Covid-19 vs. BCG Universal Immunization: Statistical Significance in the Early Phase of the Pandemics

Serge Dolgikh [1*]

[1] *National Aviation University, 1 Lubomira Huzara Ave, Kyiv 03058, Ukraine.*

Abstract

A possible correlation between the impact of Covid-19 and universal immunization program against tuberculosis with Bacillus Calmette-Guérin (BCG) vaccine was suggested previously, based on apparent correlation between lower impact of the epidemics and a record of national BCG immunization program. In this work a time-adjusted dataset of Covid-19 statistical data by national and subnational health jurisdictions at the time point of six months after the local arrival of the epidemics was used to perform a statistical analysis of the significance of the correlation hypothesis between universal BCG immunization and a milder early phase Covid-19 scenario. With the data accumulated up to the point of the analysis, the significance of the correlation hypothesis was evaluated both qualitatively and quantitatively with the conclusion that it has achieved statistically significant level of confidence. The conclusions of this research can be used in development of epidemiological policy as well as the rationale to investigate the origin and mechanisms of a broad immunity protection that can be associated with an early-age exposure to BCG vaccine.

Keywords: Infectious Diseases; Epidemiology; Statistical Analysis; Covid-19.

1. Introduction

As the novel coronavirus Covid-19 epidemics has emerged and penetrated multiple national jurisdictions, a possible correlation between the impact of Covid-19 and universal immunization program against tuberculosis with Bacillus Calmette-Guérin (BCG) vaccine was suggested by Miller et al. (2020) [1], based on apparent correlation between lower rate and impact of the epidemics and a record of national BCG immunization program (UBIP). Due to significant variation in multiple potential factors of influence between jurisdictions and stochastic and unpredictable character of development of the Covid-19 epidemics, a confident conclusion about such correlation can be far from trivial and for this reason was investigated in a large and growing number of studies [2-5].

Escobar et al. (2020) [2], a strong correlation between the BCG immunization record and Covid-19 mortality in a number of culturally and socially similar European countries was observed ($R^2 = 0.88$; $P = 8 \times 10^{-7}$), indicating that every 10% increase in the BCG index was associated with a 10.4% reduction in Covid-19 mortality. The results imposed strong constraints on the null hypothesis (that is, of no correlation between a current or previous UBIP in the jurisdiction and the recorded Covid-19 impact), suggesting that BCG may have a certain broad protective effect resulting in a milder epidemiological scenario. A similar conclusion is supported by the results in Sharma et al. (2020) and Yitbarek et al. (2020) [3, 4] establishing a strong correlation between a record of UBIP, current or previous, and lower factors of Covid-19 impact in the jurisdiction measured by infection incidence and the resulting mortality "*The*

---

results ... show that countries without a universal BCG policy (such as Belgium, Italy, the United States, and the Netherlands) have increased incidence of COVID-19 (2810.9 ± 497.1 (mean ± SEM) per million) *compared with countries with ongoing national BCG policy* (570.9 ± 155.6 (mean ± SEM) per million)" [4].

Dolgikh (2020) [5] studied qualitative and quantitative analysis of distribution of Covid-19 impacts among national and subnational jurisdictions in Europe, North America and Middle East was performed with a number of observations consistently pointing to a possibility of a correlation between UBIP and a milder type of the epidemiological scenario in the initial phase of the Covid-19 epidemics. These findings were consistent with a number of further results in support of the correlation hypothesis [6-8].

On the other hand, in a number of studies arguments against correlation were presented [9, 10] indicating complex character of the relationship with multiple influencing factors that make confident determination of the correlation challenging.

Given the complexity and importance of the problem, the intent of this work was to analyze publicly available Covid-19 epidemiological data reported by national and subnational public health jurisdictions with respect to the hypothesized induced immunity population-scale protection resulting from a universal BCG vaccination policy, current or previous, and attempt both qualitative and quantitative analysis of the hypothesis of a correlation between a current or previous UBIP in the jurisdiction and a milder scenario of Covid-19 epidemics; to verify the assumptions, the results and conclusions of the earlier studies with a specific objective to determine, in a quantitative analysis, the constraints and confidence of the correlation and null hypotheses. To this end, a dataset of cases with different time of local arrival of the epidemics was compiled and analyzed with statistical methods to evaluate statistical significance of the UBIP – Covid-19 correlation hypothesis. Whereas an earlier work of the authors [5] was intended to point out a number of observations consistent with the correlation hypothesis, in this work these observations are verified at a different time point, indicating stability of the established connection; and a quantitative statistical significance analysis of the correlation and null hypotheses performed with the compiled dataset of epidemiological cases.

It needs to be noted and emphasized that the analysis in this study applied to the early, initial phase of development of Covid-19 pandemics and may not be assumed to be applicable unconditionally to the subsequent phases, due to a number of factors, including mutability of the infection agent, difference in the standards of quality of the public health care system, availability of resources, epidemiological management and control policies including vaccination and other factors.

1.1. Terminology

Timing considerations can be critical in the analysis of the development of an epidemiological scenario. For this reason, an effort was made to ensure that the data was recorded at a similar phase in the development of the epidemics in the reporting jurisdiction. To emphasize the need for synchronization of epidemiological data, the zero time of the start of the global Covid-19 pandemics was defined in Dolgikh (2020) study [5] as *31.12.2020*; along with the global Time Zero (TZ) we also define local Time Zero point (LTZ) indicating the time of arrival of the epidemics in the given locality. It can be sensibly defined as the date of the first confirmed case in the local area, subnational or national jurisdiction.

The impact of the epidemics was measured by Covid-19 caused mortality per 1 Million capita in a reporting jurisdiction as a function of time:

$$m(case, t) = \frac{Mortality\ (jurisdiction,\ time\ point)}{Population\ (Million)} \qquad (1)$$

It was judged that especially in the initial phase of the epidemics this parameter could be a more current and accurate measure of the epidemiological impact than the number of cases that strongly depends on the testing and other practices that may not be consistent between jurisdictions, on the assumption that policies and protocols in the selected public health administration allowed reasonably accurate identification of cause and accurate and timely reporting.

Evidently, as defined, the impact of the epidemics in a jurisdiction, *m(c, t)* would be a function of the jurisdiction factors *F*, including demographics, geographical distribution, prosperity, social customs and traditions, lifestyle, public health administration and epidemiological policy, including records of universal immunization programs and not in the least, the time. As the experience of the earlier period shows, the considerations of timing can be of critical importance for the correctness of the conclusions of the analysis. For that reason, an attention was given to comparing impacts relative to the time of the first exposure by using time-adjusted data at the same local time point in the development of the epidemics.

Throughout this work two related measures of the epidemiological impact were used as well: relative impact $m_R(t)$ measured as a ratio of the impact recorded in the locality to the world's highest value (at the time of evaluation); and the impact expressed in a logarithmic scale $m_L(t)$:

$$m_R(c,t) = \frac{m(c,\ t)}{max(m)}; \ \ m_L(c,t) = \log\big(m(c,t)\big) \tag{2}$$

At the time point of the analysis in this work, that is, approximately six months after of local development, it is expected that the epidemics had developed, and the relationships of factors of influence would have a more stable nature than in the very early phase that in many observed cases was subject to random factors and fluctuations; whereas in the later phases differences between jurisdictions in a number of essential factors, including not in the least, epidemiological policy and availability of resources, could become more pronounced.

The rest of this paper is organized as follows. In Section 2, the methods of the research and the dataset used in the study are described. Section 3 contains the results of qualitative and quantitative analysis of the correlation hypothesis. Section 4 contains detailed definition of the method and the results of the statistical significance analysis. In Section 5 a discussion of the findings is provided. Finally, in the concluding Section 6 the results of the study are summarized and a brief discussion of possible applications offered.

2. Research Methodology

Qualitative and quantitative methods such as: comparison of cases with similar socio-economic factors; trend analysis with respect to duration of UBIP; and evaluation of statistical significance based on observed values of sample means in the identified groups of cases were applied to analyze trends in development of the epidemiological situation in the selected jurisdictions with the intent to evaluate statistical significance of the correlation hypothesis between the impact of Covid-19 epidemics and a record of universal BCG immunization and impose quantitative constraints on the null hypothesis.
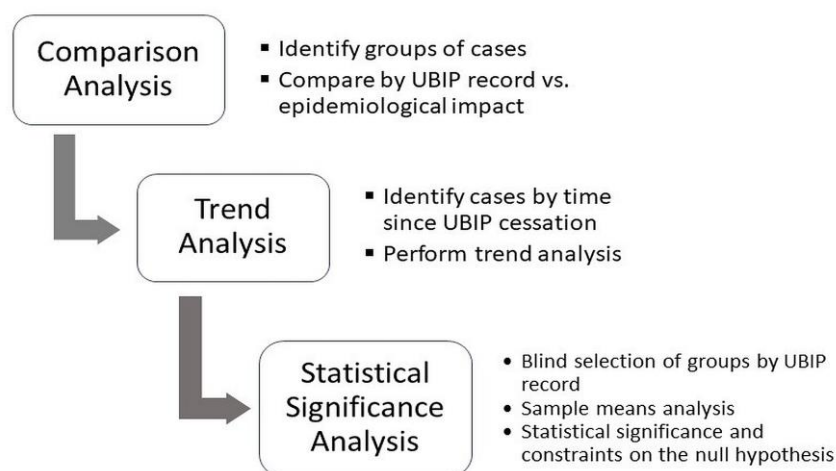

Figure 1. Research methodology

2.1. UBIP Correlation Hypothesis

The hypothesis that will be evaluated in the study is that of an induced broad protection effect against viral infections including Covid-19 at a population-scale level (importantly to note, not necessarily a complete individual protection as in the case of regular vaccinations; and with further qualifications e.g. the immunization program was implemented consistently, with adequate quality and without significant interruptions). The hypothesis was proposed in Dolgikh (2020) [5] based on the analysis of early statistical data and the earlier results of immunology studies suggesting a possibility of such link in a number of human health conditions [11-13]. To the best of our knowledge, the exact mechanism of such a population-wide protection in the immune system has yet to be determined and will be addressed elsewhere; however, a logical possibility that such an induced general immunity protection may have an influence on the epidemiological scenario via providing some level of population-wide mitigation of negative Covid-19 impacts in the initial phases of the epidemics can be tested, in our view, with the publicly available epidemiological data.

To evaluate the statistical significance of the correlation hypothesis, we created groups of cases based on UBIP record. The selection of jurisdiction cases in the groups was blind, based only on the factors of UBIP policy for the jurisdiction and irrespective of the current Covid-19 impact statistics. Under the null hypothesis, the cases in different groups should have no significant correlation to the outcomes and therefore, have similar distribution; whereas

detecting a significant variation in the outcomes between the groups defined by UBIP policy could place constraints on the null hypothesis. The detailed method of evaluation of the statistical significance of the correlation hypothesis is described in Section 5.

2.2. Data

A time-adjusted selected jurisdictions dataset was compiled from public sources with data of two groups of cases, Group 1 and Group 2 adjusted by the time of first exposure to Covid-19 as suggested by earlier studies [5]. Specifically, the dataset is comprised of the Group 1 cases with local arrival of the pandemics in January / early February 2020, recorded at the time point of TZ + 7 months and Group 2 cases with local arrival in February / March 2020 at TZ + 8 months, i.e. with approximately the same local exposure of six months. It is expected that by this stage, the epidemiological situation has developed to an expressed state in the analyzed jurisdictions.

A number of criteria were applied to selecting the cases in the dataset to minimize the uncertainty factors due to vast variation of conditions among the reporting jurisdictions worldwide:

- A reasonable expectation of the accuracy, consistency and timeliness of the reporting from the national public health administration;

- A reasonable level of exposure to Covid-19, e.g. certain minimum number of reported cases and / or impact;

- A compatible level of social development and specifically, certain minimum standard of public health administration with respect to universal policy administration including importantly, the quality and the coverage.

The aim of the above criteria was to reduce the uncertainty related to the quality of administration of mass public health policy even when such has been declared.

Categories or bands of cases by the reported epidemiological impact were defined based on the logarithmic scale as follows:

Very Low (VL): relative mortality per capita, m_R. in the region of 0.001;

Low (L): $m_R \sim 0.01$;

Medium: (M): $m_R \sim 0.1$

High (H): Relative epidemiological impact greater than 0.2.

While the data for a number of smaller jurisdictions with population under 5 million was recorded in the dataset, they were not included in the statistical analysis due to higher probability of fluctuations related to unpredictable character of cluster development.

BCG universal vaccination record is described by the following bands as defined by Zwerling et al. (2011) [14] and summarized, with certain modifications, in Table 1.

Table 1. BCG universal immunization categories

Category	Description
A	Ongoing universal or near-universal BCG vaccination program.
A2	Has a current UBIP with some limitations or qualifications (for example: a late start; inconsistencies in application practice, possible significant interruptions due to social factors and other).
B	Had a UBIP in the past covering significant part of population.
B2, B3	Immunization program was offered for a limited time interval or specific groups and cannot be qualified as universal over multiple age cohorts; immunization practice inconsistent with the hypothesis of early age induced immunity protection for example, delivered at an older age.
C	Jurisdiction never had a universal BCG immunization program.

Notes and qualifications:

- Consistency and reliability of data reported by the national, regional and local health administrations;

- Alignment in the time of reporting may be an issue due to reporting practices of jurisdictions;

- Availability, consistency and reliability of historical data and statistics on the ad-ministration of immunization programs in the national, regional and so on, jurisdictions can be an issue.

The case dataset can be found in the Appendix I [15-22].

3. Results

3.1. Regional Variation Analysis

In certain jurisdictions significant regional variability in administration of UBIP can be noted, providing essential input relevant to the correlation hypothesis. In the analyzed cases many social parameters, such as living standard, age distribution, social traditions and practices were similar that can be expected to exclude or mitigate the influence of such factors and provide ground for a more confident conclusions of the correlation analysis.

Northern Europe: adjusted to the same time of local exposure, the four cases of Northern Europe show strong correlation between Covid-19 impact and the time of cessation of BCG UIP. These countries share similar levels of prosperity, lifestyle and traditions, climate that allows to eliminate many potential influencing factors.

Table 2. Covid-19 impact vs. cessation of UPIB, Northern Europe

Country	Sweden	Denmark	Norway	Finland
Covid-19 epidemiological impact [1]	568.6	111.8	48.9	60.22
Cessation of UBIP in the jurisdiction	1975	1986	1995	2006

[1] Adjusted to the time point of 6 months of local exposure

A similar pattern can be observed in the cases of Portugal (Group A) and Spain (Group B2), with the relative impact, at the time, of 0.064 and 0.22, respectively. These cases provide anecdotal but consistent over an extended period [5] support for the correlation hypothesis.

3.2. UBIP Cessation vs. Epidemiological Impact

In the group B, where a BCG immunization program existed but was ceased earlier, a strong correlation can be observed between the time of cessation of the UBIP and the severity of Covid-19 impact as shown in the diagram of Figure 2 (the data was time-adjusted to LTZ + 6 months).
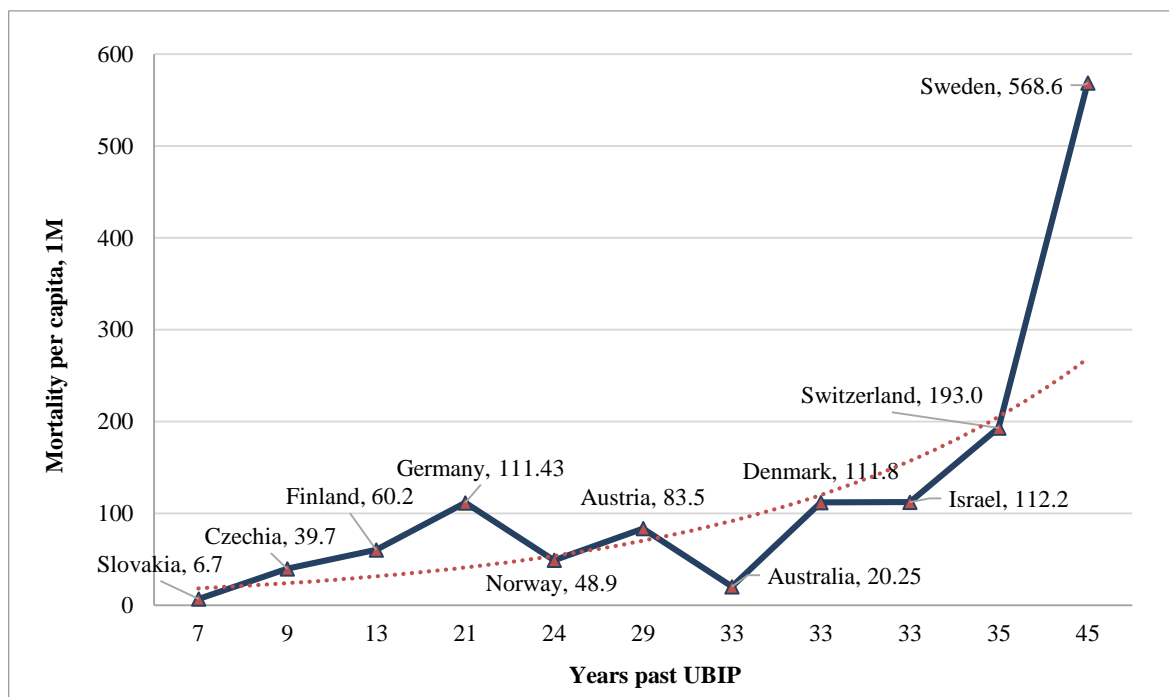

Figure 2. Epidemiological impact vs. time past UBIP, Group B

The data in Figure 2 can be found in Table A2, Appendix I. It can be noted that most of the cases in this analysis had similar factors of demographics, levels of prosperity and overall quality of public health administration with the trend of correlation between the time of cessation of UBIP and the epidemiological impact of Covid-19 impact clearly observable in the diagram.

3.3. Heavy Onset Cases in the UBIP Group A

Several cases of rapid onset of Covid-19 disease were reported in countries with a current BCG UIP, including but probably not limited to, the following: Brazil; Mexico, India, South Africa, Russia, Iran and possibly, others. Without going into specific details of each case that can be done in another study, some general observations can be made here.

In the situations where universality and quality of administration of a UBIP policy in the daily practice could not be ascertained, large group of population may remain with limited or without effective protection even with a formally declared universal policy in place. In cases where these vulnerable groups would happen to be more exposed to the infection, it is possible to see higher impact of the epidemics. The factors such as levels of poverty; quality and access to public healthcare; a record of prolonged social disorder, wars, economic collapse and similar have a strong potential to compromise the administration of UBIP and with it, the hypothetical general protection effect. Most of the observed rapid onset cases in group A with a current UBIP fall into one of these categories, though certainly a more detailed analysis of these cases is warranted. Without a specific and detailed study of a jurisdiction it is not possible, in our view, to determine how essential and influential these factors can be and in this analysis we will limit ourselves to stating that given the generally available records for most if not all of the observed exceptions, they may not be found to be in a strong contradiction with the correlation hypothesis.

Delayed onset. An interesting observation that can be made about the cases in this group is ostensibly significantly delayed onset of the epidemics from the first introduction to the peak of the impact. Comparing the cases of the first arrival of the pandemics in Europe and Far East it can be observed that a development period of approximately $1.5 - 2$ months was common (for example, from end of January to March 2020, when the epidemics was in full development in the European jurisdictions [1, 3, 5]). This can be contrasted to the development period of four months and longer in many cases discussed in this section: Mexico, Brazil, India, South Africa, Russia and other similar cases. Whilst this observation at the time of writing may not have statistical significance, it can be seen as an indirect argument for the correlation hypothesis, as the possibility that the immunized part of the population could delay the transmission of the epidemics to the unprotected groups. This hypothesis requires further analysis and will be addressed elsewhere.

3.4. Possible Mechanisms

The hypothesis of induced early age immunity protection from the exposure to BCG proposed by Dolgikh (2020) [5] based on a number of reports pointing at a possible association between early delivery of BCG vaccine and a broad immunity against several conditions, including infectious [12, 13]. It is further supported by a study indicating a possible mechanism for increased production of immune cells in infants following vaccination with BCG [11, 23] as well as reported gender differences of Covid-19 [24]. A number of works looked at potential mechanisms for such an induced population-level hypothetical protection effect that can be associated with an exposure to BCG vaccine [25, 26].

The research in this important area is ongoing and will be addressed in more detail elsewhere.

4. Statistical Significance of the Correlation Hypothesis

In this section the statistical significance analysis is performed at the time point of six months after the first local exposure to the infection. The analysis is based on evaluation of statistical parameters such as mean and standard deviation [27] of the epidemiological impact between groups of cases blindly selected based on the record of UBIP.

The null hypothesis in this case would dictate that immunization should carry no statistical significance for the epidemics impact, and therefore distributions in all of BCG group sample points (A, B, C) as defined above were described by a single distribution with, possibly, time-dependent parameters: mean epidemiological impact $\mu(t)$, standard deviation $\sigma(t)$ that can be estimated from the overall dataset under the assumption of normal distribution.

The basis for the analysis that follows is the observation of a strong disparity between the sample means in groups A and C formalizing observations in Miller et al. (2020), Escobar et al. (2020) and Dolgikh (2020) studies [1, 2, 5]. Under the null hypothesis, these groups should be treated as the randomly drawn samples of a given size, for which distribution parameters can be estimated with the law of distribution of sample means [28].
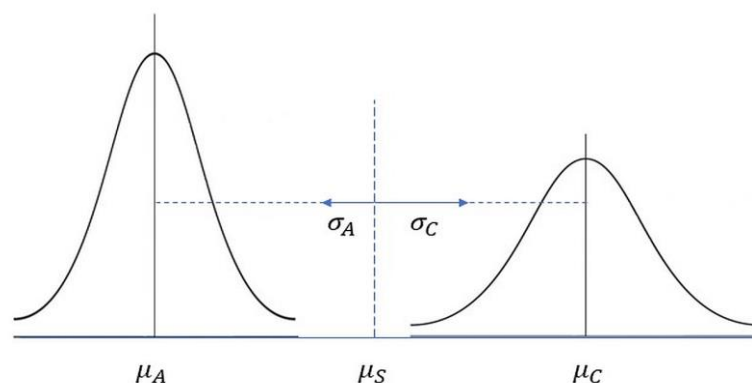

Figure 3. Sample mean distributions, UBIP groups A – C

The groups of cases in the analysis were selected based on the record of UBIP following the criteria outlined earlier in Section 2. The resulting groups were:

Group A (current UBIP and equivalent, protected cohort over 30 years of age): Taiwan, Japan, South Korea, Singapore, Slovakia, Poland, Ukraine, Greece, Portugal, Czechia (2010), Finland (2008), Germany (1998), Norway (1995), Austria (1990).

Group B (previous UBIP, significant unprotected cohort): Sweden (1975), Switzerland (c. 1985), Australia (c. 1985), Israel (1986), Denmark (1987).

Group C (no record of a previous UBIP, and equivalent): Italy, Belgium, Netherlands, United Kingdom (see Notes below), Spain (Notes), USA, Ontario, Quebec (provinces of Canada), California, Florida, New York (states of USA).

For further detail on the composition of the groups refer to the Section A3, Appendix I.

Under the assumption of the null hypothesis all group samples would be drawn from the same distribution and the rule of sample means dictates that the means of the samples of groups A − C with the number of samples N_G will be distributed with the same mean and a standard deviation σ_G as:

$$\sigma_G = \frac{\sigma}{\sqrt{N_G}} \tag{3}$$

where σ is the overall standard deviation of the dataset. From (1) based on the size of each group, one can estimate sample mean standard deviations for the groups A − C samples. For the selected groups of cases, statistical parameters of the epidemiological impact distribution measured in logarithmic mortality per capita $m_L(case, t)$ were obtained from the dataset (the Appendix I) as shown in Table 3.

Table 3. Statistical parameters of the epidemiological impact by UBIP group

Sample	Cases	Sample mean	Sample STD	Sample mean σ	Sample mean distance from set center, σ
Group A	14	4.50	2.46	0.75	− 2.76
Group B	5	6.94	1.74	1.25	0.28
Group C	11	9.09	0.81	0.86	2.90
Combined	30	6.59	2.80	-	0

To satisfy the null hypothesis, the means of UBIP groups A − C would need to independently satisfy the normal distribution laws with the same mean μ_S that can be assumed to be equal to the overall dataset mean and σ_G defined by (1). Then one obtains:

$$P(\mu_A, \mu_B, \mu_C) = P(\mu_A \mid \mu_S, \sigma_A) \times P(\mu_B \mid \mu_S, \sigma_B) \times P(\mu_C \mid \mu_S, \sigma_C) \tag{4}$$

where the first term on the right is the probability of μ_A within the observed range below μ_S with a standard deviation σ_A and the second, similarly, of μ_C within the observed range above μ_S with a standard deviation σ_C. Then from (2) the p-value of the null hypothesis can be estimated as:

$$p \le N(x < -2.76\,\sigma) \times N(x > 2.9\,\sigma) \approx 5.5 \times 10^{-6} \tag{5}$$

excluding the null hypothesis at a confidence level of at least 10^{-5}. The mean of the sample B was close to the dataset mean μ_s and for that reason did not contribute significantly to the p-value constraint.

This result can be illustrated by a histogram of the epidemiological impact for cases in UBIP groups A and C (Figure 4) in the analysis above.
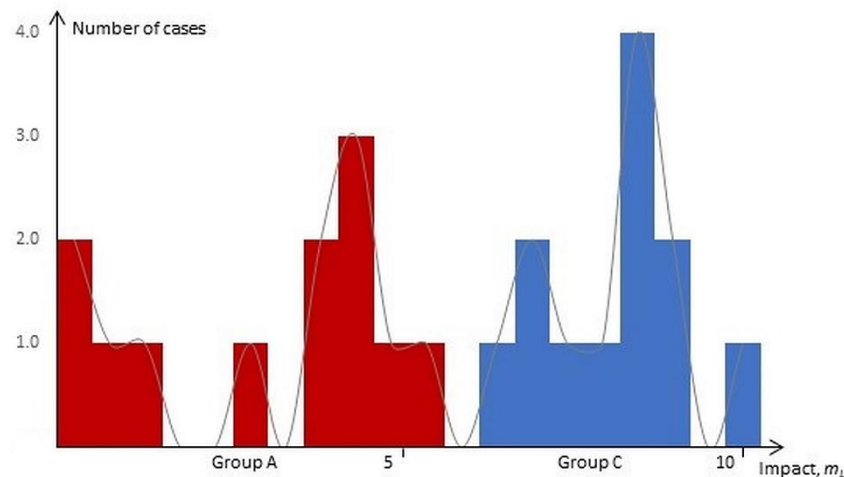
Figure 4. Case histogram by epidemiological impact, UBIP groups A and C

The trend of the groups A (red) and C (blue) to the opposite ends of the impact range in logarithmic mortality per capita can be seen clearly, supporting the results of the statistical significance analysis.

To summarize the results of this section, if the group samples, selected blindly according to the record of UBIP had no correlation with the epidemiological impact and therefore, considered as independent random samples under the null hypothesis, repeated observations of sample means as far apart as in the analyzed groups would lead to strong constraints on the p-value of the null hypothesis.

5. Discussion

A confident determination of a correlation between universal BCG immunization and the type of Covid-19 epidemiological scenario in the national and subnational jurisdictions in the early phase of the development of the pandemics presents a serious challenge due to significant number of potential factors of influence as outlined earlier.

The analysis of observations of distribution of national and subnational cases by epidemiological impact supports the findings in a number of results indicating compatibility of observed data and statistical preference for the correlation hypothesis [3-5]. An original contribution of this work is the regional variation analysis and the analysis of past UBIP cases, Sections 3.1 and 3.2, which demonstrated a clear trend of observed epidemiological impact with respect to the time since cessation of UBIP in the early phase of the epidemics. This finding, observed in jurisdictions with similar socio-economic factors allowed to isolate significant number of possible factors of influence in a broader analysis and provided strong support for the correlation hypothesis.

Direct determination of statistical significance of the null hypothesis (Section 4) provided another strong argument in favor of the correlation. Indeed, under assumption of the null hypothesis, blind selection of cases based entirely on UBIP record and without any reference to the observed epidemiological scenario should have resulted in similar, excluding statistical fluctuations, values of statistical parameters among the groups, including mean and standard deviation. That assumption was found to be in a strong contradiction with the observed data and allowed to impose strong quantitative constraints (5) on the null hypothesis, in an agreement with a number of published results.

Overall, in the authors view, the arguments provided in this work strengthen the case for the correlation hypothesis in the early phases of the epidemics, before jurisdictional differences could impose measurable differences on the development of the epidemiological scenario. These observational and empirical findings are further supported by a connection to immunology studies (Section 3.4) pointing to a number of possible mechanisms for the origin of the effect of broad population-wide mitigation of Covid-19 and possibly, similar infections. This research is ongoing and has a potential to offer essential contributions to epidemiological science and policy [29].

6. Conclusions

The approaches in statistical analysis of the correlation hypothesis of a universal immunization program with Bacillus Calmette-Guérin vaccine and an early phase Covid-19 epidemiological scenario demonstrated in this work with an originally compiled time-adjusted dataset of national and subnational jurisdictions offer additional arguments in support of the correlation hypothesis, indicating a possibility of some form of general population-wide protection effect against Covid-19 resulting from universal immunization program with BCG vaccine, consistent with a number of earlier results. The findings add support to the rationale for further studies of the possible mechanisms of such general protection with potential benefits that may extend well beyond Covid-19 pandemics.

The results of this work in coordination with further research in this promising direction can be instrumental in development of effective responses and policies in the national public health care systems to minimize the impact of the infectious epidemics and protect the population against emergent epidemiological risks.

It is hoped that time-adjusted dataset compiled in this work as observations obtained with it can be useful to other researchers in the field looking for effective approaches to understanding and eventually, effectively managing and controlling this and similar infectious diseases in the future.

In conclusion, the authors would like to emphasize that the results reported in this study are related to the early, initial phases of the development of the Covid-19 epidemics in the studied cases, and cannot be assumed to be applicable in different stages without additional qualifications and / or constraints. For example, in the subsequent stages factors such as quality and effectiveness of epidemiological policy, state of the public healthcare system, availability of critical care resources and other can play more significant role. Choosing the optimal time interval for the statistical analysis in this work allowed to produce some essential insights into the early development of the pandemics.

7. Declarations

7.1. Funding

7.2. Acknowledgements

7.3. Ethical Approval

The study was conducted in accordance with the Declaration of Helsinki. The research uses exclusively publicly available information and does not contain identifiable information.

7.4. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.5. Conflict of Interest

The author declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

8. References

[1] Miller, A., Reandelar, M. J., Fasciglione, K., Roumenova, V., Li, Y., & Otazu, G. H. (2020). Correlation between universal BCG vaccination policy and reduced mortality for COVID-19. medRxiv. doi:10.1101/2020.03.24.20042937

[2] Escobar, L. E., Molina-Cruz, A., & Barillas-Mury, C. (2020). BCG Vaccine Protection from Severe Coronavirus Disease 2019 (COVID19). Proceedings of the National Academy of Sciences, 117(44), 27741-27742. doi:10.1101/2020.05.05.20091975.

[3] Sharma, A., Kumar Sharma, S., Shi, Y., Bucci, E., Carafoli, E., Melino, G., … Das, G. (2020). BCG vaccination policy and preventive chloroquine usage: do they have an impact on COVID-19 pandemic? Cell Death & Disease, 11(7), 1-10. doi:10.1038/s41419-020-2720-9.

[4] Yitbarek, K., Abraham, G., Girma, T., Tilahun, T., & Woldie, M. (2020). The effect of Bacillus Calmette–Guérin (BCG) vaccination in preventing severe infectious respiratory diseases other than TB: Implications for the COVID-19 pandemic. Vaccine, 38(41), 6374–6380. doi:10.1016/j.vaccine.2020.08.018.

[5] Dolgikh S. (2020). Further evidence of a possible correlation between the severity of Covid-19 and BCG immunization. Journal of Infectious Diseases and Epidemiology, 6(2). doi:10.23937/2474-3658/1510120.

[6] Gursel, M., & Gursel, I. (2020). Is global BCG vaccination‐induced trained immunity relevant to the progression of SARS-CoV-2 pandemic? Allergy, 75(7), 1815-1819. doi:10.1111/all.14345.

[7] Berg M.K., Yu Q, Salvador C.E., Melani I. & Kitayama S. (2020). Mandated Bacillus Calmette-Guérin (BCG) vaccination predicts flattened curves for the spread of COVID-19. Science Advances, 6(32) eabc1463. doi:10.1126/sciadv.abc1463.

[8] Meena, J., Yadav, A., & Kumar, J. (2020). BCG Vaccination Policy and Protection against COVID-19. The Indian Journal of Pediatrics, 87(9), 749. doi:10.1007/s12098-020-03371-3.

[9] Paredes J.A., Garduno V. & Torres J. (2020). COVID-19 Related Mortality: Is the BCG Vaccine Truly Effective? medRxiv. doi:10.1101/ 2020.05.01.20087411.

[10] Szigeti R., Kellermayer D. & Kellermayer R. (2020). BCG protects against COVID-19? A word of caution. medRxiv. doi:10.1101/2020.05.01.20087411.

[11] Smith S.G., Kleinnijenhuis J., Netea M.G. & Dockrell H.M. (2017). Whole blood profiling of bacillus Calmette–Guérin-induced trained innate immunity in infants identifies epidermal growth factor, IL-6, platelet-derived growth factor-AB/BB, and natural killer cell activation. Frontiers in Immunology 8, 644. doi:10.3389/fimmu.2017.00644.

[12] Brandau S. and Suttmann H. (2007). Thirty years of BCG immunotherapy for non-muscle invasive bladder cancer: a success story with room for improvement. Biomedicine and Pharmacotherapy, 61 (6), 299–30. doi:10.1016/j.biopha.2007.05.004.

[13] Roy P., Vekemans J., Clark A., Sanderson C., Harris R.C. & White R.G. (2019). Potential effect of age of BCG vaccination on global paediatric tuberculosis mortality: a modelling study, The Lancet, 7(12) e1655-1663 (2019). doi:10.1016/s2214-109x(19)30444-9.

[14] Zwerling A., Behr M.A., Verma A., Brewer T.F., Menzies D., Pai M. (2011). The BCG World Atlas: a database of global BCG vaccination policies and practices. PLOS Medicine. doi:10.1371/journal.pmed.1001012.

[15] BCG World Atlas. Available online: http://www.bcgatlas.org (accessed on September 2020).

[16] Coronavirus map Google. Available online: https://www.google.com/covid19-map (accessed on September 2020).

[17] Canada Covid-19 situation update. Available online: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.html?topic=tilelink (accessed on September 2020).

[18] New York City Covid-19 Updates. Available online: https://www1.nyc.gov/site/doh/covid/covid-19-data.page (accessed on September 2020).

[19] Taiwan Center for Disease Control Covid-19 information. Available online: https://www.cdc.gov.tw/En/Category/ListContent/bg0g_VU_Ysrgkes_KRUDgQ (accessed on September 2020).

[20] National Tuberculosis Advisory Committee. (2006). The BCG vaccine: information and recommendations for use in Australia. Communicable Disease Intelligence 30(1) 109-15.

[21] WHO immunization estimates, BCG by country. Available online: https://apps.who.int/gho/data/node.main.A830 (accessed on September 2020).

[22] Health Ministry of Ukraine. Available online: https://www.kmu.gov.ua/en/news/v-ukrayini-zareyestrovanij-vipadok-zahvoryuvannya-na-covid-19-moz (accessed on September 2020).

[23] Brook, B., Harbeson, D. J., Shannon, C. P., Cai, B., He, D., … Ben-Othman, R. (2020). BCG vaccination–induced emergency granulopoiesis provides rapid protection from neonatal sepsis. Science Translational Medicine, 12(542), eaax4517. doi:10.1126/scitranslmed.aax4517.

[24] Channappanavar R, Fett C, Mack M, Ten Eyck P.P, Meyerholz D.K & Perlman S. (2017). Sex-based differences in susceptibility to Severe Acute Respiratory Syndrome coronavirus infection. Journal of Immunology, 198(10), 4046–4053. doi:10.4049/jimmunol.1601896.

[25] Covián A.R-D., Bueno S.M & Kalergis A.M. (2020). Could BCG vaccination induce protective trained immunity for SARS-CoV-2? Frontiers in Immunology, 11. doi:10.3389/fimmu.2020.00970.

[26] Sinha S. (2020). Possible role of BCG-induced trained immunity to reduce the prevalence and severity of Covid-19 in high-risk population. Journal of Infectious Diseases and Travel Medicine, 4(S1). doi:10.23880/jidtm-16000s1-005.

[27] Freedman D. (2005). Statistical Models: Theory and Practice. Cambridge University Press.

[28] Montgomery D.C. and Runger G.C. (2014). Applied Statistics and Probability for Engineers (6th Ed.). Wiley.

[29] Zeka A., Tobias A., Leonardi G., Bianchi F., Lauriola P. et al. (2020). Responding to COVID-19 requires strong epidemiological evidence of environmental and societal determining factors. The Lancet Planetary Health, 4(9), e375-376. doi:10.1016/s2542-5196(20)30169-8.

[30] Rousseau M.C., Conus F., Ka K., El-Zein M., Parent E.-M. & Menzies D. (2017). Bacillus Calmette-Guérin (BCG) vaccination patterns in the province of Québec, Canada, 1956–1974. Vaccine 35(36), 4777–4784. doi:10.1016/j.vaccine.2017.06.064.

[31] Sweeney E., Dahly D., Seddiq N., Corcoran G., Horgan M. & Sadlier C. (2019). Impact of BCG vaccination on incidence of tuberculosis disease in southern Ireland. BMC Infectious Diseases, 19(1), 397. doi:10.1186/s12879-019-4026-z.

[32] France Loi n° 50-7 du 5 janvier 1950, (1950).

Appendix I

A1. Case Dataset

Table A1 describes the characteristics of the dataset of epidemiological cases at the time of approximately six months after the local arrival of the Covid-19 pandemics. The dataset is available from the corresponding author upon request.

Table A1. Epidemiological analysis case dataset, national and subnational jurisdictions

Region	Number of cases	UBIP groups, A, B, C	Impact, range [1]	Population, range (M)	Local arrival, groups 1, 2 [2]
Europe	23	9; 11; 3			7; 16
Americas	13	4; 3; 6			5; 8
Asia	4	4; 0; 0			4; 0
Other	3	1; 2; 0			1; 2
Total	43	18; 16; 9	0.3 – 2751	2.9 – 327.2	17; 26

[1] In mortality per 1M capita (m.p.c.) at the time point of approximately 6 months after local arrival.

[2] Group 1: c. TZ + 1 month, 01.02.2020; group 2: c. TZ + 2 months, 01.03.2020.

A2. UBIP Groups

Notes and clarifications on selection of UBIP groups A – C for statistical significance analysis.

- Equivalent in Group A were the cases with a recent cessation of UBIP, with unprotected under the correlation hypothesis age cohort under 30 years of age.

- Equivalent in Group C were the cases with a very short or otherwise affected UBIP as further noted below.

- Group A selection: only cases with a reliable and consistent application of UBIP were selected. The case of Ireland was not added in this group due to documented records of inconsistent policy application across some subnational regions [30].

- The case of Spain was placed in Group C due to a very short duration of UBIP (less than 20 years overall) making it negligible for any hypothesized effect of population-wide protection. The same argument was applied in the case of Quebec, Canada.

- The case of United Kingdom placed in Group C due to incompatibility of the UBIP administration practice with the early age induced immunity hypothesis as it was administered at a school or early adolescence age [14].

- Canada, where no UBIP was provided except for a short duration in the province of Quebec [31] was represented by the cases of Ontario and Quebec, with the highest population and epidemiological impact (two subnational cases).

- Due to large population and high regional variation, United States was represented by three state cases California, Florida, New York, and the overall statistics for the country (three subnational cases).

- The case of France where UBIP was in place till 2007 was not included in any UBIP group due to uncertainty about the administration practice. Sources provide inconsistent information about the age of administration of BCG, in the infancy or school age: "BCG was mandatory for school children between 1950 and 2007" [14, 32] indicating a possibility that in the least the practice was not consistent and the decision on the placement in the appropriate group could not be made without further detailed analysis.

Table A2 describes the cases with previous UBIP, Group B.

Table A2. BCG universal immunization categories

Jurisdiction	Impact [1]	UBIP type	UBIP cessation
Czechia	39.7	B	2010
Finland	60.2	B	2006
Germany	111.4	B	1998
Norway	48.9	B	1995
Austria	83.5	B	1990

Australia	20.2	B	1987
Denmark	111.8	B	1986
Israel	112.2	B	1986
Switzerland	193.0	B	c.1985
Sweden	568.6	B	1975
Quebec (Canada) [2]	678.1	B2	1956 – 74
Spain [2]	609.1	B2	1965 – 81
United Kingdom [3]	695.8	B3	2005
France [4]	451.7	B3	2007

[1] In mortality per 1 M capita (m.p.c.) at the time point of approximately 6 months after local arrival

[2] Considered no-UBIP equivalent due to a very short duration of UBIP [14].

[3] Considered no-UBIP equivalent due to immunization practice incompatible with the early age immunity hypothesis [14].

[4] Inconsistent data about the immunization practice [32].