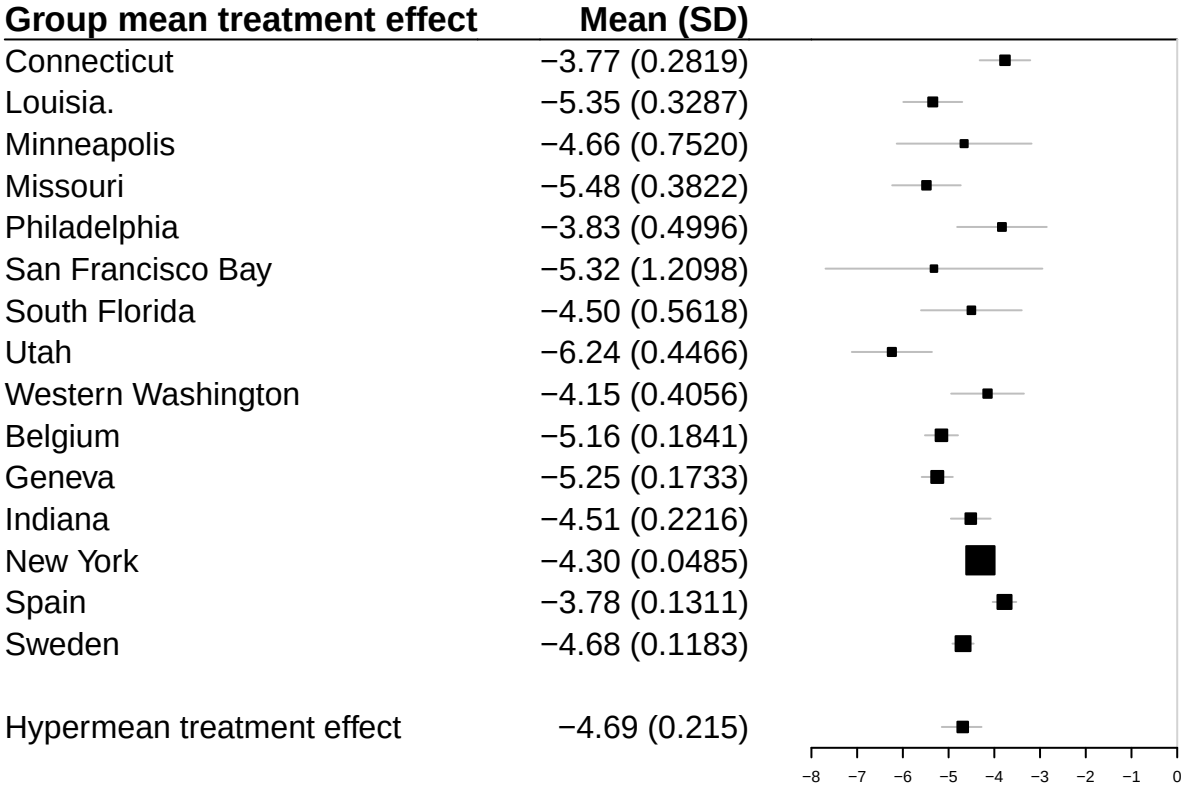
Basic pooling metric (1 - I^2) for the partially pooled model suggests low pooling:

```
##      2.5%      mean     97.5%
## 0.1474570 0.3382654 0.5724073
```

In conclusion, the pooled IFR in general population in the included studies is as follows:

```
##   2.5%   mean  97.5% median     sd
## 0.0058 0.0093 0.0137 0.0092 0.0020
```

We can also summarise this as a forest plot:




### Model with age-specific IFRs

We can modify the above model to include some covariates. A basic structure could include study setting and age of participants. For simplicity we start with median age variable (**to be refined**). If only summary data are used, this model can be written as a modification of the previous one, where

$$\theta_i = \alpha_i + \beta(age_i - 2.5) + \gamma study_i$$

where *age* is median age in the study (in decades). We center age at 25 years of age, so that the main estimate is for the 20-29 age group. Variable *study* is a location indicator (we use Belgium as reference). This simplistic model assumes that each extra decade of life has the same impact in terms of *odds ratios* of dying. (**This can be modified in the future.**) The rest of the model is the same as the previous one.

Data for this model is the same dataset, but without merging of IFRs across age groups:

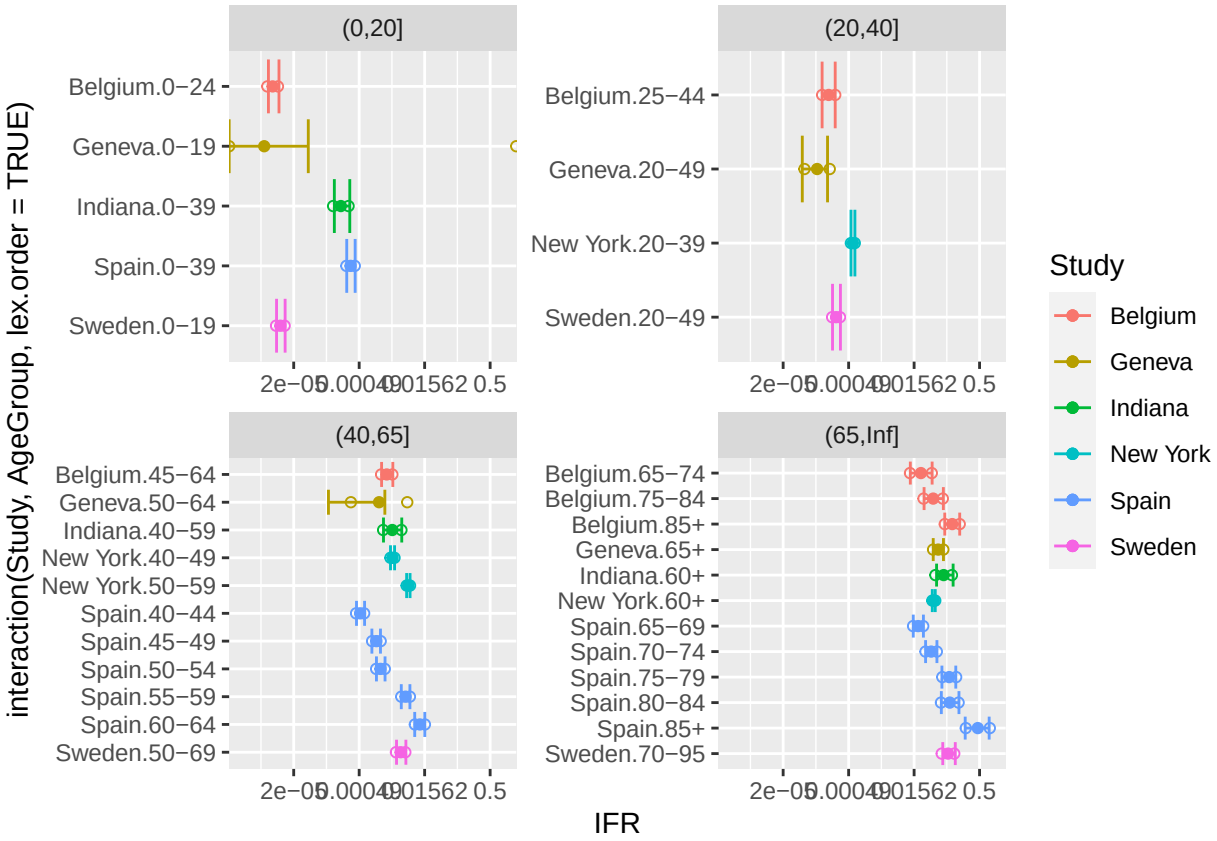
```
## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Transformation introduced infinite values in continuous x-axis
```

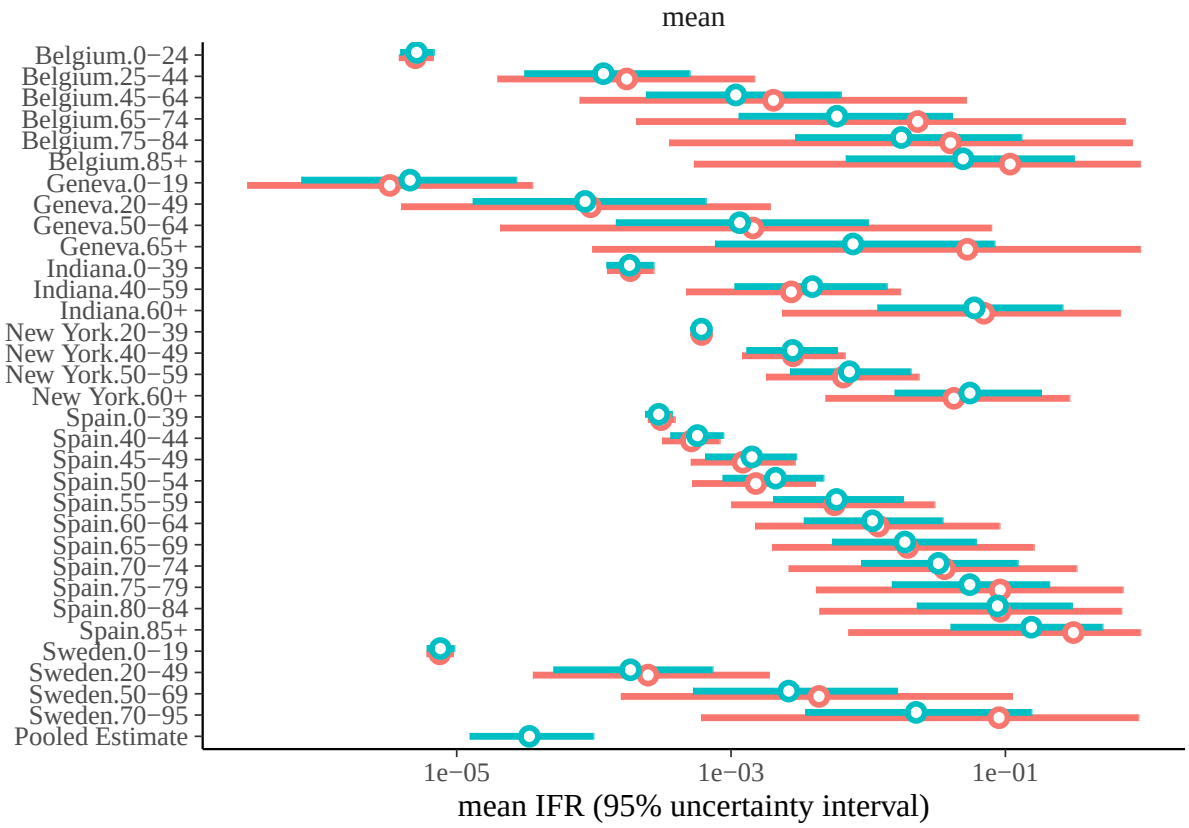



The results are as follows (**will be fixed to inv logit**):

```
## Model type: Rubin model with aggregate data
## Pooling of effects: partial
##
## Aggregate treatment effect (on mean):
## Exponent of hypermean (exp(tau)) =  4.0e-05 with 95% interval 1.2e-05 to 9.9e-05
##
## Group effects omitted, as number of groups is > 20.
##  Use print.baggr() with group = TRUE to print them.
## Covariate (fixed) effects on mean (converted to exp scale):
##            mean   lci  uci
## Median_Age  3.1 2.546  3.9
## countryGe   1.5 0.096  6.6
## countryIn  11.1 1.784 35.7
## countryNy  10.3 1.864 30.4
## countryES   5.5 1.101 15.3
## countrySE   2.1 0.351  7.0
```

By explaining part of the variation with location- and age-specific covariates, we can also see how the partially pooled estimates are narrower than their non-pooled estimates

```
## There is no treatment effect estimated when pooling = 'none'.
## There is no treatment effect estimated when pooling = 'none'.
```




## References

Levin, Andrew T., Kensington B. Cochran, and Seamus P. Walsh. 2020. "ASSESSING THE AGE SPECIFICITY OF INFECTION FATALITY RATES FOR COVID-19: META-ANALYSIS &Amp; PUBLIC POLICY IMPLICATIONS." *medRxiv*, July, 2020.07.23.20160895. https://doi.org/10.1101/2020.07.23.20160895.