Significantly Improved COVID-19 Outcomes in Countries with Higher BCG Vaccination Coverage: A Multivariable Analysis

Danielle Klinger[1], Ido Blass[2], Nadav Rappoport[3,*] and Michal Linial[1,*]

1 Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

2 The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel.

3 Department of Software and Information Systems Engineering, Faculty of Engineering Sciences, Ben Gurion University of the Negev, Israel

* Correspondence authors: N.R (nadavrap@bgu.ac.il) ; M.L. (michall@cc.huji.ac.il)

Keywords: Epidemiology, SARS-CoV-2, Multivariable regression, tuberculosis, Demography, Chain of infection

Abstract

COVID-19 started in Hubei province in China and has spread to 210 countries within 3 months. The statistic of COVID-19 across the world is reported daily, providing a dynamic perspective for each country. We tested the hypothesis that the vaccination against tuberculosis with BCG correlates with better outcomes for COVID-19 patients. To this end, we combined the information on demography, economy, major chronic diseases, and immunization policies with the COVID-19 outcomes. We filtered out at a fixed date all countries that were below a predetermined value for population size and number of COVID-19 deaths per million (DPM). Altogether, 55 countries, covering 62.9% of the world population, met our criteria and were analyzed. To allow a reliable comparison between countries, each was aligned to a critical alignment threshold date (0.5 DPM). We found that the years of BCG admission are negatively correlated with DPM at varying times post alignment. Results from multivariable regression tests with 22 quantitative properties of each country and its population substantiate the dominant contribution of BCG admission years to the COVID-19 outcomes. Analyzing countries according to an age group partition across several time-points, reveals that the strongest correlation is attributed to the coverage in BCG vaccination of the young population (<25 years), to a lesser degree the middle age group (25-64 years), while BCG coverage status of the elderly (>65 years) was insignificant. More specifically, the signal is attributed to recent immunizations (last 15 years) rather than past policies. We propose that BCG immunization coverage, especially among the most recently vaccinated, may contribute to attenuation of the COVID-19 spread and severity.

Introduction

COVID-19 spread within 3 months to 210 countries across the globe. The progression of COVID-19 pandemic in each country is estimated by the daily report compiled by the World Health Organization (WHO) and publically available through the Worldometers online resource (*1*). The country-specific reports provide statistical information on the number of tests performed, the COVID-19 confirmed cases,

the number of deaths, and the cumulative state of patients hospitalized in serious and critical condition. Along with the spread of the pandemic, most countries imposed a policy of social distancing and other measures to mitigate COVID-19. Despite the intense effort, the knowledge on the COVID-19 SARS-CoV-2 virus is fragmented and key epidemiological parameters are still missing (*2-6*). With over 170k reported deaths and an average number of 21.7 deaths per million (DPM, as of April 20th, 2020), the death toll remains a reliable measure for monitoring the spread and progression of the disease across countries (*7*). While for some European countries such as Belgium, Spain, and Italy the DPM is >400, other infected countries (e.g., Turkey, Israel, Hungary) are closer to the world average. The large differences in COVID-19 outcomes, even among neighboring countries (e.g., Spain and Portugal) are not likely to solely reflect differences in the regulations imposed by each country at the initial phase of the epidemic (*8*).

In this study, we tested the possibility that the extent and spreading of COVID-19 cases are associated with the status of tuberculosis (TB) immunization across the world. The Bacille Calmette-Guérin vaccine (BCG) contains a live attenuated strain of *Mycobacterium bovis,* is widely used to eradicate TB and was among the most broadly used vaccination throughout the 20th century in neonatal and young children. Despite its wide use in many developing countries, about 1% of the world population is infected with TB each year. While a constant reduction in death due to TB is observed in the last decade, the annual deaths exceeds 1.5M (*9*), with a substantial overlap with people that are HIV infected (*10, 11*). Currently, the BCG vaccine is provided to the entire population in most countries with high TB incidence (*12*). Over the last two decades, numerous countries changed their policy and restricted BCG immunization policy to non-native born migrants from high TB burden countries (*13*). In this study, we questioned whether the long-range impact of the BCG vaccination regimes and the varied fraction of the population that is covered by BCG in different countries are linked with an improved COVID-19 outcome.

Materials and Methods:

Data extraction:

Information regarding COVID-19 outcomes was extracted daily between January 29th and April 20th, 2020, from the Worldometers website (*14*). Demographic measures of countries were extracted from the Worldometers website on April 17th, 2020 (*15*). Information regarding the share of population >65 years and economic development indicators was extracted from the World Bank data (*16*). Prevalence of chronic diseases (e.g. obesity, type 2 diabetes, Tuberculosis), and the death rate from cardiovascular disorders were extracted from Our World In Data website. Supplemental **Table S2** provides the source for each of these country-related information.

Information regarding past and present BCG administration practices in every country were extracted from the BCG world atlas (*15*). Two vaccination status groups were considered. The first group included countries that had either a current or past national mandatory vaccination policy (49 countries). The second group included countries that have only administered BCG vaccinations to specific groups at risk (6 countries). The percentage of the people at risk that were immunized out of the total population size is negligible (*15*). In addition, the estimates for BCG vaccination coverage between years 1980-2018 were extracted from the annual WHO reports (*17*). For additional resources used to establish the years of mandatory BCG administration see Supplemental **Table S3**.

For an extended description of data collection, data imputation and data processing see Supplemental Methods **S1-text.** Data normalization was performed for each country by accounting for the relevant population size. We included in the analysis of death per million (DPM), cases of hospitalization with

serious and critical (SPM) and cases per million (CPM). Countries were included in the selected cohort if their population is >3M, and they met the criteria of >=3 deaths per 1M population on April 17th. The countries cohort can be seen in Supplemental **Table S1**.

Data Analysis:
Accounting for the varying time of the diseases across countries, we define a unified aligned key date of a country as the first date when DPM reached for the first time the value 0.5 or higher. The following analysis was conducted across changing dates (e.g. 10-30 days) for the key aligned date as defined. For country-specific groups (binary or categorized, such as BCG administration), we applied the ranked Wilcoxon test as a nonparametric statistical test for comparing any paired groups. For the continuous (i.e. the number of years BCG was administered, the prevalence of TB in the population, median age) data we applied linear regression, reporting the regression fit and the calculated statistical significance (p-value) for the COVID-19 outcomes. A correlation between the BCG by age groups was determined by partitioning the population of each country into three groups: (i) 24 years and younger; (ii) 25 to 64 years; (iii) 65 years and older. From the age partition and the BCG coverage within each age group, a value that measures the *percentage of the population in the share of the age group with BCG* is calculated (see Supplemental Materials and Methods, **S1-text**).

Data Analysis Availability:
An online tool for displaying the analytical results is available at: https://covi.shinyapps.io/COVID19/.
Code and data is available at: https://github.com/nadavrap/COVID19.

Results

Unified COVID-19 pandemic across countries by alignment key dates

Analyzing the trends of the COVID-19 pandemic across different countries is prone to variability on the actual date in which countries were exposed and where COVID-19 fatality is recognized and reported. To increase the robustness of our analyses, we focused on countries with a minimal size of >3M population and mandatory 3 DPM at the analysis date (17 April, 2020). Altogether, 55 out of 210 countries covering 62.9% of the world population complied with these thresholds. A regional partition of these countries is shown in **Fig 1.** For detailed information on the countries included in the analyses, see Supplemental **Table S1**.

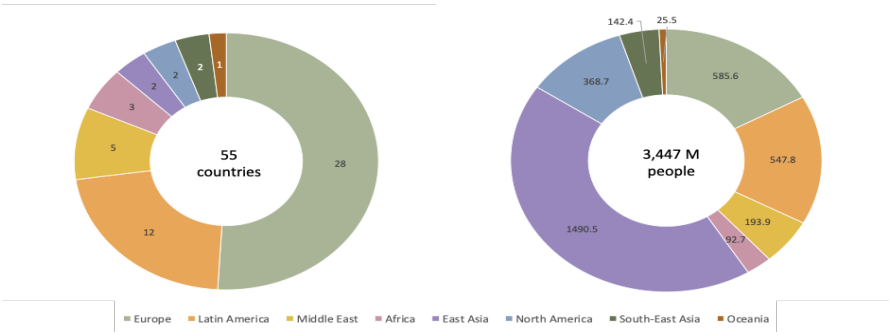

Fig. 1. Countries analyzed in this study (55 countries), partitioned by their geographical regions (left) and the cumulative population size within each region (right).

BCG Administration Years Are Negatively Correlated with COVID-19 Outcomes

First we analyze COVID-19 outcomes as the difference in the death per millions (DPM) or cases per million (CPM). **Fig. 2** shows strong and significant correlations between COVID-19 outcomes and years of BCG administration. The analysis is performed 15 days following 2 different alignment key dates (defined by DPM = 0.5 and DPM = 1.5). Among the 51 countries with data information on the 15th day, we observed a strong negative correlation with both DPM and CPM as a function of BCG admission years. For DPM outcome we observed a correlation of R = -0.46, p-value of 0.0018 when aligned at DPM at >= 0.5 (**Fig. 2a**) and R = -0.49, p-value of 0.0019 when aligned at DPM >= 1.5 (**Fig. 2b**). Similarly, considering the CPM rather than the DMP results is a similar trend with slightly lower correlation and statistical strength. Specifically, at 15 days post alignment, we observed a correlation of R = -0.31, p-value of 0.039 when aligned at DPM >= 0.5 (**Fig. 2c**) and R = -0.39, p-value of 0.016 when aligned at DPM >= 1.5 (**Fig. 2d**). Repeating the analysis at various time points along the progression of the disease (e.g., 20 days after the aligned key date) resulted in a similar trend of highly significant negative correlation between COVID-19 outcomes and years of BCG administration. Note that at 20 days post key aligned data, as many as 44 out of 55 of the countries provided information. The other countries (mainly from Latin America) were at an earlier phase of the pandemic.

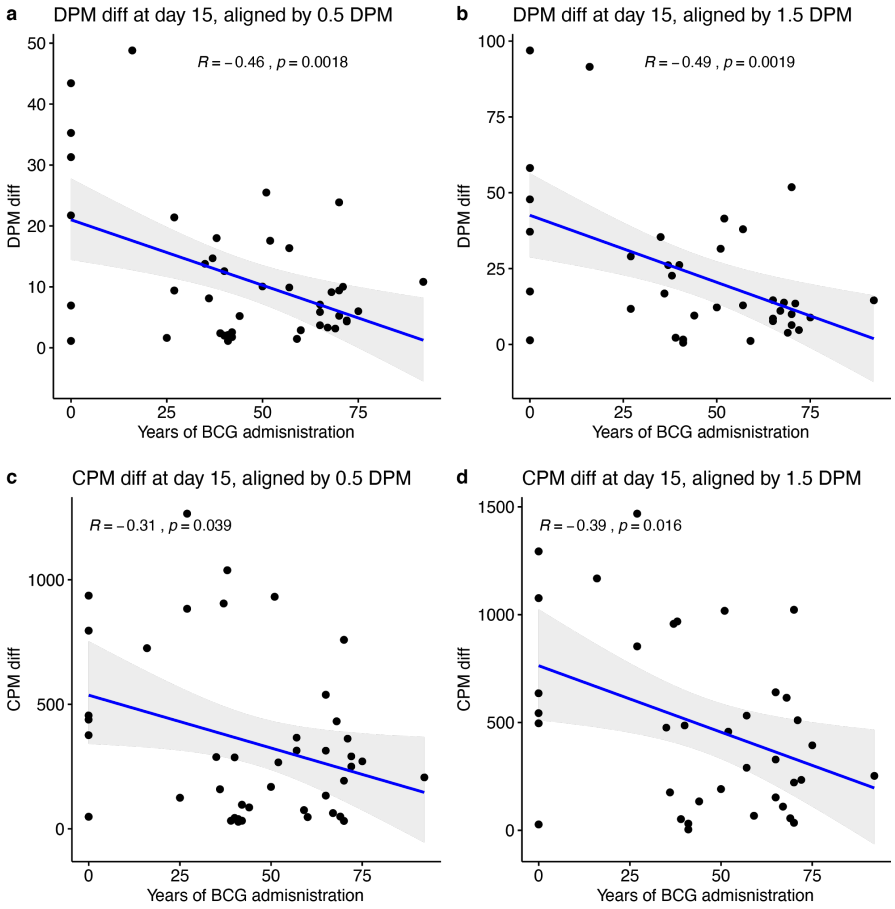

Fig. 2. Statistical analysis of the countries COVID-19 outcomes and years of BCG administration. All correlations were measured at 15 days following the alignment key date (51 countries). Correlations of years of BCG administration with DPM = 0.5 (a) and DPM = 1.5 (b). Correlation with CPM diff at 15 days

when key date was defined as CPM = 0.5 (c) and CPM = 1.5 (d). DPM diff and CPM diff are calculated by the differences in the numbers from the measured date to alignment date.

To test the generality of our observations we repeated the analysis at a broad range of time points along the progress of the disease in the different countries, starting from the 10th day post the alignment key date and showing the trends in 5 days intervals (**Fig. 3)**. For this analysis we tested the outcome of COVID-19 confirmed cases (CPM) and the number of recovered (RPM), in addition to the DPM and SPM. The results of the DPM and SPM show high significance association at all time points, excluding the last time point that presents limited statistical significance based on data from only 19 out of 55 countries.

Fig. 3. Outcomes correlations with years of BCG administered. Each row represents a different outcome, and each column represents the time interval of the outcome from alignment date. Circle size represents the number of countries, and colour represents statistical significance with darker color indicates higher significance. Values in circles are the correlation estimates. Note that all correlations are in negative values.

Multivariable analysis reveals the strong contribution of BCG administration to the statistical signal of COVID-19 outcome

Countries differ in many aspects like population size, Gross Domestic Product (GDP), life span, and more. In order to control for some of the potential confounding factors, we included numerous demographic values for a multivariable linear regression. The results show that the number of BCG administration years ranks significantly within the 3 coefficients with the larger effect (as measured by the normalized beta coefficient, and out of 22 coefficients) **(Fig 4**). The results are consistent among the different times observed: 15 days- ranks 3rd largest, p-value = 0.0355, beta = -0.67; 20 days- ranks 6th larger effect, p-value = 0.0123, beta = -0.75; 25 days- ranks 1st largest effect, p-value = 0.0235, beta = -1.18. Unsurprisingly, a strong positive beta coefficient is associated with the median age. Cancer percentage is

also significant. The contribution of comorbidities in the multivariable analysis call for future analysis.

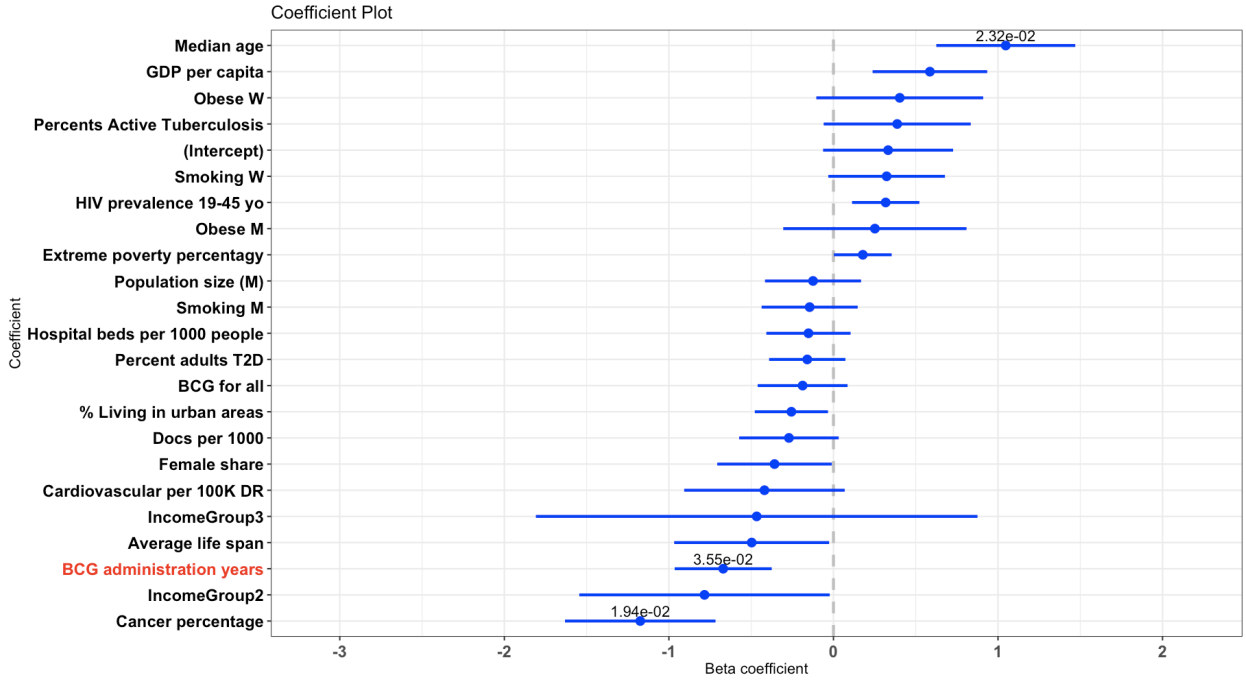

Fig. 4. Multivariable analysis. Beta coefficients of the normalized multivariable linear regression are shown. Blue lines represent the coefficients' 95% confidence intervals. P-values were added for all variables with p-value < 0.1.

Highest correlation with BCG age coverage applies to the most recently vaccinated

We next investigated the relevance of age groups to the observation showing that years of BCG admission are strongly correlated with better COVID-19 outcomes (**Figs 2,3**). We inspected the age partition of the population for each country, as it is evident that countries differ drastically in their median age. For instance, while in Ecuador the median age is 27.9, it is 47.3 years in Italy. **Fig. 5** shows the correlation of total years of BCG admission within the given timeframe of the age group, according to the country-based age composition, 15 days post alignment key dates. The population in each country was partitioned to young (<24 years of age), working class (25-64 years) and old (>65 years). Across several time points, the strongest correlation for BCG coverage for the age dependent was associated with the young group (**Fig. 5a**), followed by the working class group (**Fig 5b**). The correlation with the young age group (testing at 20 days) is highly significant with R = -0.53, p-value 0.00048. The highly statistically significant negative trend repeats among all dates tested within the 10-25 days post alignment. The correlation with the age group of 25-65 years (testing at 20 days) is also significant with R = -0.32 with a weaker significance (p-value = 0.044). The negative trend in 15 days is observed but it is not significant (e.g. at 15 days- R =- 0.28, p-value = 0.064). Interestingly, for the old age group the correlation is negligible and insignificant (**Fig. 5c**)

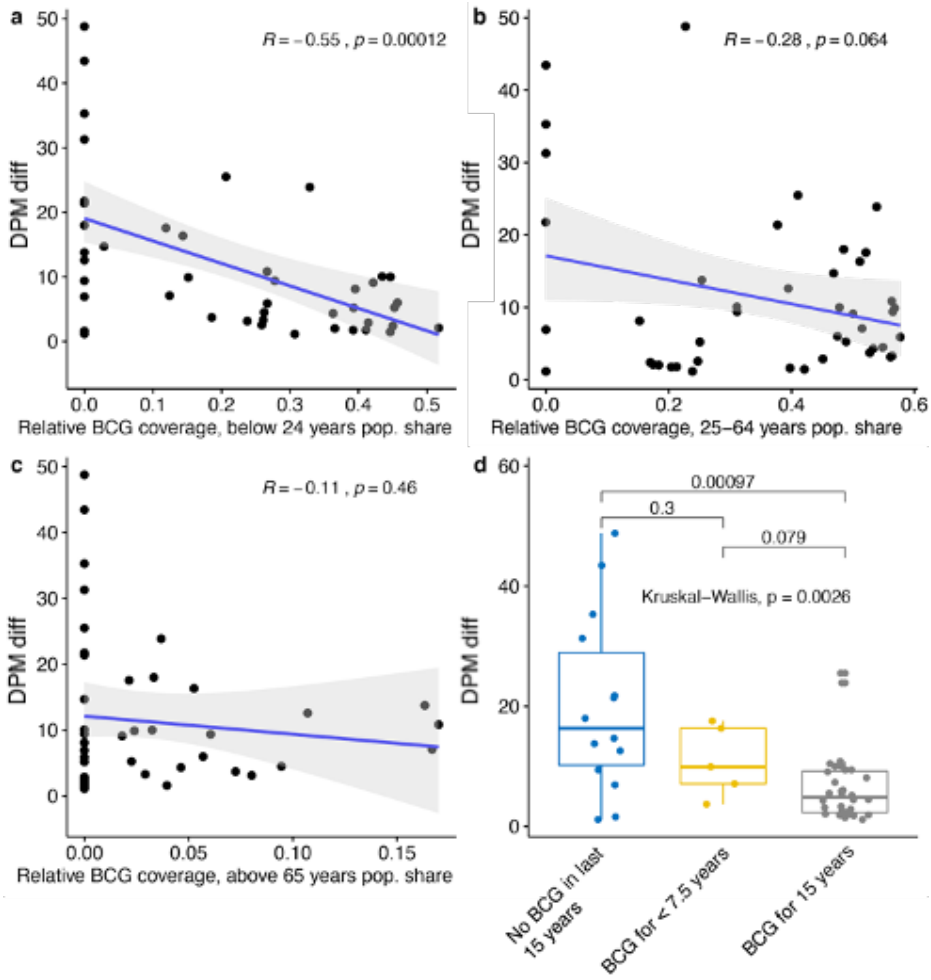

Fig. 5. Correlations of relative BCG coverage with DPM diff among varying age groups. All correlations were measured at 15 days following the alignment key date. Relative BCG coverage within the young age group (**a**); Relative BCG coverage within the working class age group (**b**); Relative BCG coverage within the old age group (**c**). The DPM diff in countries that have (gray), partially have (yellow), or have not (blue) rolled immunization programs over the past 15 years. The statistical significance values are shown (**d**).

The pronounced signal in the young age group led us to investigate whether recent immunization may have a positive effect on the outcome, and whether the signal may be due to the proximity in age group of the young to the time of vaccination. The efficacy of BCG against TB is expected to span approximately 12-15 years (*18*). We therefore divided the countries into 3 disjoined groups representing their vaccination policies over the past 15 years: (i) countries with mandatory immunization policies over the past 15 years; (ii) countries with mandatory immunization policies for less than 7.5 years within the past 15 years; (iii) countries with no mandatory immunization policies over the past 15 years. Applying the Kruskal-Wallis test to the outcome of DPM, yielded highly significant results at a range of tested dates post alignment (10-25 days), establishing that countries with immunization policies over the past 15 years have a significantly lower rate of DPM with respect with countries in group (iii). **Fig. 5d** shows results for 15 days post alignment, with a p-value of 0.00097 between the fully treated (group i) and non treated (group iii) groups.

We provide an interactive website in https://covi.shinyapps.io/COVID19/ that provides a useful analytical webtool for single variant statistics, correlations, multivariable analyses and more. The user friendly platform allows users to change parameters along the pandemic progression by selecting any predetermined outcomes as a reference date for the alignment.

Discussion

The significant strong correlation between the BCG vaccination and better outcomes for COVID-19 is shown across many countries, covering the majority of the world population (62.9%). This finding is based on an unbiased view of all countries that comply with predetermined thresholds on a minimum number of DPM and population size (see Materials and Methods). We anticipate that with the progression of COVID-19 pandemic more countries that are at an early phase of the pandemic could be included. Importantly, the strong negative correlation between the BCG admission years and DPM was sustained at a range of time-points (**Fig. 3**). The trend was very similar for DPM or for accounting people that were hospitalized and are in serious and critical condition (SPM). The negative correlation to COVID-19 positive cases (CPM) is weaker relative to DPM. It reflects the varying capacities of different countries to carry the molecular tests (or lung CT pathology as a test) and more importantly the policy of who is being tested greatly varies. Weaker confidence was observed for people that were recovered (RPM). We assume it reflects the early stage of the pandemic in many countries. Furthermore, the non-standardized definition for COVID-19 confirmed recovery is likely to reduce the reliability of the RPM and eventually affect statistical power.

Our multivariable analysis highlights the strength of combining a broad range of country-based quantitative observation (e.g., economic measures, number of doctors per 1000 people), and population composition including habits (smoking), exposure to infectious diseases (prevalence of TB) and several major comorbidities (e.g. cancer, diabetes). Many of these individual measures are correlated and may reflect confounding factors. For instance, countries that suffer from a high level of TB (*19*) are also marked by high immigration, social inequalities, high drug or alcohol abuse, and HIV infection (*11*). Most importantly, the multivariable analysis validated the importance and statistical significance of BCG immunization years given the other variables, attributing better outcomes to them. As expected the median age was also extremely significant, with higher values strongly contributing to the worse COVID-19 outcome.

The detailed information on the number of BCG admission years is used as an additional source of information that is unavailable when a binary definition of BCG immunization is considered. Detailed information on the exact years from first admission, the health policy in each country, and the actual immunization coverage (provided by the WHO) allowed a thorough inspection of the BCG vaccination by the population structure (**Fig. 5**). We explicitly tested what component of the BCG immunization is most critical out of the COVID-19 outcomes. By partitioning the population to young, middle and elderly age groups, we confirmed that the strongest signal towards COVID-19 outcome is associated with the young (<24 years), then the middle age (25-64 years) while the elderly (>65 years) that are at highest risk for COVID-19 mortality do not correlate with BCG higher coverage. The implication of this observation for COVID-19 epidemiology is evident. It is likely that in countries where the young population (<24 years) is vaccinated maximal protection is provided. Universally, the DPM among young people is very low (0.04% for <17 years (*14*)), their main contribution is expressed in sustaining the chain of infection. Higher coverage of young that leads to a lower death toll in the elderly population is most probably due to a lower rate of infection that is caused by the young group, leading to an overall lower death toll. The lack of correlation between BCG coverage of >65 years and COVID-19 outcomes is in accord with the expected negligible impact of the elderly population on the transmission within the community.

Several reports proposed that the BCG vaccinated population is partially protected against viruses in general, and the SARS-CoV-2 in particular (*20*). Despite the broad usage of BCG for almost a century, the long-term effect of BCG on the immune system remains enigmatic. BCG was reported to have a protective function on various pathogens and was argued to act as a general systemic immunological stimulator. The indirect effects of BCG on diabetes, multiple sclerosis, and all-cause mortality of children in developing countries were reported (*21*). We argue based on the age-partition analysis that the recent immunization in many of these countries rather than a long-term heterologous effect dominates. The strong statistical significance value for BCG being effective for the recently immunized population (0-15 years, **Fig 4d**) supports the hypothesis for a better outcome of COVID-19 is governs by recent BCG immunization (*22*), the alternative view suggesting a lifetime activation of the system that awaits additional evidence.
We conclude that the strong negative correlation with BCG administration years, the impact of a recent vaccination, and the validated role of the young population in the spread of COVID-19 are fundamental observations for revisiting global and national BCG immunization policy. While, at present, the WHO does not recommend BCG vaccination for prevention of COVID-19, several clinical trials with BCG are undertaken (*23, 24*).

Acknowledgement

We would like to thank the Interdisciplinary Data Science Research (CIRD), the Hebrew University of Jerusalem, for a partial fellowship to D.K.

References

1. World Health Organization, Coronavirus disease 2019 (COVID-19): situation report **74**, (2020).
2. S.-m. Jung *et al.*, Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. *Journal of clinical medicine* **9**, 523 (2020).
3. Y. Bai *et al.*, Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA*, (2020).
4. Q. Ruan, K. Yang, W. Wang, L. Jiang, J. Song, Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive care medicine*, 1-3 (2020).
5. Y. Wang, Y. Wang, Y. Chen, Q. Qin, Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *Journal of medical virology*, (2020).
6. R. M. Anderson, H. Heesterbeek, D. Klinkenberg, T. D. Hollingsworth, How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* **395**, 931-934 (2020).
7. N. Subbaraman, Why daily death tolls have become unusually important in understanding the coronavirus pandemic. *Nature*, (2020).
8. G. Onder, G. Rezza, S. Brusaferro, Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* (2020).
9. G. Sotgiu, G. Sulis, A. Matteelli, Tuberculosis - A World Health Organization Perspective. *Tuberculosis and Nontuberculous Mycobacterial Infections*, 211-228 (2017).

10. N. R. Gandhi *et al.*, Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *The Lancet* **368**, 1575-1580 (2006).
11. J.-P. Millet *et al.*, Factors that influence current tuberculosis epidemiology. *European Spine Journal* **22**, 539-548 (2013).
12. P. Glaziou, C. Sismanidis, K. Floyd, M. Raviglione, Global epidemiology of tuberculosis. *Cold Spring Harbor perspectives in medicine* **5**, a017798 (2015).
13. M. Pareek, C. Greenaway, T. Noori, J. Munoz, D. Zenner, The impact of migration on tuberculosis epidemiology and control in high-income countries: a review. *BMC medicine* **14**, 48 (2016).
14. WorldMeters, https://www.worldometers.info/coronavirus/ *Online Resource* (2020).
15. A. Zwerling *et al.*, The BCG World Atlas: a database of global BCG vaccination policies and practices. *PLoS Med* **8**, e1001012 (2011).
16. World Bank, https://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS, *Online Resource* (2019).
17. World Health Organization, Reported estimates of BCG coverage. *Online Resource* (2019).
18. P. D. Hart, I. Sutherland, BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. *Br Med J* **2**, 293-295 (1977).
19. J. A. Seddon, D. Shingadia, Epidemiology and disease burden of tuberculosis in children: a global perspective. *Infection and drug resistance* **7**, 153 (2014).
20. M. Gursel, I. Gursel, Is Global BCG Vaccination Coverage Relevant To The Progression Of SARS-CoV-2 Pandemic? *Medical Hypotheses*, 109707 (2020).
21. B. R. Bloom, in *The Value of BCG and TNF in Autoimmunity*. (Elsevier, 2018), pp. 1-10.
22. S. Salman, M. L. Salem, Routine childhood immunization may protect against COVID-19. *Med Hypotheses* **140**, 109689 (2020).
23. J. de Vriese, Can a century-old TB vaccine steel the immune system against the new coronavirus. *Science* (2020).
24. B. M. Ayoub, COVID-19 vaccination clinical trials should consider multiple doses of BCG. *Die Pharmazie* **75**, 159 (2020).
25. WorldMeters, https://https://www.worldometers.info/population/, Online Resource *(2018).*
26. World Bank indicators, *https://databank.worldbank.org/source/world-development-indicators/*. *Online Resource (2019).*
27. United Nations, Population Division, https://ourworldindata.org/grapher/population-by-broad-age-group, *Online Resource* (2017).