

Association Between BCG Policy is Significantly Confounded by Age and is Unlikely to Alter Infection or Mortality Rates

Stefan Kirov, PhD

Bristol Myers Squibb

Abstract

Recently a number of publications looked at the association between COVID-19 morbidity and mortality on one hand and countries' policies with respect to BCG vaccination on the other. This connection arises from differences in the rates of infection in countries where BCG vaccination is mandatory compared to countries where mandatory vaccination no longer exists or was never implemented in the first place. In at least 2 preprint publications the authors expressed the view that the "known immunological benefits" of BCG vaccination may be behind the biological mechanism of such observation.

One study accounted for different income levels in different groups. Another study did not attempt to do so, instead exploring the differences between countries where a booster shot is given vs others where no such practice exists (finding no connection).

Both of these studies did not explore other potential confounding factors. Meanwhile the press has focused on these headlines and pushed the narrative that BCG vaccination is causally linked to infection and mortality rates. This poses a serious challenge, demonstrated by the recently initiated clinical trials on BCG vaccination within the COVID19 context.

This study shows that population age is a very significant confounding factor that explains the rates of infections much better and has a solid biology mechanism which explains this correlation. It suggests that BCG vaccination may have little or no causal link to infection rates and advises that any follow up studies should control for several confounding factors, such as population age, ethnicity, rates of certain chronic diseases, time from community spread start date, major public policy decisions and income levels.

Introduction

BCG vaccine has been associated in multiple studies with effects beyond protection against tuberculosis, which is the original target of the intervention(1). As such, BCG vaccination has been shown to enhance the protection provided by H1N1 vaccine(2). Another study observed that prior BCG vaccination attenuates yellow fever vaccine associated viremia(3). These findings prompted two independent studies(4,5) in which a strong correlation between the presence of BCG vaccination and a reduced rate of COVID19 infections and/or death rates was observed.

These observations were also communicated by the press (6–9). While some of these publications authors contacted additional subject matter experts who cautioned that the studies are not yet peer reviewed and urged patience(7), there is danger that the observations will be over interpreted and used for policy decisions. There are currently at least 2 ongoing studies focusing on the effect of BCG vaccination in the context of COVID19 infections(10) any decisions should be taken only after conclusive evidence is presented after the end of the trials.

At the same time, some of the established risk factors for COVID19 hospitalization and death such as age(11) and BMI(12) are easy to evaluate in the same context (per country). Given the extreme urgency and the potential for serious consequences I decided to explore this matter in greater depth.

Materials and Methods

Data on average population age per country and number of infections was collected from Wikipedia. BCG policy and income level data was obtained from one of the original studies. BCG and rubella immunization rates were obtained from WHO website(13).

Country names were cleaned and data was merged in R (Rstudio 1.1.456 on Ubuntu 16 Linux, R version 3.2.3). All scripts and data files are available from <https://github.com/kirovsa/covid19-bcg>.

As with one of the original studies the countries with population of less than 1M were excluded(4). For analysis of mortality rates countries with no deaths on record were excluded as these are likely to be in the very initial phase of the epidemic and will introduce significant noise.

To determine the effect of different factors I used lm function from stats package. Evaluating factors effects on infections in the presence of random effect was done with lme from nlme package. Log likelihood was tested with lrtest from lmttest package.

BCG policy and income level were coded according to one of the studies that found the association with COVID19 infection(4):

BCG Policy

1 = current universal policy

2= used to recommend, not anymore

3 = never had universal policy

2018 FY income level:

Low income (L) -1

Lower middle income (LM)-2

Upper middle income (UM)-3

High income (H)-4

Results

Based on a significant amount of accumulated data, age is a significant factor predicting hospitalization of COVID19 patients and fatal outcome. Younger people also seem more likely to remain

asymptomatic. Therefore I decided to evaluate a linear regression model that accounts for 3 factors- BCG policy, income level and median age per country.

While the model as a whole explains very well the differences in infection rates across countries, the most significant factor was income level, followed by median age. BCG policy was significant but lagged behind the other factors (Figure 1). However, BCG immunization rates was not significant in this model at alpha level at 0.05 ($p=0.088$). The likelihood test did find that the BCG policy had an effect ($p=0.0028$) compared to the full model, however this was not true for BCG vaccination rates ($p=0.08$). If there is a causal link between BCG vaccination and COVID19 infection rates one would expect this association to hold or even get stronger, something I did not find evidence for.

The Pearson correlation between median age and infection rates was also much higher at $R=0.774$ than the reported correlation between the BCG policy and the infection rates at $R=0.521$ or the reported correlation between start date of BCG vaccination and infection rates ($R=0.21$).

The correlation between number of cases per million people with the median age in a country does not change substantially between different policy categories (Figure 2A), though there was some separation between categories 1 and 3. This can only be evaluated for countries with high rates of infection and also higher median age. When the BCG immunization rates were used instead of the policy there was no association (Figure 2B).

I also explored potential connection between countries with higher rubella immunization rates vs those with lower rates (separated in categories by 50% threshold) and COVID19 infections. While this variable on its own showed significant association ($p<0.0001$) with the observed infection rates per country, it appeared that the effect is the opposite of what would be expected (Figure 2C) with countries with low immunization rates scoring better in terms of infection rate. After the inclusion of other factors such as median age and income level this association was not significant at $\alpha=0.05$ ($p=0.056$).

Since income levels are unlikely to drive infection rates I decided to compare the performance of median age and BCG policy. The data showed that median age explains the variance in the number of COVID19 cases better than the BCG policy either with or without income level adjustment (Figure 3). The median age explained 60% of the variability vs 30% for BCG policy. In a mixed model where income levels are considered a random factor median age again appears to be more important than BCG policy (Figure 4). BCG rates were again non-significant at $p=0.0798$.

Next, I looked at the median age distribution in different income levels and BCG policy categories (Figure 5). There was a strong association between median age and BCG policy with or without income level adjustment ($p<0.0001$). The same is true for median age and income level ($p<0.0001$). I also explored associations with mortality rates (Figure 6). Again, there was demonstrably better correlation between median age and mortality rates ($R=0.653$) compared to the correlation with start date of BCG vaccination policy reported in one of the studies ($R=0.54$)(4).

BMI was another strong confounding factor in the context of mortality rates(Figure 7). Countries with normal BMI were without exception in the policy category 1 (mandatory BCG vaccination). In addition, death rates were substantially higher in countries with high BMI ($p<0.0001$).

Discussion

While observational studies are a valid and useful tool, there are also serious obstacles interpreting the data correctly(14). In the specific case of the correlation between BCG vaccination policy and COVID19 outcomes it is clear that important confounding factors may have been missed. An excellent outline of these obstacles was given by Emily MacLean(15). In addition to missing hidden factors, the critique in the blog goes further to challenge the biological plausibility of the BCG vaccine-COVID19 connection. It seems this is a very reasonable concern, given that the only established connection

between BCG vaccination and protection against viral infections seems to be within the scope of actual anti-viral vaccines(2,3). On the other hand, the biological rationale for causal link between age and COVID19 morbidity and mortality seems a lot more straightforward; if we follow the Occam's razor we should prioritize this link over BCG vaccination. I also want to emphasize that the association observed in this work between infection rates and rubella immunization are almost certainly spurious. The arguments so far is that early childhood vaccinations might be protective, which is the opposite of our observation. Prior preclinical research(16) that was done during one of the previous coronavirus crisis shows clearly that childhood vaccinations are unlikely to drive different outcomes of COVID19 infections.

This is further enhanced by the data presented in this study. I need to emphasize that I have not included a number of other potentially confounding factors such as blood pressure, public policy (mandatory travel restrictions, use of masks, etc.) or time from first infection (start of community spread). Finally, any conclusive study will need to address the disagreement between BCG policy and actual BCG vaccination rates with the first still contributing to the regression model, whereas the second did not.

References

1. Dockrell HM, Smith SG. What Have We Learnt about BCG Vaccination in the Last 20 Years? *Front Immunol* [Internet]. 2017 Sep 13 [cited 2020 Apr 3];8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5601272/>
2. Leentjens J, Kox M, Stokman R, Gerretsen J, Diavatopoulos DA, van Crevel R, et al. BCG Vaccination Enhances the Immunogenicity of Subsequent Influenza Vaccination in Healthy Volunteers: A Randomized, Placebo-Controlled Pilot Study. *J Infect Dis*. 2015 Dec 15;212(12):1930–8.
3. Arts RJW, Moorlag SJCFM, Novakovic B, Li Y, Wang S-Y, Oosting M, et al. BCG Vaccination Protects against Experimental Viral Infection in Humans through the Induction of Cytokines Associated with Trained Immunity. *Cell Host Microbe*. 2018 Jan;23(1):89-100.e5.
4. (2) (PDF) Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study [Internet]. ResearchGate. [cited 2020 Apr 3]. Available from: https://www.researchgate.net/publication/340263333_Correlation_between_universal_BCG_vaccination_policy_and_reduced_morbidity_and_mortality_for_COVID-19_an_epidemiological_study
5. (PDF) BCG vaccination may be protective against Covid-19 [Internet]. ResearchGate. [cited 2020 Apr 3]. Available from: https://www.researchgate.net/publication/340224580_BCG_vaccination_may_be_protective_against_Covid-19
6. Explainer: How an old tuberculosis vaccine might help fight the new coronavirus - Reuters [Internet]. [cited 2020 Apr 2]. Available from: <https://www.reuters.com/article/us-health-coronavirus-tbvaccine-explaine/explainer-how-an-old-tuberculosis-vaccine-might-help-fight-the-new-coronavirus-idUSKBN21K372>

7. Coronavirus deaths are fewer in countries that mandate TB vaccine | Fortune [Internet]. [cited 2020 Apr 2]. Available from: <https://fortune.com/2020/04/02/coronavirus-vaccine-tb-deaths/>
8. BCG vaccine: US scientists link BCG vaccination with fewer Covid-19 cases, Indian scientists hopeful but cautious - The Economic Times [Internet]. [cited 2020 Apr 2]. Available from: <https://economictimes.indiatimes.com/news/science/us-scientists-link-bcg-vaccination-with-fewer-covid-19-cases-indian-scientists-hopeful-but-cautious/articleshow/74931591.cms>
9. Rabin RC. Can an Old Vaccine Stop the New Coronavirus? The New York Times [Internet]. 2020 Apr 3 [cited 2020 Apr 6]; Available from: <https://www.nytimes.com/2020/04/03/health/coronavirus-bcg-vaccine.html>
10. Australian researchers to trial BCG vaccine for Covid-19 [Internet]. [cited 2020 Apr 2]. Available from: <https://www.clinicaltrialsarena.com/news/australia-bcg-vaccine-trial-covid-19/>
11. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*. 2020 Mar 28;395(10229):1054–62.
12. Zeng L, Li J, Liao M, Hua R, Huang P, Zhang M, et al. Risk assessment of progression to severe conditions for patients with COVID-19 pneumonia: a single-center retrospective study. *medRxiv*. 2020 Mar 30;2020.03.25.20043166.
13. Vaccination and Immunization Statistics [Internet]. UNICEF DATA. [cited 2020 Apr 4]. Available from: <https://data.unicef.org/topic/child-health/immunization/>
14. Nørgaard M, Ehrenstein V, Vandenbroucke JP. Confounding in observational studies based on large health care databases: problems and potential solutions – a primer for the clinician. *Clin Epidemiol*. 2017 Mar 28;9:185–93.
15. Community NRM. Universal BCG vaccination and protection against COVID-19: critique of an ecological study [Internet]. *Nature Research Microbiology Community*. 2020 [cited 2020 Apr 3]. Available from: <http://naturemicrobiologycommunity.nature.com/users/36050-emily-maclean/posts/64892-universal-bcg-vaccination-and-protection-against-covid-19-critique-of-an-ecological-study>
16. Yu Y, Jin H, Chen Z, Yu QL, Ma YJ, Sun XL, et al. Children’s vaccines do not induce cross reactivity against SARS-CoV. *J Clin Pathol*. 2007 Feb;60(2):208–11.

Figures and tables

Figure 1. Linear regression model exploring policy, age and income level effect on infection rates

Call:

```
lm(formula = log(CasesPerM) ~ Median + Policy + IncomeLevel,
    data = covid.stage2)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-3.3978 -0.7759 -0.0137  0.9219  2.9194
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.16279    0.48224  -2.411  0.01738 *
Median         0.07980    0.01968   4.054 8.86e-05 ***
Policy2        1.06411    0.38036   2.798 0.00598 **
Policy3        1.48048    0.58576   2.527 0.01276 *
IncomeLevel2   0.56466    0.39583   1.427 0.15625
IncomeLevel3   2.25019    0.43483   5.175 8.99e-07 ***
IncomeLevel4   3.14738    0.54213   5.806 5.12e-08 ***
```

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

Residual standard error: 1.222 on 123 degrees of freedom

Multiple R-squared: 0.7463, Adjusted R-squared: 0.734

F-statistic: 60.32 on 6 and 123 DF, p-value: < 2.2e-16

Call:

```
lm(formula = log(CasesPerM) ~ Median + BCG + IncomeLevel, data = covid.stage2)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-3.4935 -0.7091  0.0879  0.9235  2.9354
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.788849    0.594703  -1.326  0.1871
Median         0.083537    0.020396   4.096 7.54e-05 ***
BCG           -0.005930    0.003453  -1.717  0.0884 .
IncomeLevel2   0.609350    0.411892   1.479  0.1416
IncomeLevel3   2.334206    0.455878   5.120 1.13e-06 ***
IncomeLevel4   3.365031    0.551903   6.097 1.26e-08 ***
```

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

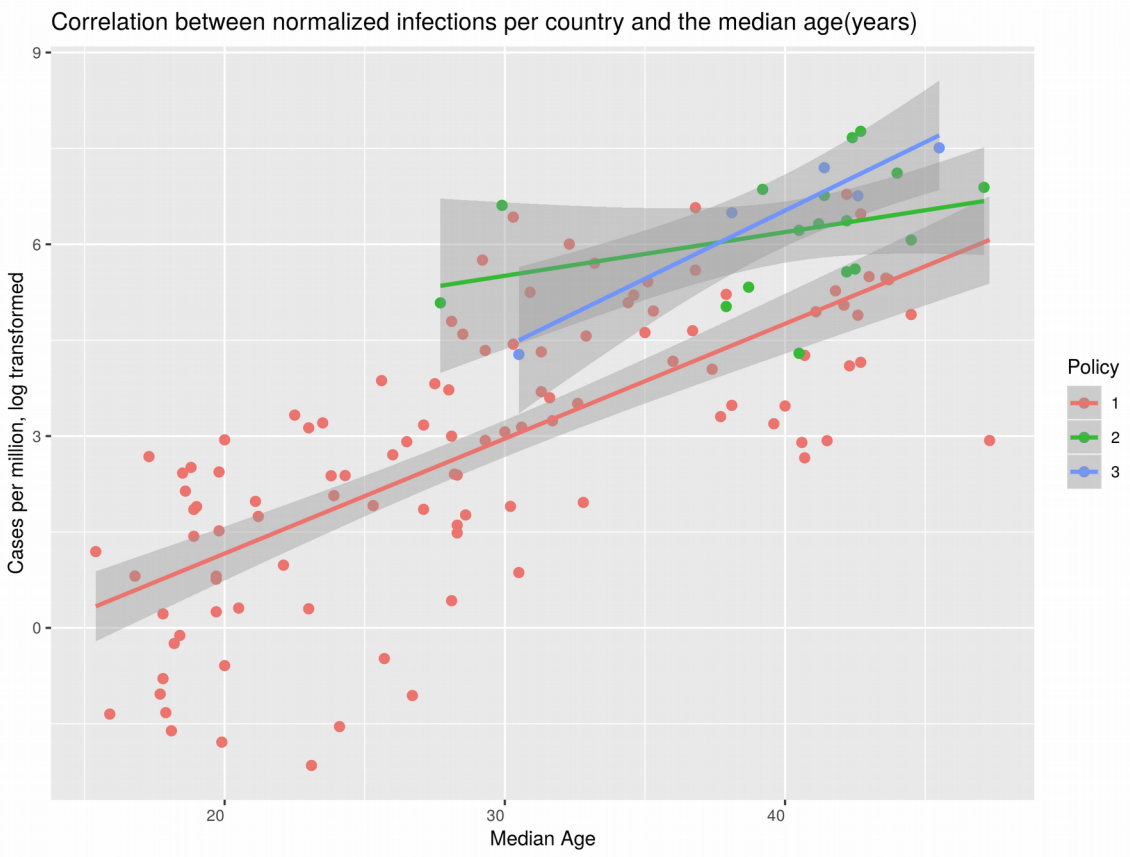
Residual standard error: 1.258 on 124 degrees of freedom

Multiple R-squared: 0.7288, Adjusted R-squared: 0.7179

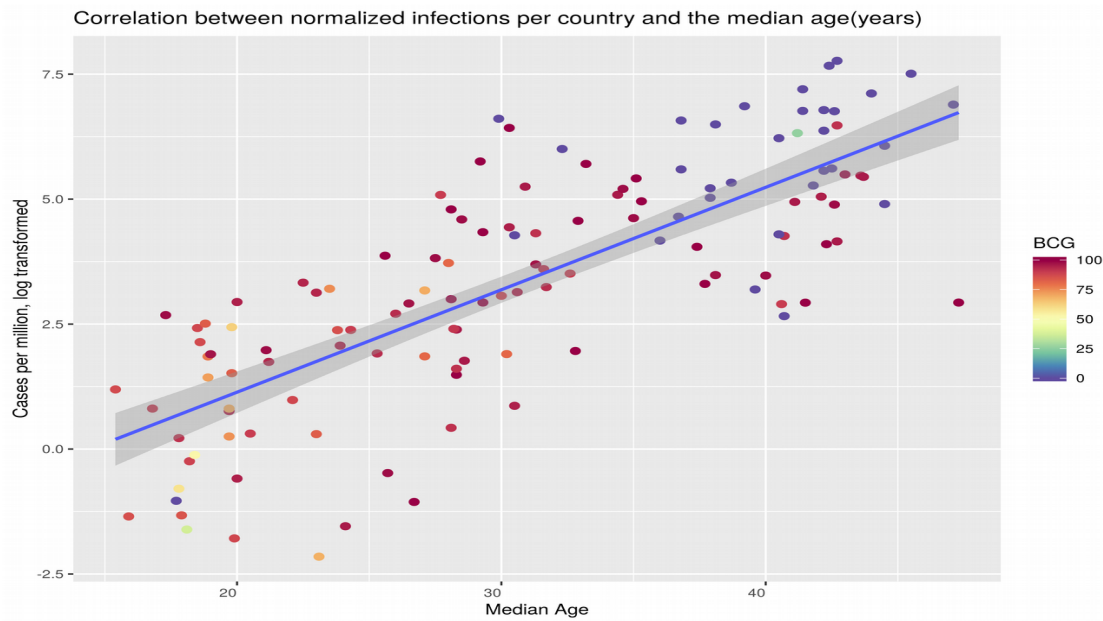
F-statistic: 66.64 on 5 and 124 DF, p-value: < 2.2e-16

Figure 2. Relationship between median age per country and infection rates in the context of BCG policy

A



B



C

Correlation between normalized infections per country and the median age(years)

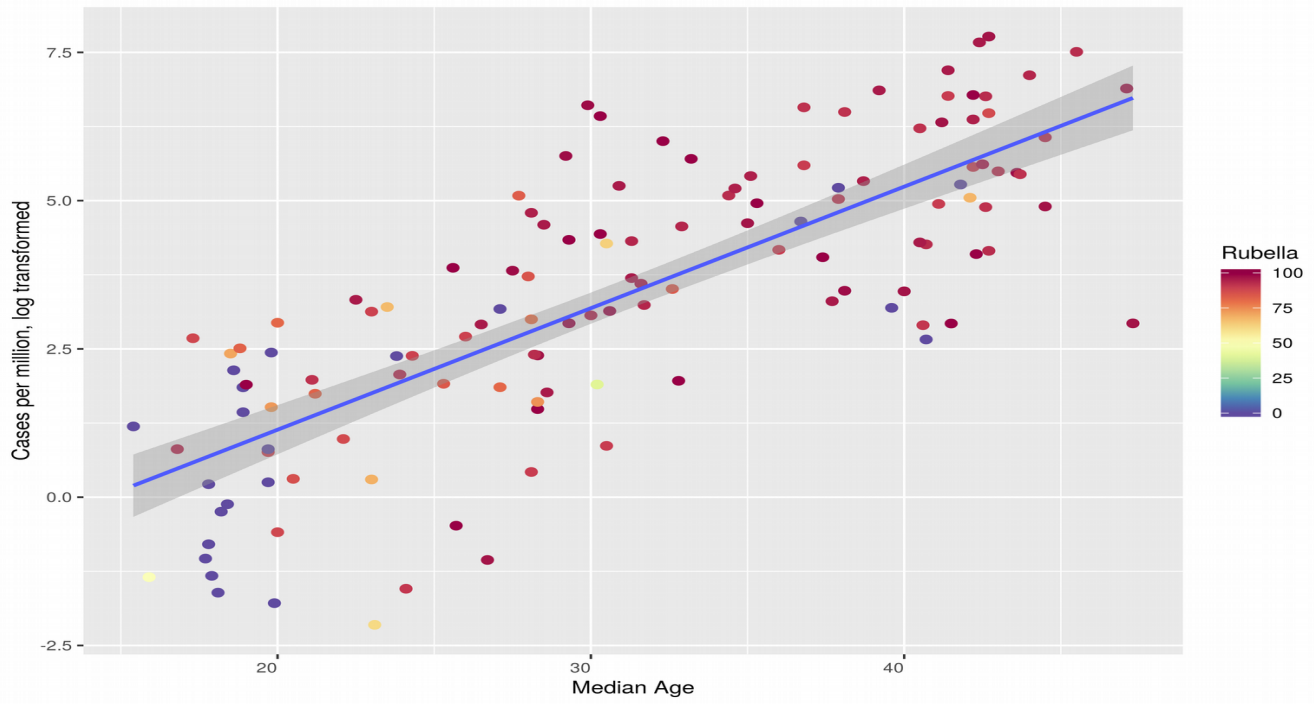


Figure 3. Effect of BCG policy or median age and infection rates in the presence and absence of income level effect

```
> anova(lm(log(CasesPerM) ~ Policy + IncomeLevel, data= covid.stage2))  
Analysis of Variance Table
```

```
Response: log(CasesPerM)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Policy	2	219.53	109.763	65.376	< 2.2e-16 ***
IncomeLevel	3	296.29	98.763	58.825	< 2.2e-16 ***
Residuals	124	208.19	1.679		

```
---
```

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

```
> anova(lm(log(CasesPerM) ~ Median + IncomeLevel, data= covid.stage2))  
Analysis of Variance Table
```

```
Response: log(CasesPerM)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Median	1	433.67	433.67	269.662	< 2.2e-16 ***
IncomeLevel	3	89.31	29.77	18.511	5.329e-10 ***
Residuals	125	201.03	1.61		

```
---
```

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

```
> anova(lm(log(CasesPerM) ~ Policy , data= covid.stage2))  
Analysis of Variance Table
```

```
Response: log(CasesPerM)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Policy	2	219.53	109.763	27.632	1.089e-10 ***
Residuals	127	504.48	3.972		

```
---
```

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

```
> anova(lm(log(CasesPerM) ~ Median, data= covid.stage2))  
Analysis of Variance Table
```

```
Response: log(CasesPerM)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Median	1	433.67	433.67	191.19	< 2.2e-16 ***
Residuals	128	290.33	2.27		

```
---
```

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

Figure 4. Mixed models used to evaluate associations between infection rates, median age, BCG policy and BCG immunization rates

Fixed effects: $\log(\text{CasesPerM}) \sim \text{Median} + \text{Policy}$

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.1470049	0.8923356	123	0.164742	0.8694
Median	0.0864087	0.0193181	123	4.472947	0.0000
Policy2	1.0936249	0.3796067	123	2.880942	0.0047
Policy3	1.5033315	0.5856559	123	2.566920	0.0115

Fixed effects: $\log(\text{CasesPerM}) \sim \text{Median} + \text{BCG}$

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.6196404	1.0028649	124	0.617870	0.5378
Median	0.0902190	0.0200332	124	4.503473	0.0000
BCG	-0.0060820	0.0034438	124	-1.766091	0.0798

Figure 5. Relationship between median age and BCG policy in the context of income levels

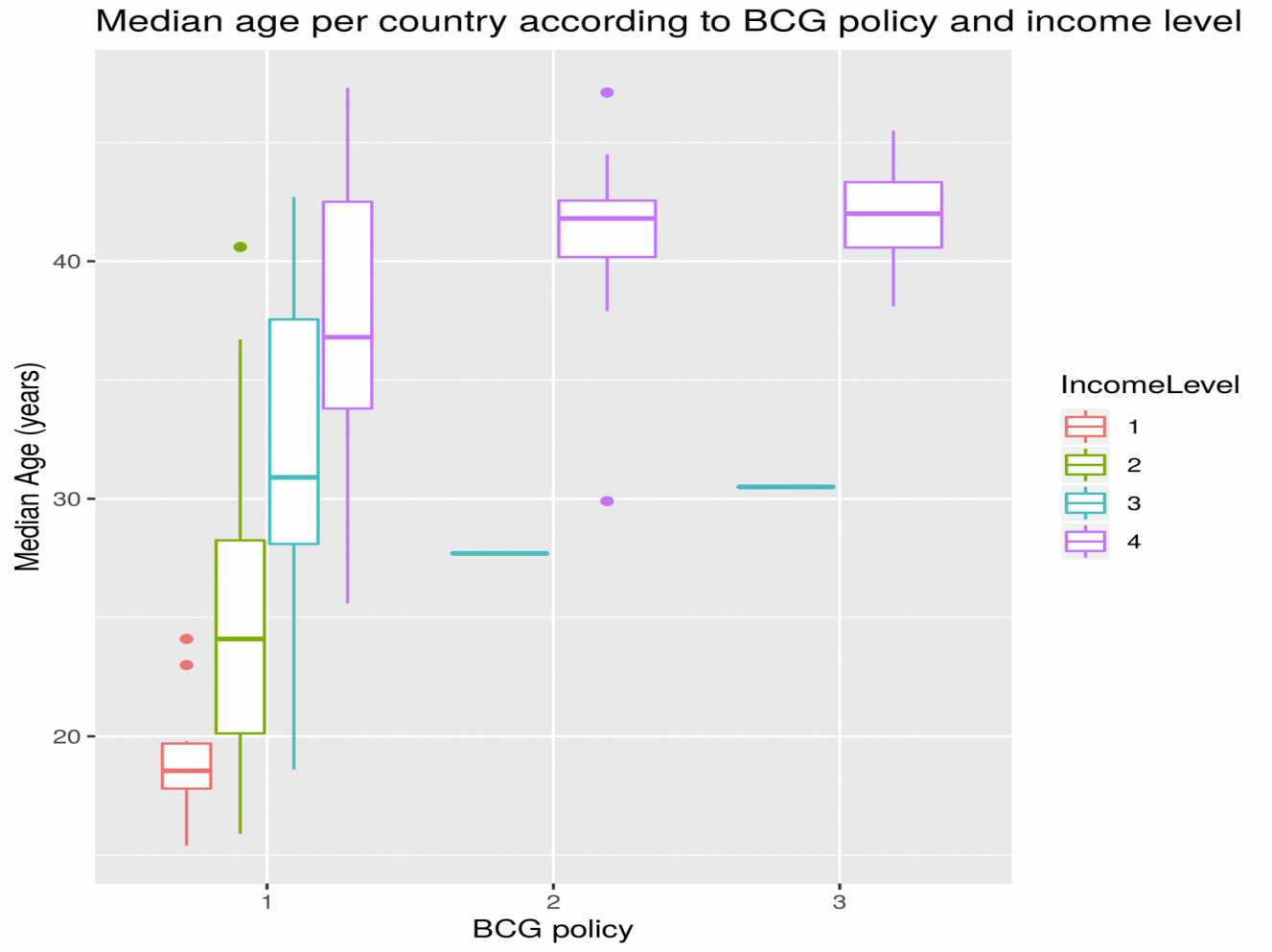


Figure 6.

Correlation between normalized deaths per country and the median age(years)

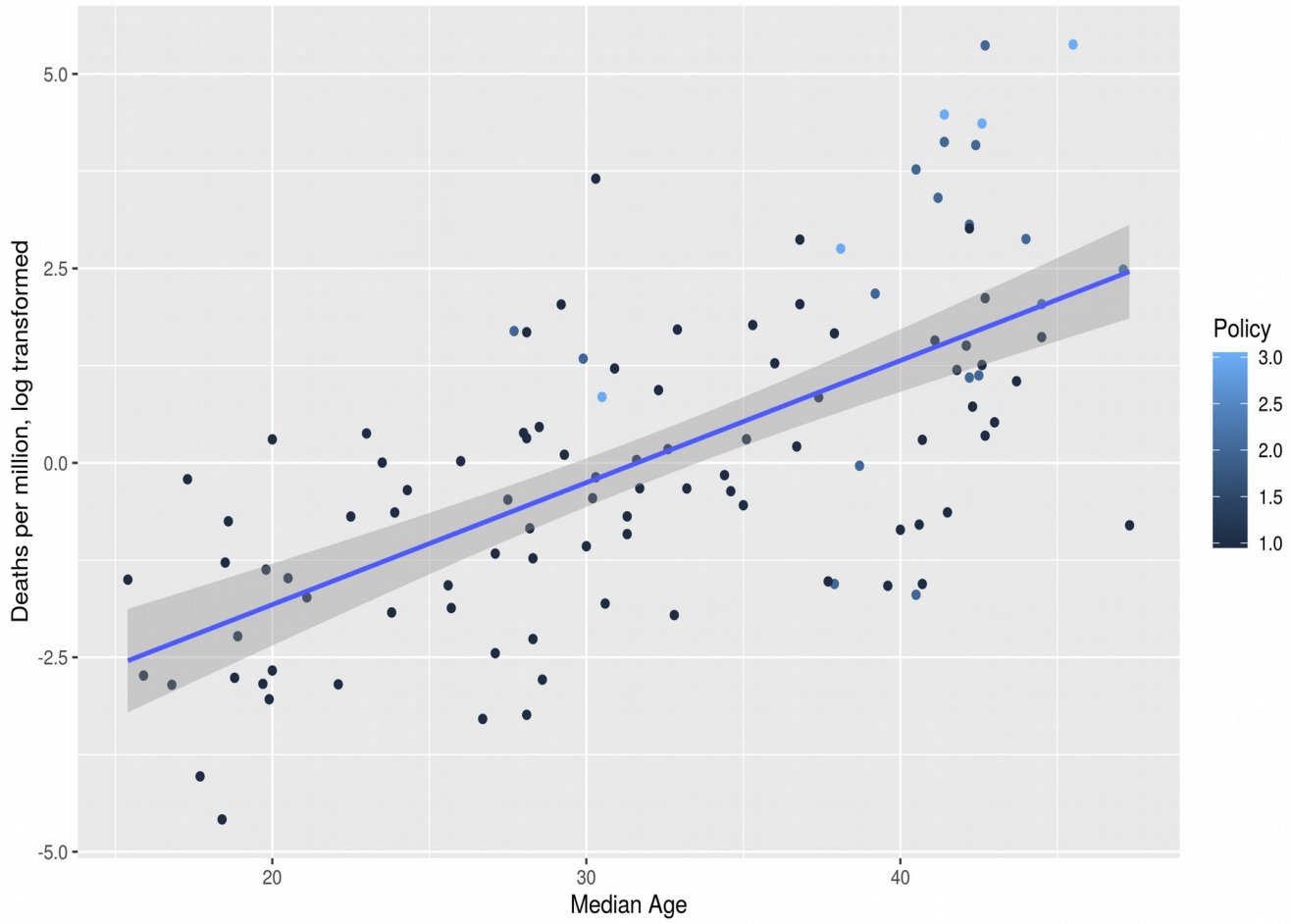
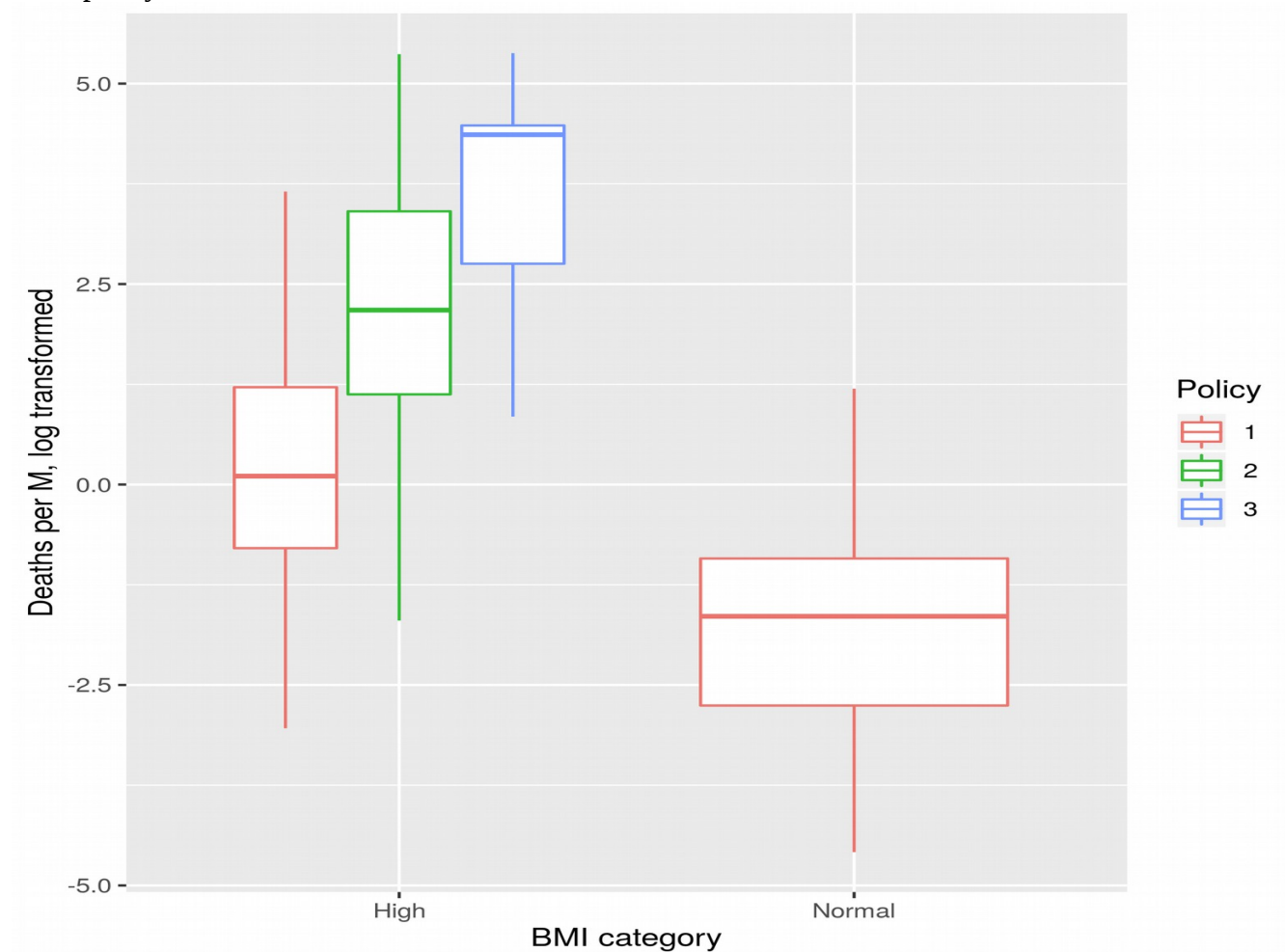


Figure 7. Association between normalized deaths from COVID19 and BMI (high>25) in the context of BCG policy



Funding: None

Note: The opinions expressed in this paper are personal and do not represent in any way Bristol Myers Squibb. No Bristol Myers Squibb resources were used to generate results or prepare this publication.

Acknowledgment: I would like to thank Max Lau, PhD for critically reading this work and Kamen Kirov for editing the text.