Bayesian model for COVID IFR

Witold Wiecek for 1 Day Sooner

Last updated 2020-10-19

1 Introduction

This is a short document describing a Bayesian model for synthesising information on many infection fatality rates (IFRs) into a single estimate that can be made specific to certain age groups or adjusted by co-morbidity status. The analysis presented here is a form of Bayesian meta-analysis, in that our primary objective is to weigh sources of evidence in a way that captures both variability (heterogeneity across different settings) and uncertainty.

Our ultimate objective is to characterise risks in a particular setting, population and time, in a way that is useful to understanding risks of human challenge trials (HCTs). Therefore, as a minimum, we want to incorporate variability into our prediction. Even better would be to understand how different factors can drive heterogeneity. Indeed, *a priori*, we can hypothesise that the three main drivers of differences in IFRs are time-specific, population-specific and otherwise country-specific.

The role of time may be due to new treatments, improvements over time in our ability to treat Covid-19 or selection pressures which may lead to more benign versions of the virus. Country-specific or location-specific factors in IFR data may be driven by under-reporting, health care factors (including access to health care services) or underlying distributions of known risk factors. Additionally, some unknown risk factors (e.g. genetic) may also be operating, in which case controlling for age and co-morbidities will be not sufficient to account for cross-location differences.

To address these drivers of differences in observed IFRs we develop a Bayesian model and apply it to publicly available summary data on IFRs from multiple countries and contexts, with particular focus on the impact of age.

2 Evidence synthesis model

What follows is an adaptation of typical methods of Bayesian evidence synthesis to analysis of IFRs. IFR is a proportion statistic, calculated as the ratio of deaths to infections in some population. Early estimates, e.g. by Verity et al. (2020), place it at over 0.6% globally. However, the risk of death is orders of magnitude higher in particular high risk groups, especially in the elderly, than in the general population.

We can use Bayesian models for repeated binary trials, accounting for the fact that different populations studies at different times have different average probability of events. We use hierarchical modelling framework to assume that the context-specific estimates of IFR_i (measured in different settings, with some uncertainty) are all linked using some common parameters.

The most straight-forward and "canonical" ways to implement such a Bayesian model is by modelling log odds of the event.[1] Deeks (2002) present a general treatment. Note, that for very rare events the odds of mortality are very similar to probability of mortality, but we model events on odds scale as a good "generic" approach to modelling binary data (in this case death following infections). Another advantage of such a

[1] It is also possible to work with IFR_i parameters and treat them as derived from Beta distribution with some "hyperparameters" α and β of Beta distribution, as done by e.g. Carpenter (2016). That approach, however, does not offer an easy way of modelling impact of covariates (e.g. age and co-morbidities) on the rates.

model is that it can use either individual-level or summary data and work with covariates (such as gender, age, time of the study, co-morbidities), captured as odds ratios or risk ratios[2].

Basic models for analysis of binary data can be implemented using existing statistical analysis packages (see, for example, *baggr* by Wiecek and Meager (2020)), by treating IFR as a logit-normal parameter to meta-analyse. However, note that when no deaths are observed, analysis of IFR (equal to observed deaths divided by modelled infections) is problematic. Therefore we propose a "custom" model that built in Stan which treats deaths and *prevalences* as data (rather than the IFRs).

Let d_k denote observed deaths for data point k and assume that logit of prevalence p_k in the population of n_k subjects is obtained from some model. We can then write:

$$d_k \sim \text{Binomial}(n_k, p_k IFR_k) \tag{1}$$

$$\text{logit}(p_k) \sim \mathcal{N}(\mu_k^{(p)}, \sigma_k^{(p)}) \tag{2}$$

where $\sigma_k^{(p)}$ and $\mu_k^{(p)}$ are parameters derived from the existing models of prevalence.

The k data points collected can span many locations (studies); we denote them by loc_k and the total number of locations by N_{loc} We can also collect other covariates impacting the IFRs, such as age groups (which we identify with median age of the population being studied, MedianAge_k). To center our X at the value of interest in our model (risk in 20-30 year olds), we use a transformation `MedianAge/10 - 2.5` to construct our matrix X. We denote all of the covariates using a design matrix X and denote by N_p the number of columns in X. We assume the impact on IFR is on logit scale, same as in the "canonical" logistic models of binary data that we mentioned above:

$$\text{logit}(IFR_k) = \theta_{\text{loc}_k} + X\beta$$

where θ is an N_{loc}-dimensional vector of location-specific (random) effects on IFR and β is N_p dimensional vector of (fixed) covariate effects.

We implement our model in Stan and assume weakly informative priors on all parameters, with prior for τ centered at 1 death per 10,000.

```
model {
  //Uncertain prevalence estimates:
  logit_prevalence ~ normal(mean_prevalence, sd_prevalence);
  //Likelihood of mortality:
  obs_deaths ~ binomial(population, prevalence .* ifr);
  //Hierarchical component of the model (location-specific theta):
  theta_k ~ normal(tau, sigma);

  //Priors:
  tau   ~ normal(logit(.0001), 5);
  sigma ~ normal(0, 10);
  beta  ~ normal(0, 10);
}
```

3 Data on age-specific mortality risk

We used estimates originally collected by Levin, Cochran, and Walsh (2020) to construct the first version of analysis dataset, which we then supplemented with more values extracted from other studies. **Write-up of data collection here.** The input data into our model consists of deaths (treated as known) and prevalences

[2]If only summary data are available, covariates can be defined as study level distributions (e.g. % male)

(treated as logit-distributed parameter with known mean and SD) in all reported age groups in all studies[3]. All of input data are reported in Table 1.

We made only minimal modifications to source data, by 1) imputing the values in Italian study (**David to describe**?), 2) imputing population size in Maranhao (as ratio of the reported number of infections and the mean infection rate) which were not reported and 3) assuming that uncertainty in prevalence 0-29 age group in Iceland is same as in the 30-39 age group since data were missing.

Our model treats number of Covid-attributable deaths as measured without error (due to lack of data) but accounts for uncertainty in infection rates, which are always model-based estimated extracted from various available data sources. We assume that logits of prevalence estimates from available studies are normally distributed, which seems to reproduce almost all of data perfectly, see Figure **??** in the Supplement.

Our construction of age variable and use of all available data (rather than the subset of data available in younger adults only) is necessitated by data limitations: out of 114 data points comprising age-specific estimates of prevalence (or IFR) and counts of deaths, 24 contain individuals aged 20-30 who are of primary interest to us. However, the populations are mixed with regards to age, with typical age groupings such as 19-49, 20-49, 20-39, 0-49 used instead. In fact, we find only one estimate out of 24 that is entirely specific to the 20-29 age group (Brazilian state of Maranhao).

4 Results: age-specific risk of Covid mortality

There were no issues with convergence of the Bayesian model. We set number of iterations to 5,000 and used 4 chains, with `max_treedepth` option set to 15. There were no divergent transitions and effective sample size was greater than 3346 for all of 596 modeled parameters. For the three main parameters in the model we obtained the following:

```
## Inference for Stan model: ifr_with0.
## 4 chains, each with iter=5000; warmup=2500; thin=1;
## post-warmup draws per chain=2500, total post-warmup draws=10000.
##
##          mean se_mean   sd  2.5%   25%   50%   75%   98% n_eff Rhat
## tau     -8.67       0 0.15 -8.97 -8.77 -8.66 -8.57 -8.36  9819    1
## sigma    0.67       0 0.12  0.48  0.59  0.66  0.74  0.96 10086    1
## beta[1]  1.07       0 0.01  1.05  1.07  1.07  1.08  1.09  4328    1
##
## Samples were drawn using NUTS(diag_e) at Mon Oct 19 15:36:26 2020.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

From these parameters we can predict average risks for subjects of any given age x, by using the posterior distribution of $\tau + (10x + 2.5)\beta$ (where 2.5 and 10 refer to the transformation we applied to `MedianAge` inputs).

4.1 Average infection fatality risk in young subjects

Since we centered our `MedianAge` at 25 years in constructing our matrix X, we can now obtain model-estimated risk for a typical HCT population (aged 20 to 30, with median 25) by ignoring the β coefficient and examining τ and σ only. We find that the average IFR for this group (equal to $\frac{\exp(\tau)}{\exp(1+\tau)}$) is 1.75×10^{-4} (with 95% interval from 1.27×10^{-4} to 2.33×10^{-4}). That means, on average, slightly under 2 deaths per 10,000 infections in the studied datasets.

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

4.2 Heterogeneity in IFRs

However, there is a considerable variability in IFRs across different locations/dataset that we should consider. To take into account parameter σ, we can generate draws from the $\mathcal{N}(\tau, \sigma^2)$ distribution, corresponding to a hypothetical IFR in a new source of data. 95% interval for such model runs from 4.41×10^{-5} to 6.93×10^{-4}. Since the model works a logistic scale, another way of interpreting the across-dataset variability is reporting the fold-impact of σ on the mean IFR; here, we obtain on average a 3.96-fold increase (decrease) in IFR per 2σ increase (decrease).

The lower end of the 95% interval, 4.41×10^{-5}, is not extreme given input data, where the "crude" mean IFR (based on mean prevalence only) is below 7 per 10,000 for all data, except for South Florida, and as low as 1.4 per 10,000 in Utah, in the population aged 19-44.

```
## # A tibble: 18 x 6
## # Groups:   Study [17]
##    Study                         AgeGroup Deaths Population       ir crude_ifr
##    <chr>                         <chr>     <dbl>      <dbl>    <dbl>     <dbl>
##  1 Belgium                       0-24        0      3228894  0.06      0
##  2 Indiana                       0-39       20      3545671  0.0305    0.000185
##  3 New York                      20-39     482      5408503  0.146     0.000610
##  4 Spain                         0-39      225     19490155  0.0373    0.000310
##  5 South Florida                 19-49      61      2512589  0.009     0.00270
##  6 Louisia.                      19-49      85      1878546  0.074     0.000611
##  7 Minneapolis                   19-49      18      1615203  0.023     0.000485
##  8 Missouri                      20-49      18      2347889  0.034     0.000225
##  9 Philadelphia                  19-49      51      1446645  0.059     0.000598
## 10 San Francisco Bay             19-49      25      3384373  0.011     0.000672
## 11 Utah                          19-44       3      1227871  0.018     0.000136
## 12 Western Washington            20-39       8      1332151  0.013     0.000462
## 13 Iceland                       0-29        0       135576  0.00409   0
## 14 Korea                         0-29        0     15623365  0.000555 0
## 15 New Zealand                   0-29        0      1911472  0.000552 0
## 16 Brazil Maranhao and Sao Luis 10-19       21      1301698. 0.43      0.0000375
## 17 Brazil Maranhao and Sao Luis 20-29       47      1266585. 0.492     0.0000754
## 18 Italy Report                  18-34      43.1   10251000  0.021     0.0002
```

We can assess this heterogeneity by inspecting the distribution of random effects in the model transformed into IFRs, i.e. the inverse logit transformation θ parameters. The largest (posterior mean) IFR value of θ is 6.29×10^{-4} in Castiglione d'Adda. The smallest posterior mean is 5.74×10^{-5} in Belgium.

4.3 Predictive checks for the model

We constructed posterior predictive distributions for number of deaths in each of the inputs by using the `generated quantities` functionality of Stan. Figure 1 compares the posterior means and 95% intervals with observed deaths. Out of 114 observations that were used to fit the model, 108 were within 95% intervals of the posterior predictive distributions. We observed the largest discrepancies occured in Spanish data. Overall, we conclude that the simple binomial model we used here is flexible enough to capture both age-specific risk increases and heterogeneity in IFRs across settings/countries.

4.4 Robustness checks

- Exclusion of small studies and test & trace data
- Median age: exclusion of 80+
- Median age: different method of age imputation
- Model with time

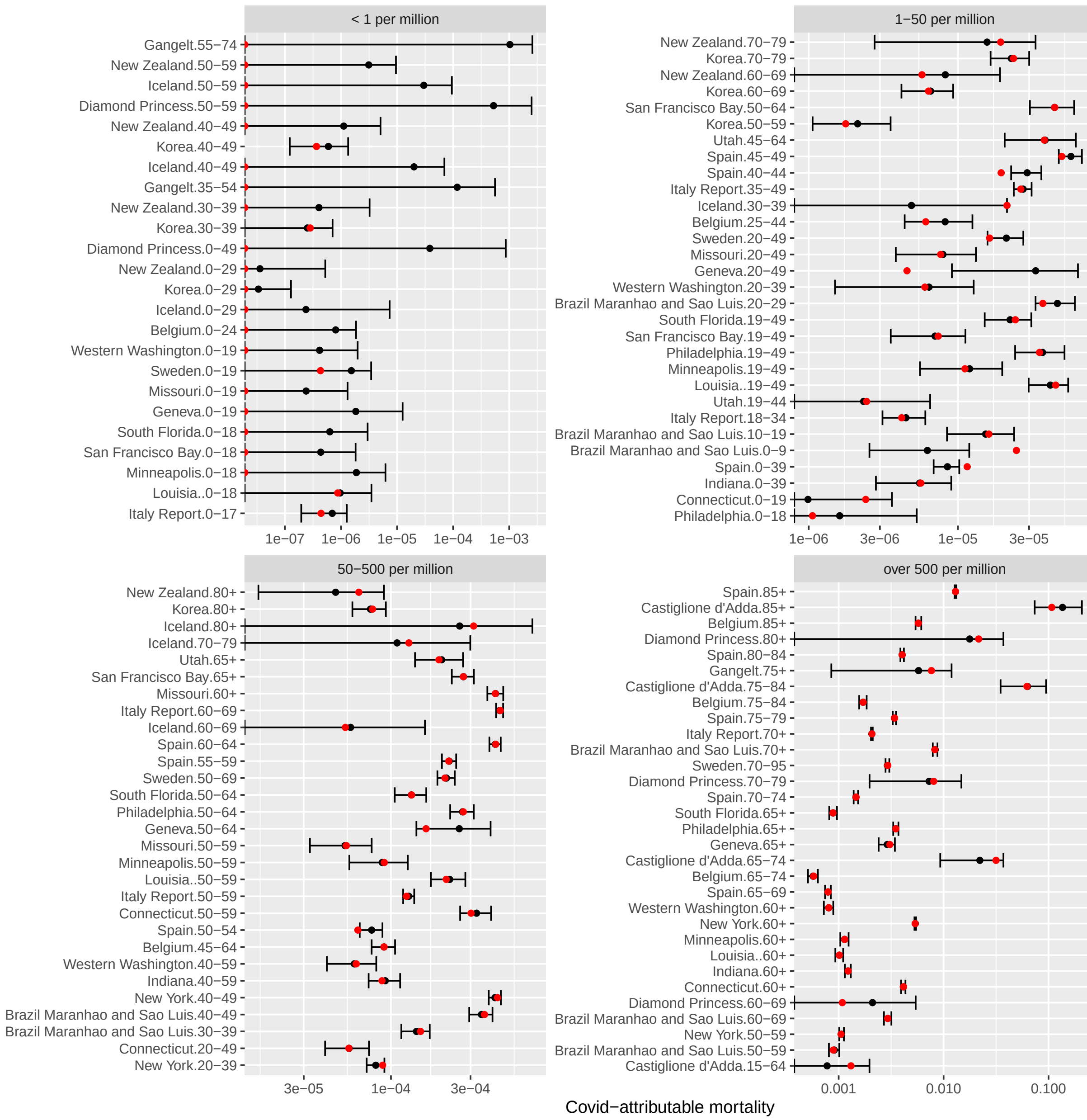


Figure 1: Comparison of model estimates (black) with observed mortality (red), compared on logarithmic scale. Bars are 95% posterior interval; point is the mean. For better clarity, we grouped the plot into four panels according to observed mortality. X axes on each panel differ. For many low-risk populations no deaths were reported: we indicate this by plotting a red point on the left-hand side of the panel plot.

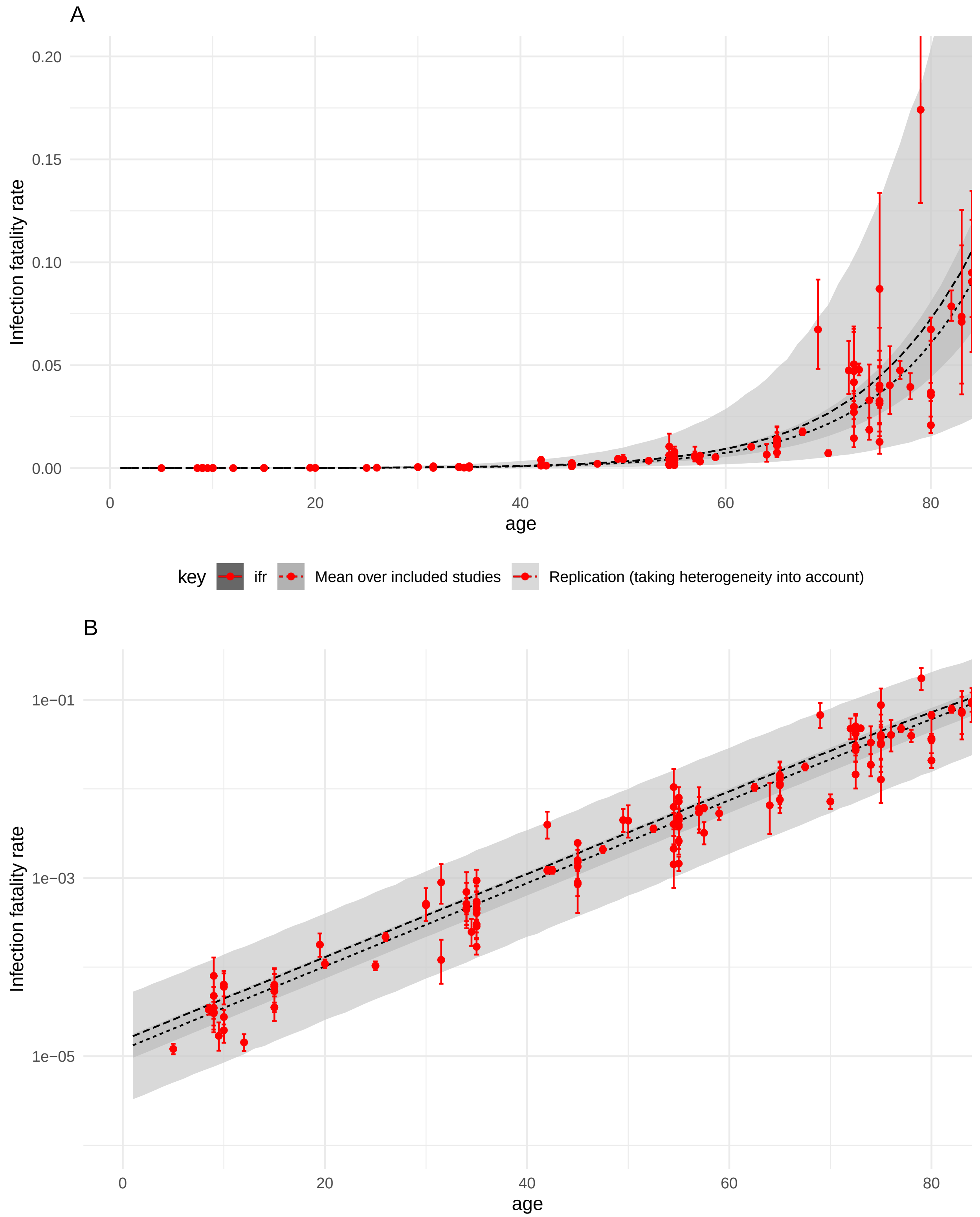


Figure 2: IFR as a function of age. Narrower ribbon corresponds to the 95% posterior interval of average across all included studies (tau parameter in the meta-analysis model), while the wider band takes into account heterogeneity (tau and sigma). Red points are model estimates of mean IFRs in partciular studies, with bars representing 95% posterior intervals. Panel A is untransformed data and panel B is same data on log 10 scale.

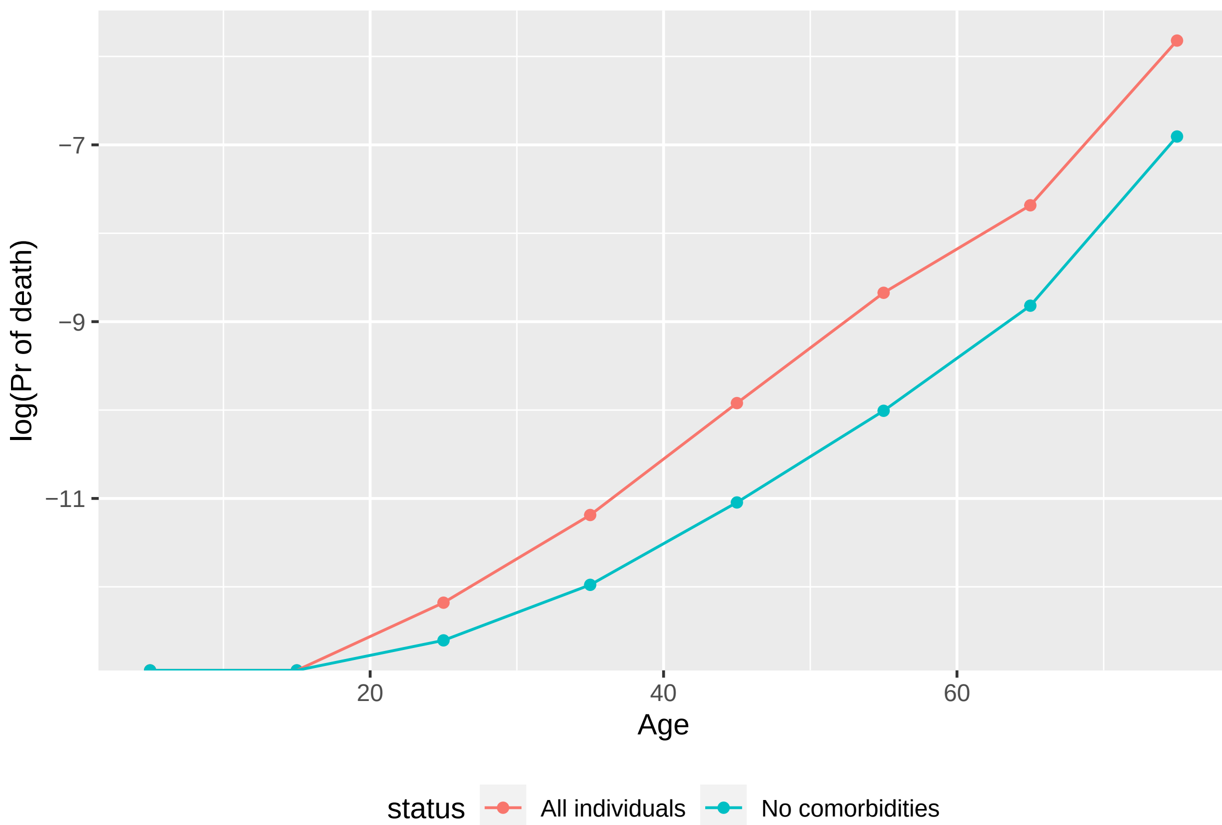
5 What is the risk in healthy individuals?

Data for this section has been provided by OpenSAFELY (https://opensafely.org/) and is as described by Williamson et al. (2020) in a recent article on Covid-19 mortality risk factors for 10,926 Covid-19 deaths in England. We group the total of 21,444,863 individuals into high and low risk groups as follows:

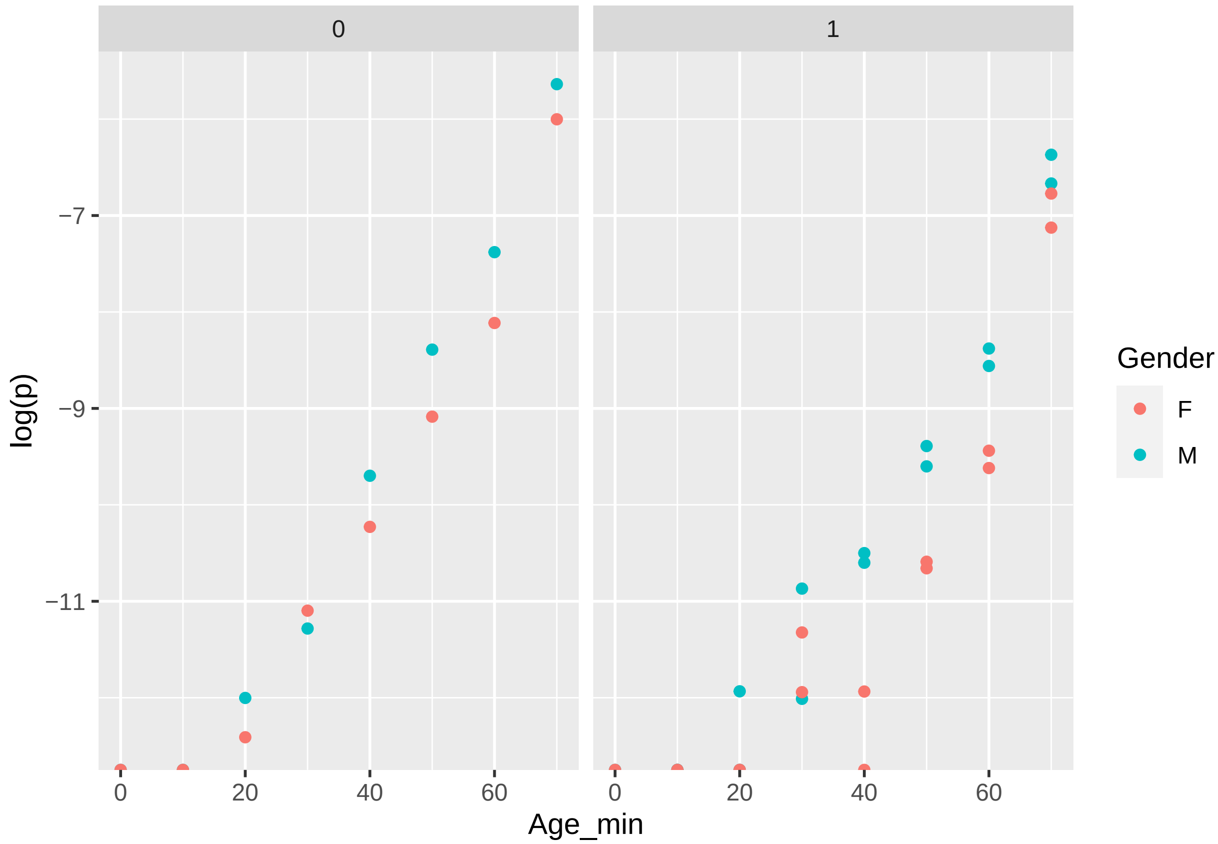
Definition of co-morbiditiy goes here

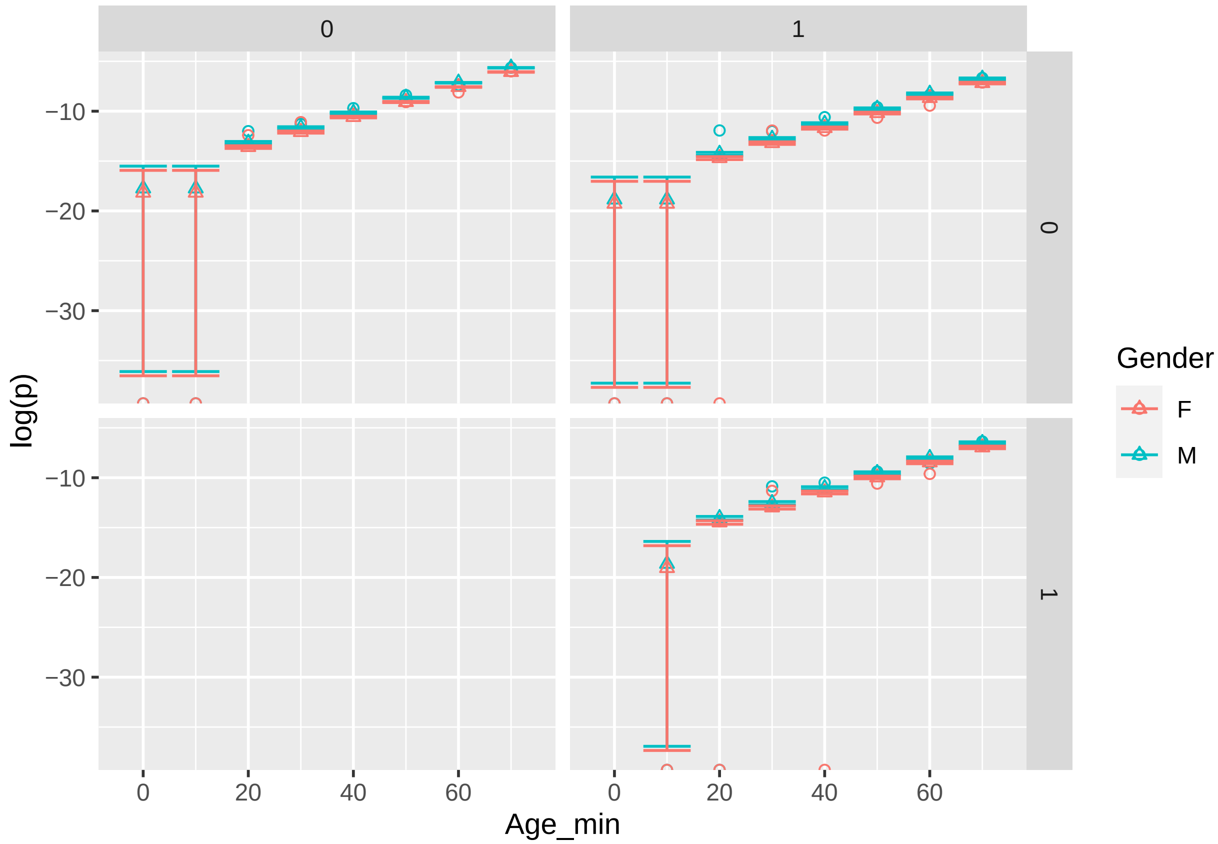
In contrast to the cited publication, we include records of individuals under 18 in our assessment.

```
## ‘summarise()‘ regrouping output by ’Age_min’ (override with ‘.groups‘ argument)
```




```
## # A tibble: 8 x 4
## # Groups:   age_group_min [8]
##   age_group_min    all_risk healthy_risk      rr
##           <dbl>       <dbl>        <dbl>   <dbl>
## 1             0 0            0           NaN
## 2            10 0            0           NaN
## 3            20 0.00000514   0.00000335   0.653
## 4            30 0.0000138    0.00000629   0.454
## 5            40 0.0000491    0.0000160    0.325
## 6            50 0.000171     0.0000450    0.263
## 7            60 0.000460     0.000148     0.321
## 8            70 0.00297      0.00100      0.338
```


0
1
−7
−9
−11
log(p)
0
20
40
60
0
20
40
60
Age_min
Gender
F
M


0
1
0
1
−10
−20
−30
log(p)
0
20
40
60
Age_min
Gender
F
M

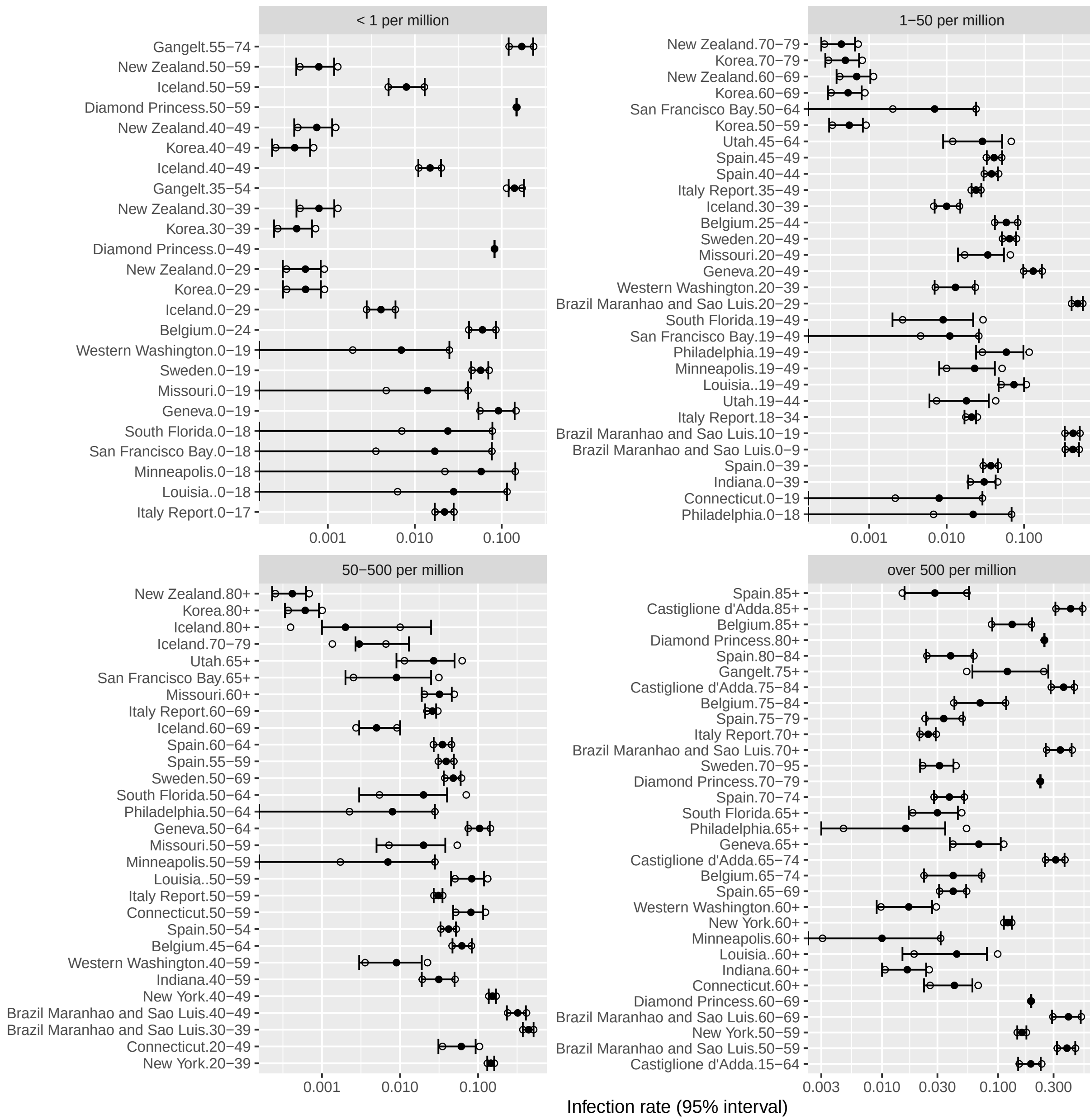


Figure 3: Comparison of model-estimated prevalences (95% CI's reported by modelling studies) collected by Levin, Cochran, and Walsh (2020) and our distributional assumptions: additional circles show 95% CIs recreated by assuming logit-normal distribution of prevalence. We group studies into 4 bands of mortality to mirror earlier figures. Please note that this approach produces discrepancies in a number of US estimates where the confidence intervals were skewed toward including 0. However, since we do not have access to source data, we decided to use the logit-normal assumption for all estimates. This assumption may have an effect of overestimating mortality risk in settings where prevalence was very low.

Table 1: Complete table of inputs used by the meta-analysis model and crude IFR's.

							Infection rate (%)			
Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	0-9	29	1.2e+06	42.60	51.30	33.80	0.06
Italy Report	Country	Seroprevalence	2020-08-01	0-17	5	1.0e+07	2.20	2.80	1.70	0.02
Louisia.	USState	IFR	2020-05-06	0-18	1	1.1e+06	2.80	11.50	0.00	0.03
Minneapolis	USState	IFR	2020-06-04	0-18	0	9.7e+05	5.80	14.30	0.00	0.00
Philadelphia	USState	IFR	2020-05-17	0-18	1	9.4e+05	2.20	6.90	0.00	0.05
San Francisco Bay	USState	IFR	2020-05-25	0-18	0	1.6e+06	1.70	7.70	0.00	0.00
South Florida	USState	IFR	NA	0-18	0	1.3e+06	2.40	7.80	0.00	0.00
Missouri	USState	IFR	2020-05-23	0-19	0	1.5e+06	1.40	4.10	0.00	0.00
Connecticut	USState	IFR	2020-05-28	0-19	2	8.3e+05	0.80	2.90	0.00	0.30
Geneva	City	IFR	2020-06-01	0-19	0	8.0e+04	9.17	14.06	5.40	0.00
Sweden	Country	IFR	2020-06-18	0-19	1	2.3e+06	5.73	7.00	4.45	0.01
Western Washington	USState	IFR	NA	0-19	0	1.0e+06	0.70	2.50	0.00	0.00
Belgium	Country	IFR	2020-05-16	0-24	0	3.2e+06	6.00	8.60	4.20	0.00
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	10-19	21	1.3e+06	43.00	52.40	33.50	0.04
Iceland	Country	IFR	2020-06-14	0-29	0	1.4e+05	0.41	0.60	0.28	0.00
Korea	Country	IFR	2020-07-11	0-29	0	1.6e+07	0.06	0.08	0.03	0.00
New Zealand	Country	IFR	2020-07-09	0-29	0	1.9e+06	0.06	0.08	0.03	0.00
Indiana	USState	IFR	NA	0-39	20	3.5e+06	3.05	4.30	1.90	0.18
Spain	Country	IFR	2020-07-15	0-39	225	1.9e+07	3.73	4.60	2.93	0.31
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	20-29	47	1.3e+06	49.20	57.30	41.10	0.08
Italy Report	Country	Seroprevalence	2020-08-01	18-34	43	1.0e+07	2.10	2.40	1.70	0.20
New York	USState	IFR	2020-05-23	20-39	482	5.4e+06	14.60	16.10	13.10	0.61
Western Washington	USState	IFR	NA	20-39	8	1.3e+06	1.30	2.30	0.70	0.46
Philadelphia	USState	IFR	2020-05-17	19-49	51	1.4e+06	5.90	9.80	2.40	0.60
Utah	City	IFR	2020-06-19	19-44	3	1.2e+06	1.80	3.50	0.60	0.14
Louisia.	USState	IFR	2020-05-06	19-49	85	1.9e+06	7.40	10.00	4.70	0.61
Minneapolis	USState	IFR	2020-06-04	19-49	18	1.6e+06	2.30	4.20	0.80	0.48
San Francisco Bay	USState	IFR	2020-05-25	19-49	25	3.4e+06	1.10	2.60	0.00	0.67
South Florida	USState	IFR	NA	19-49	61	2.5e+06	0.90	2.20	0.20	2.70
Missouri	USState	IFR	2020-05-23	20-49	18	2.3e+06	3.40	5.50	1.40	0.23

Table 1: Complete table of inputs used by the meta-analysis model and crude IFR's. *(continued)*

Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
Belgium	Country	IFR	2020-05-16	25-44	18	3.0e+06	5.90	8.30	4.20	0.10
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	30-39	163	1.1e+06	44.40	51.40	37.40	0.34
Connecticut	USState	IFR	2020-05-28	20-49	75	1.3e+06	6.10	9.30	3.10	0.92
Diamond Princess	Small	IFR	2020-04-01	0-49	0	1.2e+03	8.26	8.28	8.24	0.00
Geneva	City	IFR	2020-06-01	20-49	1	2.2e+05	13.12	17.00	9.75	0.03
Iceland	Country	IFR	2020-06-14	30-39	1	4.7e+04	1.00	1.50	0.70	2.13
Korea	Country	IFR	2020-07-11	30-39	2	7.1e+06	0.04	0.07	0.02	0.64
New Zealand	Country	IFR	2020-07-09	30-39	0	6.2e+05	0.08	0.12	0.04	0.00
Sweden	Country	IFR	2020-06-18	20-49	63	3.9e+06	6.50	7.84	5.16	0.25
Castiglione d'Adda	Small	IFR	NA	15-64	4	3.1e+03	19.10	23.24	14.86	6.86
Italy Report	Country	Seroprevalence	2020-08-01	35-49	334	1.3e+07	2.40	2.80	2.10	1.10
Spain	Country	IFR	2020-07-15	40-44	78	4.0e+06	3.80	4.60	3.00	0.51
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	40-49	290	8.0e+05	32.20	41.00	23.40	1.13
Gangelt	Small	IFR	NA	35-54	0	3.6e+03	14.00	18.00	12.00	0.00
Iceland	Country	IFR	2020-06-14	40-49	0	4.3e+04	1.50	2.00	1.10	0.00
Korea	Country	IFR	2020-07-11	40-49	3	8.2e+06	0.04	0.06	0.02	0.88
New York	USState	IFR	2020-05-23	40-49	1026	2.4e+06	15.30	17.00	13.70	2.85
New Zealand	Country	IFR	2020-07-09	40-49	0	5.9e+05	0.07	0.11	0.04	0.00
Spain	Country	IFR	2020-07-15	45-49	196	3.9e+06	4.10	5.20	3.30	1.21
Indiana	USState	IFR	NA	40-59	148	1.7e+06	3.14	5.00	1.90	2.81
Western Washington	USState	IFR	NA	40-59	69	1.1e+06	0.90	1.90	0.30	6.88
Spain	Country	IFR	2020-07-15	50-54	230	3.6e+06	4.20	5.20	3.30	1.51
Louisia.	USState	IFR	2020-05-06	50-59	126	5.9e+05	8.30	11.90	4.50	2.59
Minneapolis	USState	IFR	2020-06-04	50-59	47	5.2e+05	0.70	2.80	0.00	13.03
Missouri	USState	IFR	2020-05-23	50-59	43	8.0e+05	2.00	3.80	0.50	2.70
Philadelphia	USState	IFR	2020-05-17	50-64	290	1.1e+06	0.80	2.80	0.00	33.95
Utah	City	IFR	2020-06-19	45-64	24	6.3e+05	2.90	5.20	0.90	1.31
Belgium	Country	IFR	2020-05-16	45-64	280	3.1e+06	6.20	8.30	4.70	1.47
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	50-59	533	6.0e+05	39.10	46.10	32.10	2.27
Connecticut	USState	IFR	2020-05-28	50-59	157	5.2e+05	8.10	11.60	4.80	3.73
Diamond Princess	Small	IFR	2020-04-01	50-59	0	4.0e+02	14.82	14.90	14.70	0.00

Table 1: Complete table of inputs used by the meta-analysis model and crude IFR's. *(continued)*

Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
Iceland	Country	IFR	2020-06-14	50-59	0	4.2e+04	0.80	1.30	0.50	0.00
Italy Report	Country	Seroprevalence	2020-08-01	50-59	1196	9.6e+06	3.10	3.50	2.70	4.00
Korea	Country	IFR	2020-07-11	50-59	15	8.5e+06	0.06	0.08	0.03	3.20
New York	USState	IFR	2020-05-23	50-59	2764	2.6e+06	16.00	17.50	14.60	6.58
New Zealand	Country	IFR	2020-07-09	50-59	0	6.3e+05	0.08	0.12	0.04	0.00
San Francisco Bay	USState	IFR	2020-05-25	50-64	66	1.5e+06	0.70	2.40	0.00	6.37
South Florida	USState	IFR	NA	50-64	169	1.3e+06	2.00	4.00	0.30	6.64
Geneva	City	IFR	2020-06-01	50-64	16	9.9e+04	10.45	14.11	7.31	1.55
Spain	Country	IFR	2020-07-15	55-59	758	3.4e+06	3.90	4.90	3.10	5.69
Sweden	Country	IFR	2020-06-18	50-69	504	2.4e+06	4.81	5.98	3.64	4.38
Spain	Country	IFR	2020-07-15	60-64	1249	2.9e+06	3.50	4.60	2.70	12.14
Gangelt	Small	IFR	NA	55-74	0	3.1e+03	17.00	23.00	12.00	0.00
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	60-69	1155	3.9e+05	40.30	51.40	29.10	7.28
Diamond Princess	Small	IFR	2020-04-01	60-69	1	9.2e+02	19.18	19.30	19.10	5.65
Iceland	Country	IFR	2020-06-14	60-69	2	3.8e+04	0.50	1.00	0.30	10.66
Italy Report	Country	Seroprevalence	2020-08-01	60-69	3274	7.2e+06	2.60	2.90	2.10	17.40
Korea	Country	IFR	2020-07-11	60-69	41	6.5e+06	0.05	0.08	0.03	11.91
New Zealand	Country	IFR	2020-07-09	60-69	3	5.2e+05	0.07	0.10	0.04	8.33
Spain	Country	IFR	2020-07-15	65-69	1905	2.4e+06	4.10	5.30	3.10	19.35
Castiglione d'Adda	Small	IFR	NA	65-74	17	5.4e+02	31.30	37.30	25.40	100.95
Belgium	Country	IFR	2020-05-16	65-74	663	1.1e+06	4.10	7.20	2.30	14.10
Connecticut	USState	IFR	2020-05-28	60+	3633	8.8e+05	4.20	6.00	2.30	98.77
Indiana	USState	IFR	NA	60+	1864	1.5e+06	1.65	2.40	1.00	74.71
Louisia.	USState	IFR	2020-05-06	60+	1053	1.0e+06	4.40	8.00	1.50	23.07
Minneapolis	USState	IFR	2020-06-04	60+	928	8.1e+05	1.00	3.20	0.00	114.11
Missouri	USState	IFR	2020-05-23	60+	620	1.5e+06	3.20	4.60	1.90	13.21
Spain	Country	IFR	2020-07-15	70-74	3230	2.2e+06	3.80	5.10	2.80	38.56
Western Washington	USState	IFR	NA	60+	700	8.7e+05	1.70	2.70	0.90	47.31
New York	USState	IFR	2020-05-23	60+	24376	4.5e+06	12.10	13.10	11.20	44.33
Geneva	City	IFR	2020-06-01	65+	257	8.4e+04	6.82	10.53	3.83	45.09
New Zealand	Country	IFR	2020-07-09	70-79	7	3.6e+05	0.04	0.07	0.02	44.30
Diamond Princess	Small	IFR	2020-04-01	70-79	8	1.0e+03	23.05	23.20	23.00	34.90

Table 1: Complete table of inputs used by the meta-analysis model and crude IFR's. *(continued)*

Study location	Scale	Type	End date	Age	Deaths	N	Mean	2.5%	97.5%	IFR/100k
Iceland	Country	IFR	2020-06-14	70-79	3	2.3e+04	0.30	1.30	0.27	42.71
Korea	Country	IFR	2020-07-11	70-79	84	3.6e+06	0.05	0.07	0.03	47.89
Philadelphia	USState	IFR	2020-05-17	65+	2374	6.8e+05	1.60	3.50	0.30	219.23
San Francisco Bay	USState	IFR	2020-05-25	65+	333	1.2e+06	0.90	2.50	0.20	30.20
Utah	City	IFR	2020-06-19	65+	71	3.7e+05	2.70	5.00	0.90	7.19
South Florida	USState	IFR	NA	65+	1060	1.2e+06	3.00	4.50	1.70	29.38
Spain	Country	IFR	2020-07-15	75-79	6175	1.8e+06	3.40	5.00	2.40	100.28
Sweden	Country	IFR	2020-06-18	70-95	4485	1.5e+06	3.12	4.12	2.13	94.03
Castiglione d'Adda	Small	IFR	NA	75-84	25	4.0e+02	36.60	44.90	28.30	170.34
Belgium	Country	IFR	2020-05-16	75-84	1182	6.9e+05	7.00	11.70	4.20	24.45
Brazil Maranhao and Sao Luis	Region	Seroprevalence	NA	70+	2788	3.4e+05	34.30	42.90	25.70	24.07
Gangelt	Small	IFR	NA	75+	9	1.2e+03	12.00	27.00	6.00	63.83
Italy Report	Country	Seroprevalence	2020-08-01	70+	21271	1.0e+07	2.50	2.90	2.10	83.00
Spain	Country	IFR	2020-07-15	80-84	5192	1.3e+06	3.90	6.10	2.40	103.33
Diamond Princess	Small	IFR	2020-04-01	80+	5	2.2e+02	25.00	25.10	24.90	86.42
Iceland	Country	IFR	2020-06-14	80+	4	1.3e+04	0.20	2.50	0.10	156.56
Korea	Country	IFR	2020-07-11	80+	144	1.9e+06	0.06	0.09	0.03	127.21
New Zealand	Country	IFR	2020-07-09	80+	12	1.9e+05	0.04	0.06	0.02	153.85
Spain	Country	IFR	2020-07-15	85+	21248	1.6e+06	2.85	5.60	1.56	455.37
Castiglione d'Adda	Small	IFR	NA	85+	16	1.5e+02	42.10	53.10	31.10	255.07
Belgium	Country	IFR	2020-05-16	85+	1878	3.3e+05	13.20	19.60	8.90	43.55

References

Carpenter, Bob. 2016. "Hierarchical Partial Pooling for Repeated Binary Trials." https://mc-stan.org/users/documentation/case-studies/pool-binary-trials.html.

Deeks, Jonathan J. 2002. "Issues in the Selection of a Summary Statistic for Meta-Analysis of Clinical Trials with Binary Outcomes." *Statistics in Medicine* 21 (11): 1575–1600. https://doi.org/10.1002/sim.1188.

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.

Verity, Robert, Lucy C. Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, et al. 2020. "Estimates of the Severity of Coronavirus Disease 2019: A Model-Based Analysis." *The Lancet Infectious Diseases* 0 (0). https://doi.org/10.1016/S1473-3099(20)30243-7.

Wiecek, Witold, and Rachael Meager. 2020. "Baggr: Bayesian Aggregate Treatment Effects Package." Zenodo. https://doi.org/10.5281/zenodo.3813443.

Williamson, Elizabeth J., Alex J. Walker, Krishnan Bhaskaran, Seb Bacon, Chris Bates, Caroline E. Morton, Helen J. Curtis, et al. 2020. "Factors Associated with COVID-19-Related Death Using OpenSAFELY." *Nature* 584 (7821): 430–36. https://doi.org/10.1038/s41586-020-2521-4.