

Opinions – For Debate

Cite this article: Biswas RK, Afiaz A, Huq S (2020). Underreporting COVID-19: the curious case of the Indian subcontinent. *Epidemiology and Infection* **148**, e207, 1–5. <https://doi.org/10.1017/S0950268820002095>

Received: 14 July 2020

Revised: 3 September 2020

Accepted: 8 September 2020

Key words:

Data validity; health system; pandemic; SARS-CoV-2; underreporting

Author for correspondence:

Raaj Kishore Biswas,

E-mail: raajkishore.biswas@student.unsw.edu.au

¹Transport and Road Safety (TARS) Research Centre, School of Aviation, University of New South Wales, Sydney, New South Wales, Australia; ²Institute of Statistical Research and Training, University of Dhaka, Dhaka, Bangladesh and ³Child Health Research Foundation, Dhaka, Bangladesh

Abstract

COVID-19 has spread across the globe with higher burden placed in Europe and North America. However, the rate of transmission has recently picked up in low- and middle-income countries, particularly in the Indian subcontinent. There is a severe underreporting bias in the existing data available from these countries mostly due to the limitation of resources and accessibility. Most studies comparing cross-country cases or fatalities could fail to account for this systematic bias and reach erroneous conclusions. This paper provides several recommendations on how to effectively tackle these issues regarding data quality, test coverage and case counts.

Since the inception of the COVID-19 pandemic, both the media and research focus were on China, Europe and the USA primarily due to the large cluster of cases in these regions during the early days. However, despite low fatality rates and total cases, the focus then shifted to the low- and middle-income countries (LMICs) soon after the outbreak had hit the Indian subcontinent (ISC) and their COVID-19 response dynamics. In the meantime, academic studies started making inferences on the COVID-19 response effectiveness through comparing the disease prevalence and fatality rates between higher and lower income nations in order to investigate the curious case of low COVID-19 infection rates among the LMICs.

Conducting research on LMICs with limited data could often lead to erroneous findings and biased interpretations, which is becoming a concern with the avalanche of studies published daily. The reasons behind inadequate data in LMICs or even low fatality/detection rates could be qualitatively discussed. While these would not impede academicians in conducting research, caution should be exercised during interpretation and in proposals of ‘evidence-based’ policies and the development of operational plans focusing on mitigation and response in respective contexts. For a greater focus and authors’ area of expertise, this paper is limited to the countries in the ISC, which is a sample representation of LMICs; however, country-wise discussion of LMICs based on the respective socioeconomic context is required for a reasonable generalisation.

Rate and quality of testing

India, Pakistan and Bangladesh are among the worst 20 countries affected by the COVID-19 pandemic in terms of total number of cases; however, they are ranked 138, 139 and 147, respectively, in tests per million population, as of 18 June 2020 [1]. It is worth noting that countries such as Bangladesh reached the threshold of 10 000 tests per day on 20 May, an astonishing 74 days after the detection of the first confirmed case which was still a mere fraction of the number of tests conducted by countries such as the USA, UK and Italy. This lack of testing capabilities during the early days accompanied by the limited protective gears for health personnel and low implementation capacity related to the response of such pandemics could have concealed the true rate of infection and disease spread in the LMICs of the ISC.

Due to limited testing facilities and availability of trained personnel, long delays in both sample collection and dissemination of test results are observed in these countries [2, 3]. Furthermore, these inadequate facilities could compromise sample handling and storage as they require strict low temperature preservation for an optimum result [3], which is a challenge for LMICs including the ISC. Evidently, the official press releases in Bangladesh reflected that all collected samples were not tested daily with long backlogs leading to curbing sample collection [4] and resulting in public distress as many non-COVID medical facilities require certification of a negative test result before admitting new patients.

Another issue for maintaining a rigorous score of transmission rates is to adequately define both cases and deaths from COVID-19, which varies across borders resulting in inconsistencies among reports [5]. For example, countries such as China and Bangladesh, have changed the definition of confirmed cases or recoveries during the on-going pandemic [6, 7].

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Moreover, leadership and political goodwill during such crisis play a crucial role in data collection, testing quality and country-wide coverage [8].

Testing coverage

Testing coverage in the Indian subcontinent has yet to reach the peripheral areas. The government testing facilities are mostly free but are consequently overwhelmed with backlogs, whereas the costs of private tests are well out of the reach of most people [9]. Particularly with countrywide lockdowns and reduced patient transport facilities, it is hard to acquire free-public amenities as an observed 60% increase in extreme poverty in Bangladesh and 90% of people (around 400 million) working in the informal economy in India are at the risk of deeper poverty due to the lack of work during lockdowns [10, 11]. In such scenarios, people living in urban slums or rural areas are likely to prioritise wage-earning activities meeting the heightened unmet basic needs in the midst of low economic flow considering the risks of COVID-19 infection instead of a 14-day post-test quarantine.

Furthermore, given that the testing centres are mostly located in metropolitan areas, the coverage is often centralised to a few locations. For example, as of 18 June 2020, 64.1% of all tests in Bangladesh were conducted in Dhaka district compared to 0.01% in Jessore district [12], and 38.3% tests in Pakistan were conducted in Sindh compared to 0.50% in Gilgit-Baltistan [13]. This decidedly undermines the coverage across these countries and limits the cases to few privileged cohorts challenging identification of the disease at the community level and under-representing cluster transmission for an extended period of time. The lack of testing facilities in peripheral areas requires samples to be transported over long distances in a limited timeframe, which could jeopardise the reliability of samples and test results. The presence of centralised laboratories in few locations contributes to inadequate risk assessments across the country and impacts subsequent decision making.

Funeral statistics

The BBC has put forward an interesting idea of calculating the underreported deaths of COVID-19 using fatality data from previous years [14]. While this could be a way forward, the burden of diseases varies yearly. One particular example could be the expected decrease in road fatalities in 2020 and an increase in suicide rate due to the adverse effects on mental health from stress during lockdown. Similarly, as COVID-19 is not identical with regard to seasonal outbreaks such as measles or dengue in Bangladesh or wild polio in Pakistan, we cannot predict COVID-19 fatality rate from the mortality of the previous seasonal outbreaks, which is likely to lead to a dubious understanding of COVID-19 numbers in these countries. Instead the total increase in death count after adjusting for typical seasonal diseases observed in the previous years could provide a better crude estimate of the impact of COVID-19.

Another avenue of estimating some of the deaths by COVID-like symptoms is data from graveyards and crematoriums. In West Bengal, India, the number of bodies cremated is nearly seven times more than the typical rate, whereas in Dhaka and Narayanganj districts of Bangladesh have seen twice or three times more burials in May 2020 compared to March or April [15, 16]. Disease misclassifications based on the differentiation in information regarding COVID-19 mortality apart from

its comorbidities across multiple sources can adversely impact comparative analyses. This can also undermine subsequent resource mobilisation and evidence-based decision making in generating appropriate COVID-19 management and response worldwide. Moreover, the deceased with COVID-like symptoms are often untested in these countries, which although is understandable considering the resource limitations, but again considerably undermines the overall death tally.

Spread among prominent public figures

Another method of scrutinising the underreporting of cases is to assess the data of frontline workers since they are more likely to be tested alongside the politicians. As of June 2020, 3.28%, 2.88% and 1.45% of COVID-19 fatalities in Bangladesh, Afghanistan and Pakistan were health workers respectively, whereas they were 0.37%, 0.55% and 0.50% in the USA, UK and Italy, respectively. These indicate that the health workers lacked adequate protective gear and knowledge about infection prevention and control measures in the ISC during the preparation phase as well as the fact that they were likely to be over-represented in the tests conducted, leading to an overall underreporting.

As of 19 June 2020, a total of 14 members of the Bangladesh parliament out of 350 had tested positive with two fatalities and over 100 staff of the parliament secretariat, which resulted in a truncated budget discussion in the parliament [17]. Moreover, 8.1% of the COVID-19 infected belongs to the Bangladesh police with 28 fatalities so far [18]. Their testing is expectedly prioritised and the rapid increase in these numbers indicates that community transmission has been severely underreported in the ISC.

Cultural differences

In the ISC, people over 60 generally do not go out of home much, and often their external visits are limited to their familial circle [19]. Furthermore, a considerable number of women are home-makers [20]. Thus, a large portion of the society is used to staying at home, where able males mostly go out for work. While these scenarios are gradually changing, it could partially explain the slower transmission in ISC compared to the developed nations where all household members are more likely to go out increasing the speed of infection.

Social dogma regarding the COVID-19 is also playing a role in these nations. People in Pakistan and India are typically religious. There exist concerns among them that some of the funeral rituals, such as bathing the deceased, cannot be performed if they died of COVID-19 which has created public resentment towards testing [15]. Furthermore, neighbourhood protests were observed in Bangladesh where locals denied the COVID-19 deceased to be buried in their local graveyards [21]. Thus, comparison of case prevalence and fatalities across countries need to consider the cross-cultural and demographic factors.

Biased findings?

There exists a major cause for concern regarding the data quality in the ISC. The subsequent use of these data in their raw form could lead to biased findings [22]. Davies *et al.* rightly found that younger age could be a protective factor in LMICs [23]; however, it is still too early to extrapolate any generalised conclusions. Non-random sampling has been conducted in the ISC, and their limited capacity forces them to test mostly the symptomatic

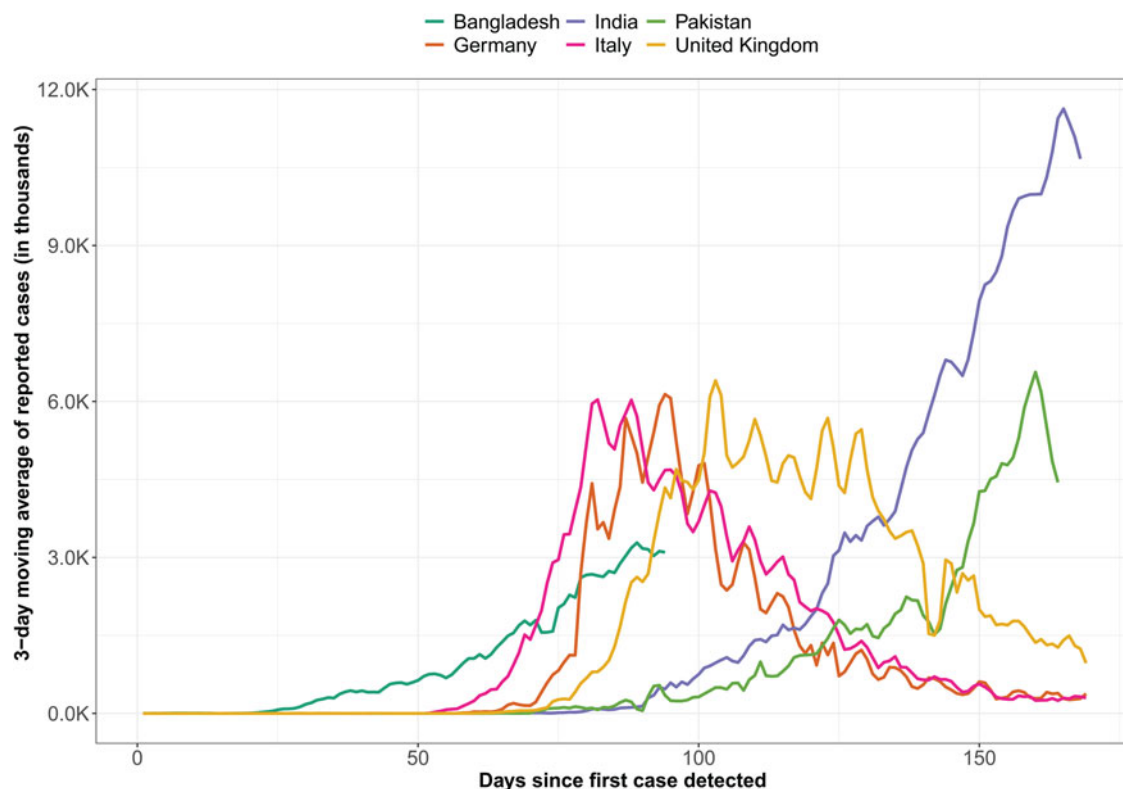


Fig. 1. Distribution of reported COVID-19 cases in three higher income countries (UK, Italy and Germany) and three countries of ISC (Bangladesh, India and Pakistan) for the first 169 days (until 16 June 2020).

individuals and foreign returnees from the high risk countries in the early days. Thus, statistical or epidemiological modelling might be statistically unprincipled with marginalised results when not taking into account the weaknesses of the data generating mechanism [24].

Some findings are developing on a regular basis. For example, in the early days of the pandemic, a hypothesis was shown to be ‘statistically significant’ that temperature is associated with the infection rate [25] without adequate information on true R_0 value and its impact over COVID-19 transmission, which gave a misleading hope to politicians who used it to assure the general mass in Bangladesh leading to a sense of nationwide complacency [26]. Another assumption was that they are genetically immune to the coronavirus, or BCG vaccine might work as a protective factor [27]; however, the large death tolls of Bangladeshi expatriates in Saudi Arabia, Singapore and New York have evidently debunked it [28]. This false optimism has led to relaxation of social distancing policies in public transport and consideration of opening schools across the country [29, 30].

Recommendations

While it is inevitable that modelling with data from LMICs on COVID-19 would continue, a few cautions should be exercised:

- An appropriate definition of ‘death from COVID-19’ is essential before collapsing deaths from the COVID-like symptoms with the COVID-19 fatalities. For example, Bangladesh changed the definition of COVID-19 recovery a month after the detection of the first case [7].
- For validating the COVID-19 fatality scores of a region, specific mortality causes of comorbid conditions such as respiratory and

cardiovascular complication or communicable diseases representing similar manifestation of symptoms linked to COVID-19 could be coded to calibrate from the total deaths during the pandemic period. However, this also needs to consider the seasonal outbreaks of diseases in specific regions.

- Comparison among seemingly random countries based on convenience or data availability might lead to a systematic bias. A comparative assessment on countries with similar testing coverages, analogous socioeconomic context, close geographical borders (e.g. EU or ISC), cultural resemblances and context-specific priorities with homogeneous health systems might be more insightful.
- Contrasting country-wise performances and COVID-19 infection timeline, where the COVID-19 prevalence curve has started to flatten with countries that are yet to reach its peak (Figs 1 and 2), are unreasonable as true propensity of the pandemic is yet to be observed in LMICs such as ISC.
- Instead of modelling COVID-19 incidence rates across borders, cultures and demographics, this could be limited to regions with homogeneous attributes. Comparison between neighbourhoods in the same locality might be better suited for hypothesis testing on mortality and disease prevalence, with the utmost care in avoiding narratives that might mislead the public opinion.
- Validating the COVID-19 official data with random sampling, hospital data, disease burden trends and local news outlets could account for some of the underreporting biases.

With thousands of PhD dissertations and research articles developing the evidence-base are expected on COVID-19 in years to come, it is imperative that the data validity is constantly

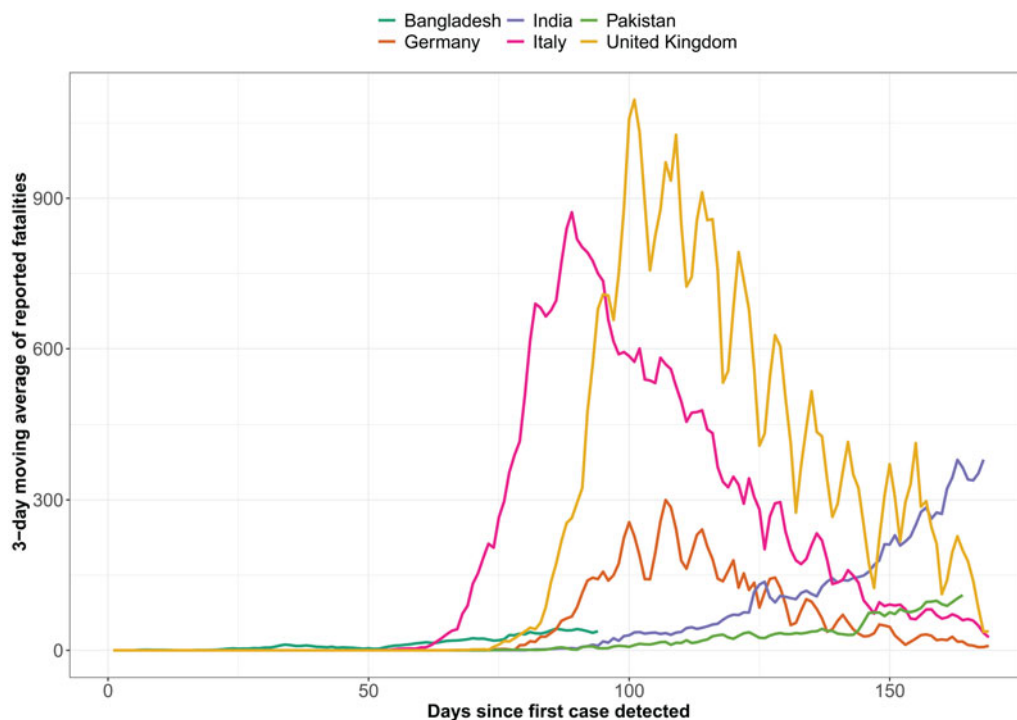


Fig. 2. Distribution of reported COVID-19 fatalities in three higher income countries (UK, Italy and Germany) and three countries of ISC (Bangladesh, India and Pakistan) for the first 169 days (until 16 June 2020).

questioned, and cross-border comparisons are routinely scrutinised given the definition of fatalities from COVID-like symptoms and quality of non-random data vary worldwide.

Acknowledgements. The authors acknowledge European Centre for Disease Prevention and Control, who regularly published worldwide COVID-19 data. We convey thanks to the media for the situational analysis. We extend our appreciation to the three anonymous reviewers who have significantly improved the focus and clarity of the paper.

Author contributions. R.K. Biswas conceptualised the study and drafted the manuscript. A. Afiaz coded the figures, and critically revised the manuscript. S. Huq critically reviewed and edited the manuscript. The final manuscript was read and approved by all the authors.

Financial support. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest. There was no conflict of interest among the authors.

Ethical standards. Publicly available data were used in this study. The compiled data would be made available upon request.

Data availability statement. The data that support the findings of this study are openly available at <https://www.ecdc.europa.eu/en>.

References

1. **Worldometer.** Coronavirus Cases (Accessed: 19 June 2020). Worldometer. 2020.
2. **The DAWN.** Lower Dir residents complain of delay in virus test results (published: 12 June 2020) [Internet] [cited 2020 Jun 19]. p. 2020. Available from: <https://www.dawn.com/news/1562889>
3. **The Business Standard.** As tests pile up, infections spread, patients suffer (published: 15 June 2020) [Internet] 2020 [cited 2020 Jun 20]. Available from: <https://tbsnews.net/coronavirus-chronicle/covid-19-bangladesh/tests-pile-infections-spread-patients-suffer-93376>
4. **Dhaka Tribune.** COVID-19: Bangladesh lags behind in sample testing despite being among 20 worst affected; Dhaka Tribune (published: 5 June 2020) [Internet]. 2020 [cited 2020 Jun 19]. Available from: <https://www.dhakatribune.com/health/coronavirus/2020/06/05/covid-19-bangladesh-lags-behind-in-sample-testing-despite-being-among-20-worst-affected>
5. **Onder G, Rezza G and Brusaferro S** (2020) Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA – Journal of the American Medical Association* [Internet] [cited 2020 Aug 22];323:1775–1776. Available from: <https://jamanetwork.com/journals/jama/article-abstract/2763667>
6. **Singhal T** (2020) A review of coronavirus disease-2019 (COVID-19). *Indian Journal of Pediatrics.* 87, 281–286.
7. **The Daily Star.** Explaining the jump in number of COVID-19 recovered patients in Bangladesh (published: 4 May 2020) [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://www.thedailystar.net/coronavirus-deadly-new-threat/news/explaining-the-jump-number-covid-19-recovered-patients-bangladesh-1899127>
8. **Biswas RK et al.** (2020) A systematic assessment on COVID-19 preparedness and transition strategy in Bangladesh. *Journal of Evaluation in Clinical Practice* [Internet]. 2020 Aug 21 [cited 2020 Aug 22];1–13. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jep.13467>
9. **The Caravan.** India's private COVID-19 tests cost highest in South Asia; free in Bangladesh (Published: 23 April 2020) [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://caravanmagazine.in/health/india-private-covid-19-tests-cost-highest-in-south-asia-free-in-bangladesh>
10. **BRAC Bangladesh.** COVID-19 Awareness and Economic Impact [Internet]. 2020. Available from: <http://www.brac.net/program/wp-content/uploads/2020/04/Rapid-Perception-Survey-On-COVID19-Awareness-and-Economic-Impact.pdf>
11. **International Labor Organization** (2020) COVID-19 and the World of Work (7 April 2020) [Internet], 4th Edn. Geneva, Switzerland: ILO Monitor. 2020 [cited 2020 Aug 22]. Available from: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/briefingnote/wcms_745963.pdf

12. **Government of Bangladesh.** Coronavirus info Bangladesh: Press release [Internet]. 2020. Available from: <https://corona.gov.bd/press-release>
13. **Government of Pakistan.** Coronavirus in Pakistan [Internet]. 2020 [cited 2020 Jun 20]. Available from: <http://covid.gov.pk/>
14. **BBC World.** Coronavirus: What is the true death toll of the pandemic? [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://www.bbc.com/news/world-53073046>
15. **The Telegraph.** Asia's hidden deaths: coronavirus fatalities are being covered up and undercounted (published: 22 May 2020) [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://www.telegraph.co.uk/global-health/science-and-disease/asias-hidden-deaths-coronavirus-fatalities-covered-undercounted/>
16. **Dhaka Tribune.** Burials in Dhaka rose by a third in May (published: 10 June 2020) [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://www.dhakatribune.com/bangladesh/dhaka/2020/06/10/burials-in-dhaka-rose-by-a-third-in-may>
17. **The Daily Star.** Budget session to be cut short as MPs, JS staff infected with COVID-19 (published: 18 June 2020) [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://www.thedailystar.net/budget-2020-21-session-be-cut-short-mp-jatiya-sangsad-staff-infected-covid-19-1916665>
18. **United News of Bangladesh.** Cop dies of COVID-19 in Bogura (published: 19 June 2020) [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://www.unb.com.bd/category/Bangladesh/cop-dies-of-covid-19-in-bogura/53324>
19. **Shwalb DW et al.** (2018) Grandparents in Bangladesh, India, and Pakistan. In: Grandparents in Cultural Context [Internet]. 2018 [cited 2020 Aug 22]. New York, pp. 159–186. Available from: <https://doi.org/10.4324/9781315642284>
20. **Bhopal K** (2019) Gender, 'race' and patriarchy: a study of South Asian women [Internet]. 2019 [cited 2020 Aug 22]. Available from: <https://doi.org/10.1177/030981689906700107>
21. **Dhaka Tribune.** Panic continues to obstruct burials of coronavirus patients (published: 2 April 2020) [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://www.dhakatribune.com/bangladesh/2020/04/02/panic-stricken-people-continue-to-obstruct-burials-of-coronavirus-patients>
22. **Huq S and Biswas RK** (2020) COVID-19 in Bangladesh: data deficiency to delayed decision. *Journal of Global Health* [Internet]. 1. Available from: <https://dx.doi.org/10.7189%2Fjogh.10.010342>
23. **Davies NG et al.** (2020) Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine* 26, 1205–1211 [Internet]. 2020 [cited 2020 Jun 19]; 2020.03.24.20043018. Available from: <https://www.nature.com/articles/s41591-020-0962-9>
24. **Meng X-L** (2020) COVID-19: a massive stress test with many unexpected opportunities (for data science). *Harvard Data Science Review* [Internet]. 2020 [cited 2020 Jun 20]; Available from: <https://hdr.mitpress.mit.edu/pub/l7a2t45s>
25. **Wang J et al.** (2020) High temperature and high humidity reduce the transmission of COVID-19 [internet]. *SSRN Electronic Journal*. 2020 [cited 2020 Jun 20]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3557152
26. **New Age.** Warm weather may buy Bangladesh some time to fight coronavirus (published: 16 Mar 2020) [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://www.newagebd.net/article/102375/warm-weather-may-buy-bangladesh-some-time-to-fight-coronavirus>
27. **The Daily Star.** Could TB vaccine protect medics from COVID-19? (Published: 04 April 2020) [Internet] [cited 2020 Jun 20]. Available from: <https://www.thedailystar.net/world/news/could-tb-vaccine-protect-medics-covid-19-1889419>
28. **The Daily Star.** 375 expatriate Bangladeshis died of COVID-19, symptoms in Saudi Arabia (published: 19 June 2020) [Internet]. 2020 [cited 2020 Jun 20]. Available from: <https://www.thedailystar.net/375-expatriate-bangladeshis-died-covid-19-symptoms-in-saudi-arabia-1917101>
29. **The Daily Star.** Road transport association leaders want to operate vehicles at full capacity (published on: 19 August 2020). 2020 [cited 2020 Aug 22]; Available from: <https://www.thedailystar.net/bangladesh-road-transport-association-leaders-want-operate-vehicles-full-capacity-1947641>
30. **Dhaka Tribune.** Bangladesh plans to reopen schools combining online, offline education (published on: 7 August 2020). 2020 [cited 2020 Aug 22]; Available from: <https://www.dhakatribune.com/bangladesh/2020/08/07/bangladesh-plans-to-reopen-schools-combining-online-and-offline-education>