COVID-19: Data collection and transparency among countries

16

Erwin Calgua

Research Center of Health Sciences, School of Medicine, University of San Carlos of Guatemala,
Guatemala City, Guatemala

Introduction

The coronavirus disease 2019 (COVID-19) pandemic is a public health problem with extensive implications for health systems and economies around the globe. The first cases were reported as a pneumonia of unknown origin in Wuhan, China, on December 31, 2019.[1] The World Health Organization (WHO) announced through social media outlets about these cases during the first week of January, and in the same week, the agent was identified as a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).[2] This new coronavirus showed immediately the attribute to spread efficiently to new countries and regions outside China, such as Philippines, France, Italy, Spain, and the United States.

It is important to note that the first data was collected and analyzed by local Chinese authorities where the outbreak was taking place, and the timing that took identifying the first case to the moment that they shared the information with WHO was reported as immediate, a claim that was not challenged in the first phase of the pandemic. Another interesting fact is that, in the very beginning, WHO or the scientific community did not have any doubt about the authenticity of the data being shared by Chinese officials. The data began to flow through the weekly WHO bulletins, and later, the Johns Hopkins University (JHU) developed the Coronavirus Resource Center that kept track of the coronavirus cases based largely on data provided by 188 different countries and regions of the world.[3]

As the pandemic developed, most individuals turned to the JHU resource to have a timely and reliable data of the COVID-19 cases. As of the preparation of this chapter, this resource has been the most popular (even more popular that the official site of WHO for COVID-19[4]), despite the fact that a few actually know where the data for its dashboard comes from, and even less, take the time to verify the authenticity of the source of the data used for this information, which includes media reports. For a while, this information was not challenged, but as the dynamic of the pandemic developed, questions arose about the data that were collected and about the information provided after the analysis of such data. Here is where the issue of transparency emerged, especially triggered by the reports that questioned the information

provided by Chinese officials in regard to the start of the pandemic. At the same time, similar concerns were raised in other high-income countries,[5] including lower middle-income countries like Guatemala, that challenged the transparency of the data and information presented by local authorities.

With this chapter, we want the reader to have a brief overview, from the perspective of a clinical epidemiologist, of the lessons learned on two key epidemiological aspects in the process to inform authorities and society related to COVID-19 being those: (1) data collection and (2) transparency in the data.

Lessons learned about COVID-19 data collection

The WHO is the leading institution when it comes to data and information related to an epidemiological emergency.[6] For the purposes of this chapter, it is important to understand that WHO's key role is to coordinate data collection and provides information based on analysis of the data in a global context. Based on this information, technical and financial support is provided to countries around the globe. Of utmost importance is to note that WHO monitors constantly the data and information that is provided by the different countries and regions of the world.[7] Also, interesting to note is that collecting the data in a dynamic global context is both a strength and a weakness, because what WHO gains from obtaining data of authorities, losses in the quality of the data that is provided by Ministries of Health, that respond directly to the highest authorities in government who decide which data and when the data is released, something that is of key importance when making decisions in the midst of an epidemiological emergency, as the current COVID-19 pandemic.

During the first days of the COVID-19 pandemic, there were no questions, at least in the general public, about the data and information that was provided through WHO bulletins or the JHU site, especially the ones that came from China. Then, as the virus spread around the globe, data from France, Italy, Spain, and the United States emerged, and in a similar manner, there were no questions about the data and information. But things began to change as the pandemic evolved, and questions began to raise in regard to the data and information provided by China, after the concerns that were expressed by the highest authorities in the United States government.

As a response, China issued a statement which confirmed that they have followed the protocols and processes required by WHO in matters pertaining to data collection and reporting, making emphasis that they have complied with all these standards based on international agreements, in a timely manner. We must draw the attention that at this point, there were no questions from WHO or other key institutions leading to believe that the data collection and analysis provided by China or any other government around the world were flawed. However, the media and news outlets, through investigative journalist methods,[8] inquired about the authenticity of the data provided by governments and centered their attention in the question of hidden data. The investigations lead to findings such as that the first case could have been in the mid of November 2019.[9] In addition, independent investigators from Harvard

University conducted a study based on satellite pictures and online searches around the Wuhan area and reported that the first cases could have happened as early as August 2019[10]. Through these findings, we have learned two important lessons which are: first, media and news outlets play a key role in creating questions for the public opinion about the way data are obtained and how this information is communicated, and second, the importance of trust in local and global authorities about the transparency in management of data collection and generation of information, as it is mentioned in WHO fundamental guidelines on how to collect data during epidemiological emergencies.

Lessons learned about COVID-19 data transparency

Transparent is defined as being "free from pretense or deceit",[11] and it is expected that by all means, a government is committed to be obedient to the law, and even more, comply with high ethical standards in the way that manage data and communicate information to increase the benefits for the general public and avoid risk that could harm their lives. This, in short, should be the main aspects that any government should follow, so the mitigation directions issued are followed by society and in that way prevent the impact and high cost in morbidity and mortality of vulnerable populations during an epidemiological emergency.

So far, COVID-19 pandemic has been a challenge in regard to empower governments to lead the actions that could mitigate the impact of this virus in their countries. China, being a high-income country, has been questioned about the transparency of the data and information that provided to the world.[12] One of the errors they have made was the decision to control and limit the scientific information that was requested by the scientific communities[13] to understand better how SARS-CoV-2 appeared and behaved. No information could be provided without government authorization. That simple decision created a distrust in what was reported by Chinese institutions, something that has remained to date and could take time to overcome. Therefore the sole reaction of limiting the information was perhaps one of the most damaging actions of the Chinese government, and the worst part is that the media was not the only one reporting the lack of transparency from the Chinese government but also the scientific networks around the world raised their concern that Chinese scientists were coerced to not to share their data and information and even seek collaborations with scientists who wanted to understand better COVID-19 as the SARS-CoV-2 was spreading to other countries.

Unfortunately, similar attitudes were shown governments of lower middle-income countries, like Guatemala where the first case was reported on March 13. During the first month, the information provided by the Ministry of Health was openly received by media outlets, population, and scientific community, and in short, no questions were made as of the reliability and timely manner that this was received. However, as the pandemic developed, the scientific community and society had concerns about the number of COVID-19 tests that were performed. Certainly,

the acquisition of tests was a very difficult and important issue to address, from the government perspective, and even more challenging was to communicate to the general public about the lack of tests around the globe[14]. Among the general public, it is a common belief that a test has been validated by the time that a pandemic starts, and just a few could understand that such is not the case in certain epidemiological emergencies as was the case for SARS-CoV-2. Societies believe that funding is enough to secure a test, and in certain way it does, but with SARS-CoV-2, the issue of having funds did not secure the possibility of having a reliable test in which the health systems could trust to identify the cases among the general population. Therefore, there was a key communication issue that needed to be addressed, which was to explain the relation that has the availability of a test and the possibility to confirm a case (no test, no data), which limits the possibility to understand how a highly infectious disease behaves from the beginning.

Another situation that happened in Guatemala was the presidential daily briefing that presented to the public the COVID-19 information. The format of such briefings was inconsistent, and there was a lack of clarity for the general public. Therefore, the Guatemalan population did not completely understand the data and even noticed when contradicted data was presented to them. Based on that, as in China, news outlets[15] in Guatemala began to search deeply into the information[16] and finally, the government acknowledged that there was a sub-registry in the data that they were providing during the briefing and error not directly related to the lack of tests, but of information that was not transferred to them, especially, in regard to the number of deaths occurring in Guatemala. What should be considered as simple situation turned out to be a key issue that has been the center of what is considered as the main driver of the lack of success in the mitigation measures implemented by the government, which was the perception of lack of transparency and fear of corruption that was reported in the media,[17] and eventually, this ended up in the distrust on government planning and decisions, which lead to an increased number of COVID-19 cases, something that has been the trend to the day that this chapter was written.

What could be implemented to improve data collection for COVID-19

If data collection is a fundamental aspect that must be defined in the beginning of an epidemiological emergency, then it is important that Ministries of Health, as the highest authorities, must clearly define the procedures that not only locally but also globally must govern the data collection and subsequent storage and analysis to provide timely and reliable information. One main limitation that exists around the globe is that there is not a universal validated tool that could be used and applied in local and international contexts to collect data with the purpose of informing in a timely manner of the cases who are identified during an epidemic crisis.

There have been efforts to develop such tools, but the initiatives have moved only to the stage of a proposal, never becoming a universal tool that could be used by

Ministries of Health. There could be many reasons that could explain why the implementation of such tools might be challenging. First, there are "political" reasons, which usually are more economical reasons, that intentionally avoid to create a collaborative environment. It has been well established that when a country faces an epidemiological emergency, the economic impact becomes the highest priority relegating the provision of adequate health services to a last place. In many ways, COVID-19 is teaching us that prioritizing the economic factors over the health will eventually lead to a more complex health problem. Hiding the data from the international scientific community, and not giving time to carefully analyze it, will eventually create a situation that will affect not only the country that was primarily affected but also the globe to a point that even commerce could stagger and deepen the economic inequalities.[18] Governments, and specifically Ministries of Health, must carefully begin to outline clear policies and plans that will guide actions during such epidemiological emergencies, and the first public health policy in epidemiological emergencies should be the one about data collecting and sharing with the international community and the general public.[19]

Technical teams at Ministries of Health must know that if there is a need to share data, an important issue should be what data must be collected. That takes us again to the universal tool that we have mentioned before. The tool selected must be developed in consensus with other governments, and a guideline with a clear process must be prepared to be used in epidemiological emergencies. It will be expected that such guidelines should be developed in consultation with the WHO, and a clear statement should be made within the United Nations in order to adopt such tool in a universal manner. There are immediate limitations in this proposal, mainly due to political distinctions among countries and regions, especially between world potencies, that ultimately have the possibility to vote and decide if they will adopt the tool and share this data in the areas where they exercise an economical and military influence (we must keep present that an epidemiological emergency as the SARS-CoV-2 pandemic is considered a national security concern for world potencies). Undoubtedly, there is a primary institution that should promote the development of such a tool and that should be the WHO; however, if we argue that today's pandemics have a heavy impact in international commerce, then another instance that should be considered to coordinate these efforts is the World Trade Organization (WTO). This second option should not be a surprise,[20] because we have argued before that at the beginning of a pandemic, the priorities of governments are usually commerce over health, and COVID-19 clearly has become a commerce problem with local and international implications.

Once the world has consented the organization from where this tool will be developed, then it is important to define what kind of data should be collected. This again should not be a difficult issue to define. A common mistake of Ministries of Health is trying to collect a vast amount of data that ends up in large quantities of bytes stored in some electronic device that is never analyzed. But this should change. As mentioned earlier, of key importance is to be able to accurately count the cases. COVID-19 was difficult in the aspect that there was no diagnosis test available,

and once this was developed, such test was inaccessible for two reasons. First, because the technologies used to identify the antigen usually requires highly trained human resources and specialized equipment that is not easily available, especially in lower middle-income countries. Second, once the technology has been developed, there is the possibility that patents and licenses for such technologies are too expansive, that acquisition becomes prohibitive for countries that do not have the means to fund immediately the acquisition of such tests. This directly impacts the possibility to have accurate data (remember: no test, no data) that could be reported in a local and global context. Therefore every effort should be made toward obtaining and having the possibility to have availability and accessibility to a diagnosis test that could be applied globally, regardless of being in a high-, middle-, or lower-income country. Certainly, an economy based on free market will challenge this proposal; nevertheless, COVID-19 is teaching us that in a context where a pandemic exists in the free market around the world will be as strong as the weakest of the countries.

One important aspect also is to define the data that should be collected.[21] It is possible that the first data that must be collected will be the cases confirmed by an accurate diagnostic test. The second data that should be obtained is the swabs performed.[22] Interestingly, most of the countries report the number of positive cases per day; however, this might be accurate for high-income countries where there is the technology readily available and in a short amount of time the results are available, but it is difficult to find such situations in a lower middle-income countries. Guatemala's experience was that in some cases it took 8 days to report the test result. So, the fact that this was presented as positive in a given date, which did not reflect the exact moment of when the case was happening, giving an inaccurate picture of the situation in real time.[23] Is in that regard that it would be preferable to report the number of positive cases based on the daily swabs than the daily performed diagnosis.

The third data will be the number of deaths related to COVID-19, which is another important and challenging data that must be collected and reported. The challenge of this data is that it also needs to have a positive test for the diagnosis; however, it would be even better if the death could be also confirmed with an autopsy. But, the fact that most health services lack a test at the beginning of the pandemic, and it is not readily available for a long time, as previously stated, it might be difficult for countries to have a positive test on one end. On the other end, most health services, once the patient dies, immediately proceed to bury the cadaver without an autopsy. The lack of an autopsy limits the possibility to better understand the viral action in all the systems that comprise the human body. What would be better and should be suggested, is to conduct autopsies of the cases[24] and report the number of autopsies that finally confirms that the death was directly related to COVID-19.[25] Other mandatory data must be collected, like demographic data from the patient, length of hospitalization, etc. These suggestions of data that could be collected have the potential to inform more than a vast number of pages with information that none reads or finds practical use in order to inform decisions at all levels of public health and clinical settings.

In a global context, it would be suggested to have few data that could be shared immediately to a central organization that could collect timely and reliable data from all Ministries of Health around the world and have an accurate image of the global situation in a given epidemiological emergency.

What could be implemented to improve data transparency for COVID-19

Given the sensitive nature of the data and information that is generated during an epidemiological emergency (something that was previously stated), one of the most important challenges is how to manage this information once it reaches the government authorities. The decision on how and when information will be released to the general public ends up at the top level of a government, usually the President or Prime Minister. The decision will be driven by many factors, overall, the economic impact that could have the release of such information to the public. However, governments need to rethink and define immediately laws and policies based on firm moral grounds that clearly state that, under an epidemiological emergency, key data will be collected and shared through a central organization and that will finally create the possibility to report at least the above-mentioned indicators.

As of today, there are guidelines, mostly weak, in regard on how to proceed with providing the information to the entities that coordinate the global efforts of a pandemic, but this has not been reinforced. Undoubtedly, there is need that countries agree to develop such laws and policies within their countries, and after passing these laws locally, this must be communicated to a centralized organization body and abide by such laws and policies to move forward and begin to have a universal system that collects such data and communicates such data around the globe.

An important aspect of all this is that transparency is deeply rooted in moral arguments more than legal. The implications to abide such laws and policies will be of key importance in order to develop trust in the general public, and again, trust is of key importance when managing an epidemic crisis. The teams who will need to discuss the laws and policies must write them based on firm arguments that will stand even in the most difficult and perilous situations. COVID-19 has shown us how the lack of transparency can undermine any effort directed to mitigate the effect of pandemic. Even more important, transparency could create an environment of peace instead of conflict between nations, which has been an important subject with deep effects, like the one raised by the United States and Chinese governments[26, 27] or more locally like the one in Guatemala where the lack of transparency by the government ended up in the distrust of the population and therefore the increased number of cases and unmanageable situation of the pandemics.

Conclusion

In conclusion, data collection and transparency are two important topics closely related with strong impact in the appropriate management of an epidemiological emergency. COVID-19 has thought us globally and locally that economic impact drives the decision on how the data is obtained and what is released to the public. So far, both in high-income countries like China or low-middle–income countries like Guatemala, there is a perception that the data was flaw and inaccurate[28], and therefore, that lead to a situation of distrust that worsened the crisis. It is proposed that based on the lessons learned, that a centralized organization, whether this be WHO or WTO, collects few but substantial data that could be used for global and local decisions, data like number of positive tests, positive swabs and autopsies per day should be considered, but is not limited to this three, as there are other important demographic and clinical data that could be obtained and reported. In addition, local laws and policies related to transparency should be discussed and approved by governments in order to develop trust in the public opinion, which plays a key role in the ethical, legal, and appropriate way to manage an epidemiological emergency like the one we have experienced with COVID-19.

References

1. Archived: WHO Timeline—COVID-19. *Retrieved July 4, 2020, from*; 2020. https://www.who.int/news-room/detail/27-04-2020-who-timeline- - -covid-19.
2. WHO. *Novel coronavirus (COVID-19). Retrieved July 4, 2020, from*; 2020. https://www.who.int/bulletin/online_first/COVID-19/en/.
3. COVID-19 Map - Johns Hopkins Coronavirus Resource Center. *Retrieved July 4, 2020, from*; 2020. https://coronavirus.jhu.edu/map.html.
4. WHO Coronavirus Disease (COVID-19) Dashboard. *WHO Coronavirus Disease (COVID-19) Dashboard*; 2020. Retrieved July 4, 2020, from https://covid19.who.int/?gclid=EAIaIQobChMI_LmtvIGz6gIVCgSRCh1lSgMTEAAYASABEgK96PD_BwE.
5. Should we believe the reported COVID-19 data, statistics and headlines? - Coronavirus (COVID-19) - Australia. *Retrieved July 4, 2020, from*; 2020. https://www.mondaq.com/australia/operational-impacts-and-strategy/928868/should-we-believe-the-reported-covid-19-data-statistics-and-headlines.
6. Rao C. Medical certification of cause of death for COVID-19. In: *Bulletin of the World Health Organization*. Vol. 98, Issue 5. World Health Organization; 2020. https://doi.org/10.2471/BLT.20.257600.
7. Moorthy V, Restrepo AMH, Preziosi MP, Swaminathan S. Data sharing for novel coronavirus (COVID-19). In: *Bulletin of the World Health Organization*. Vol. 98, Issue 3. World Health Organization; 2020:150. https://doi.org/10.2471/BLT.20.251561.
8. Can We Believe Any of China's Coronavirus Numbers? *Time*; 2020. Retrieved July 4, 2020, from https://time.com/5813628/china-coronavirus-statistics-wuhan/.
9. 1st known case of coronavirus traced back to November in China. *Live Science*; 2020. Retrieved July 4, 2020, from https://www.livescience.com/first-case-coronavirus-found.html.

10. Coronavirus: Fact-checking claims it might have started in August 2019. *BBC News*; 2020. Retrieved July 4, 2020, from https://www.bbc.com/news/world-asia-china-53005768.

11. Transparent Definition of Transparent by Merriam-Webster. Retrieved July 4, 2020, from (n.d.). https://www.merriam-webster.com/dictionary/transparent.

12. Western calls for greater Chinese transparency on the coronavirus reflect a clash of culture. *South China Morning Post*; 2020. Retrieved July 4, 2020, from https://www.scmp.com/comment/opinion/article/3081931/western-calls-greater-chinese-transparency-coronavirus-reflect.

13. COVID-19 research checks could deter global collaboration. *Retrieved July 4, 2020, from*; 2020. https://www.universityworldnews.com/post.php?story=20200415141352492.

14. Why America's coronavirus testing problem is so difficult to solve. *Vox*; 2020. Retrieved July 4, 2020, from https://www.vox.com/recode/2020/4/24/21229774/coronavirus-covid-19-testing-social-distancing.

15. China says it backs WHO in tracing COVID-19, denounces U.S. "lies". *Reuters*; 2020. Retrieved July 4, 2020, from https://www.reuters.com/article/us-health-coronavirus-china/china-says-it-backs-who-in-tracing-covid-19-denounces-u-s-lies-idUSKBN22J10G.

16. Coronavirus in Guatemala: 38 deaths per covid-19 had not been recorded due to a "technical failure". *Free Press*; 2020. Retrieved July 4, 2020, from https://www.prensalibre.com/guatemala/comunitario/coronavirus-en-guatemala-38-fallecidos-por-covid-19-no-habian-sido-registrados-debido-a-una-falla-tecnica/.

17. Corruption and the Coronavirus. *News*; 2020. Transparency.org. Retrieved July 4, 2020, from https://www.transparency.org/en/news/corruption-and-the-coronavirus#.

18. Ludovic J, Bourdin S, Nadou F, Noiret G. Economic globalization and the COVID-19 pandemic: global spread and inequalities. *Bull World Health Organ*. 2020. https://doi.org/10.2471/BLT.20.261099. E-pub (April).

19. Goldacre B, Harrison S, Mahtani KR, Heneghan C. *WHO consultation on Data and Results Sharing During Public Health Emergencies. Background Briefing. Sep 2015 Centre for Evidence-Based Medicine Background Briefing: WHO Consultation on Data and Results Sharing During Public Health Emergencies Background Briefing for WHO consultation on Data and Results Sharing During Public Health Emergencies*; 2015.

20. WTO - WHO coordinate action on COVID-19 pandemic - Caribbean News Global. Retrieved July 4, 2020, from. (n.d.) https://www.caribbeannewsglobal.com/wto-who-coordinate-action-on-covid-19-pandemic/.

21. Department of Health. *7.5 What data should be collected during the early stages of a pandemic?*; 2006. Retrieved July 4, 2020, from https://www1.health.gov.au/internet/publications/publishing.nsf/Content/mathematical-models~mathematical-models-discussion.htm~mathematical-models-7.5.htm.

22. Better COVID-19 data without universal testing ImmunizationEconomics.org. (2020). Retrieved July 4, 2020, from http://immunizationeconomics.org/recent-activity/2020/4/20/better-covid-19-infection-data-without-universal-testing.

23. London JW. *Covid-19: Leading statistician slams UK's reporting of swab tests as "travesty of science"*; 2020. https://doi.org/10.1136/bmj.m1664.

24. Salerno M, Sessa F, Piscopo A, et al. No autopsies on COVID-19 deaths: a missed opportunity and the lockdown of science. *J Clin Med*. 2020;9(5):1472. https://doi.org/10.3390/jcm9051472.

25. Analysis of COVID-19 autopsies reveals many new details about this disease. (n.d.). Retrieved July 4, 2020, from https://medicalxpress.com/news/2020-05-analysis-covid-autopsies-reveals-disease.html.

26. Trump claims to have evidence coronavirus started in Chinese lab but offers no details. *US news | The Guardian*; 2020. Retrieved July 4, 2020, from https://www.theguardian.com/us-news/2020/apr/30/donald-trump-coronavirus-chinese-lab-claim.

27. US urges China's "full transparency" on COVID-19. *Retrieved July 4, 2020, from*; 2020. https://www.aa.com.tr/en/americas/us-urges-china-s-full-transparency-on-covid-19/1806425.

28. They report constant discrepancies in COVID-19 information. *The Hour*; 2020. Retrieved July 4, 2020, from https://lahora.gt/senalan-discrepancias-constantes-en-informacion-de-covid-19/.