The BCG dilemma: linear versus non-linear correlation models over the time of the COVID-19 pandemic

Guillermo Lopez-Campos¹*, PhD; Christopher Hawthorne¹, MSc; and Miguel A. Valvano¹, MD

¹ Wellcome-Wolfson Institute for Experimental Medicine, Queen’s University Belfast, 97 Lisburn road, Belfast, United Kingdom BT9 7BL
Abstract:

In this work we have analysed the use of two different correlation metrics and two different datasets, detecting differences in the extent and directionality of the correlation between COVID-19 infection and BCG immunization. Correlation between COVID-19 infection, death rates and BCG immunization is strongly affected by the correlation metrics and so will be their biological interpretation.
Introduction:

BCG immunization has been suggested to have a protective role for COVID-19 since, due to its immunomodulatory properties, it would appear to reduce the severity of other viral infections\(^1\). This was supported by several different analytical approaches and datasets analyzing the relationship between BCG vaccination rates around the world, COVID-19 cases and death rates\(^2-4\). Previous correlation analyses highlighted a negative correlation between BCG immunization and COVID-19, providing impetus for randomized clinical trials to evaluate if the BCG vaccine can reduce the severity of COVID-19\(^5\).


The results show that in 1 year old children a negligible (or weakly positive) correlation between BCG immunization rates and COVID-19 death or case rates (15\(^{th}\) May 2020) using Spearman’s correlation. Whereas the opposite is observed using Pearson’s correlation with a statistically significant weak to moderate negative correlation. Spearman’s correlation
allows the identification of monotonic trends compared to Pearson’s linear correlation. The latter as well tends to fail in skewed distributions and is more sensitive to extreme values such as those observed in some non-vaccinated countries. Similar opposite results were observed when we used the alternative BCG immunization rates (Figure 1A).

We also analyzed fortnightly COVID-19 correlations during the pandemic and immunization rates in two different years (2000 and 2018). As before, Spearman’s correlation data show a weak positive correlation over time (Figure 1C). However, correlation increased until mid-April when many countries started implementing different protective measures, it proceeded to declined slightly and seems to have stabilized in May.

The datasets used for these analyses lack some countries that either had BCG immunization policy at later ages, did not have them or it was discontinued or non-reported (E.g. UK, USA and Australia). To address this aspect, we included an oversimplified replacement of missing BCG coverage data with a 0% coverage. This replacement introduces important changes in the results obtained, inverting the previously discussed results showing a negative correlation between death rates and COVID-19 for both correlation metrics (Figure 1B and 1D).

This change also inverts the correlation between BCG immunization and tuberculosis incidence (Table 1). The inclusion of additional global variables highlighted aspects already observed in medical clinics, such as positive correlations between the percentage of the population with a BMI greater than 25 and COVID-19 death and case rates (Table 1).
Our analyses did not include any confounding factors that affect the evolution of the outbreak (e.g. implementation of social distancing measures, demographics or other social determinants of health).

Our results expose the limitations of using different analytical approaches and partial datasets in the context of an ongoing outbreak, illustrating how data can be interpreted very differently. Apart from any biological conclusions from our analytical approach regarding the BCG immunization rate, these limitations should be considered before making definitive associations based on correlations that may change over the time of an outbreak.

Acknowledgements:
This research was supported in part by the Biotechnology and Biological Sciences Research Council (grant BB/S006281 to MAV & GLC)

References
Figure 1. Correlation between BCG immunization rates and COVID-19 death and case rates.

In blue statistically significant correlations (p-value <0.05)
Table 1. Correlations between other variables of interest.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Correlation variables</th>
<th>Correlation method</th>
<th>Correlation value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>COVID-19 case rate and incidence of BMI &gt; 25</td>
<td>Spearman</td>
<td>0.57</td>
<td>1.4E-16</td>
<td></td>
</tr>
<tr>
<td>COVID-19 death rate and incidence of BMI &gt; 25</td>
<td>Pearson</td>
<td>0.39</td>
<td>9.5E-8</td>
<td></td>
</tr>
<tr>
<td>COVID-19 case rate and life expectancy in 2019</td>
<td>Spearman</td>
<td>0.70</td>
<td>4.1E-28</td>
<td></td>
</tr>
<tr>
<td>COVID-19 death rate and life expectancy in 2019</td>
<td>Pearson</td>
<td>0.48</td>
<td>8.3E-12</td>
<td></td>
</tr>
<tr>
<td>1yo BCG immunization rates without replacement</td>
<td>Immunization rate in 2000 vs Estimate incidence of Tuberculosis (per 100k people) in 2018</td>
<td>Spearman</td>
<td>-0.32</td>
<td>5.2E-5</td>
</tr>
<tr>
<td>1yo BCG immunization rates introducing 0% for NA data</td>
<td>Pearson</td>
<td>-0.10</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>General coverage without replacement</td>
<td>Immunization rate in 2000 vs Estimate incidence of Tuberculosis (per 100k people) in 2018</td>
<td>Spearman</td>
<td>-0.41</td>
<td>4.4E-7</td>
</tr>
<tr>
<td>General coverage introducing 0% for NA data</td>
<td>Pearson</td>
<td>-0.38</td>
<td>3.2E-6</td>
<td></td>
</tr>
</tbody>
</table>