BCG and COVID-19: Correlation or Causality or neither?

TEACHING NOTE

Manoj Dayal Chiba
University of Pretoria Gordon Institute of Business Science, Johannesburg, Gauteng, South Africa

Abdullah Verachia
Faculty, University of Pretoria Gordon Institute of Business Science, Johannesburg, South Africa

CASE SYNOPSIS

The case is set during the period of the COVID-19 pandemic, globally a search for effective treatments were underway. An initial forerunner that was being considered was Bacille Calmette-Guerin (BCG), given its effectiveness in the treatment of Tuberculosis and other pulmonary related infections. While there were a lack of randomised controlled trials, initial data from publicly related secondary data sources indicated that, in countries with BCG inoculation policies, the severity of the spread and mortality of COVID-19 was muted. The case is centered around the available information on BCG and COVID-19.

METHODOLOGY:

The case is built on publicly available information. Information from a host of different types of sources was collated in order to create the narrative. This includes organisation websites and academic journal articles (peer-reviewed and non-peer-reviewed). The case begins with an article published on the website of GlobalData Healthcare on 23 April 2020, on BCG vaccination and protection against COVID-19. Additionally, The Lancet published an article on 30 April 2020, on BCG and reduction on the impact of COVID-19. Considering these two articles as starting points, further information from other sources had begun publishing articles on the effectiveness of BCG on COVID-19. Most notably of the publications is the article by Ricco, Gualerzi, Ranzieri and Bragazzi (2020), which had done a review of 13 articles.

The work by Ricco et al., (2020), reports that there was low ecological validity in all the articles, and raises the concern of drawing conclusion from academic sources (peer reviewed or not) that are publicly available. Initial data visualisation as contained in Exhibit 10 in the main case, seemingly provides evidence that countries with BCG inoculation policies have reduced mortality in comparison to countries without such or targeted policies. The inoculation policies therefore raises a discussion on the different policies adopted by countries. The differential policies adopted by countries are contained in exhibit 3, which provides an overview of the countries and the inoculation policies adopted.

The inoculation policies however only present a single dimension. There are six known strains of BCG that are being used by countries currently, with some countries using more than one strain. While the case highlights this, what is purposefully omitted is the consequence of the effectiveness of some BCG strains over others. This is done in order not to speak about the effectiveness of a specific strain, rather at a conceptual level what these relationships indicate.

Assignment questions:

1. Do you believe that a causal relationship between BCG and various COVID-19 metrics exists? Why? Why not?
2. What are the necessary conditions to prove causality?

TARGET AUDIENCE:

Post-graduate students learning statistics as part of a degree programme. The case assumes no prior statistics knowledge and therefore is aimed at teaching the difference between correlation and causation.

LEARNING OBJECTIVES:

1. Correlation; and
2. Causation

TEACHING PLAN AND OBJECTIVES:

10-15 Minutes: INTRODUCTION TO COVID-19

The course instructor should spend 15-20 minutes providing the context of COVID-19 and a brief history (see supporting material). The pandemic timeline is important to place the case in context. Points to be highlighted which are directly linked to the case:

1. 31 December 2019: Cluster of cases related of Pneumonia identified in Wuhan province of China;
2. 11 March 2020: The World Health Organisation (WHO) characterizes COVID-19 as a pandemic. This is important in the context of the case as any growth plans by Numbers Incorporated would have been set in motion by this date.

3. There are number of vaccine trials that are set in motion to provide a defense towards COVID-19.
4. An investigation by researchers indicated that there may be some veracity in BCG assisting in decreasing the spread, mortality and morbidity of COVID-19. A critical article to be read by the instructor is by Ricco et al., (2020), where a review of the articles alluding to a relationship is provided.

10 Minutes:

The course instructor should provide a synopsis of the case, and a brief on the methodology of the how the case was constructed using publicly available information. This is contained in the methodology section of this teaching note.

25 – 30 minutes: Objective 1: CORRELATION

The instructor should then provide a definition of correlation. A definition that may be used "a single number that describes the degree of the relationship between two variables" (Trochim, 2020). The definition should highlight that this is a statistical definition, and that there are a number of statistical tests that allow for the calculation of the "single number" and includes Pearson's correlation co-efficient and Spearman's Rank correlation. The instructor may want to delve deeper into these and the differences at this junction, however is not a central requirement of the case – and may be best to be discussed post the completion of the case.

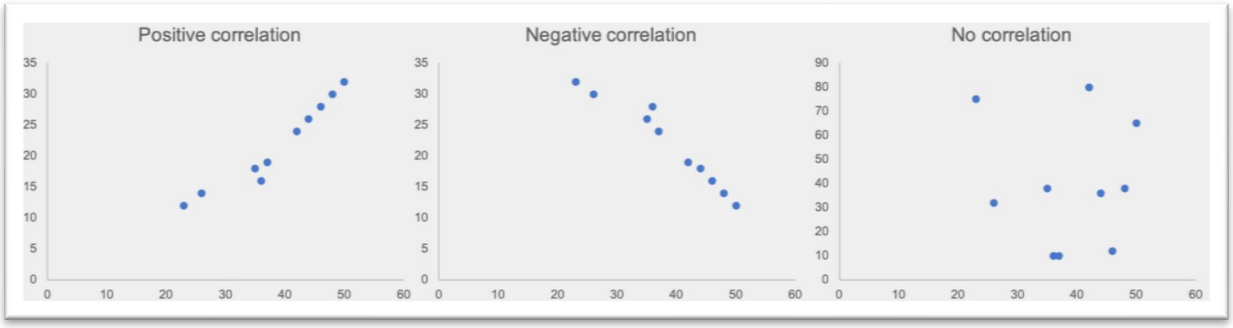
The instructor should next ask the students for examples of correlations (relationships). Once these have been established the instructor needs to highlight that if correlations are present statistically, that this does not indicate that they are true. Examples of spurious correlations that may be used to highlight this can be found at: https://www.tylervigen.com/spurious-correlations

The main aim of the discussion with correlations is to highlight the following:

1. Generally, correlations are when looking at two variables; and
2. The establishment of a correlation does not indicate that it is true.

The course instructor should describe visually the different types of relationships that may exist: Positive correlation, negative correlation and no correlation. A graphical display of this may be used as per the below image:

Figure 1: Positive, negative and no correlation



Finally, the course instructor should highlight that in statistics, there are a number of synonyms used to describe correlation, as apparent in the case, these are: link, relationship, correlation.

30-35 Minutes: Objective 2: CAUSALITY

The course instructor should begin the discussion by probing the students to provide examples of causality. Students would have come across numerous instances where the word causality is used, for example: Smoking causes cancer. The instructor should then probe the students on how would one establish causality.

To establish causality the following criteria need to be met:

1. Cause occurs before the effect;
2. Relationship between the cause and effect
3. No mediating or moderating variables between the cause and effect.

The instructor should apply each of the above to the case, and be clear that before causality can be established that all three criteria need to be established not one or two of the three, all three.

Applied:

1. Cause occurs before the effect – this cannot be established – which are the causal variables? And which are outcome (effect) variables? There is insufficient evidence in the case. The instructor should highlight this, as many students will be fixated on extrapolation of evidence that does not exist in the case;

2. Relationship between BCG and COVID-19: Yes – the numerous articles indicate that a relationship between BCG and COVID-19 exists, and specifically a negative correlation. However, when BCG vaccinations are present there is no evidence of no COVID-19 – that is, the evidence of a relationship is a negative relationship/correlation.
3. There are a host of mediating factors that are presented in the case and these should be highlighted by the students. These include: socio-economic demographics, testing rates, density of individuals per square kilometer, seasonality. These may all be considered as confounding variables.

At this point, the students should be clear that while there may be a relationship between BCG and COVID-19, there is certainly an absence of causality. The instructor should be clear that this is not a lesson on the verification of the truthfulness of BCG and COVID-19, rather how to interpret some of the evidence.

CLOSING:

In closing the instructor should clearly note the following differences between correlation and causation in the least.

1. Causation is uni-directional a → b, that is a causes b
2. Correlation is bi-directional a <-> a is related to b and b is related to a

NOTE: The instructor may want to highlight more differences between the correlation and causality, however as related to the case these are the most pertinent.

SUPPORTING MATERIAL:

Note: In order to teach the case, a brief history of the COVID-19 should be understood, a good timeline of events can be found at: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19.