Unprecedented situation, unprecedented official data and unprecedented quality of official data

**Keywords:** Data, COVID-19, Official Statistics, Random Sample, Data Quality

*Tracks and topics: Producing Official Statistics in Emergency Situations: Reflections and challenges posed by the Covid-19 pandemic*

# Introduction

The COVID-19 outbreak dominated the beginning of 2020. Almost every country and every socio-economic sector has been facing this unique situation. Official statistics need to confront new challenges, both internally and externally. Internally, every national statistics office needs to protect their workers, reorganize their way of working, and ensure the regular data production. Externally, National statistics offices are called to make a statistical description of an unprecedented complex reality.

The aim of a national statistics office is not only to produce official data but also to provide a realistic ‘picture’ of our world, even in time of crisis. However, an official statistics system should not limit itself to those standard tasks. National statistics offices should provide and participate to the production of data and information on COVID-19. For instance, the United Nations Statistics Division is actively recommending guidelines, sharing experiences and collecting data for COVID-19 (United Nations Statistics Division 2020\*). However, under the threat of COVID-19 many countries are considering official statistics as a non-essential service (Cheung 2020a). Are official statistics a non-essential service?

The aim of this letter is: (i) to analyze the role of national statistics offices during the COVID-19 health crisis, (ii) to summarize challenges and opportunities lying ahead, (iii) to trigger the discussion about national statistics offices.

For the sake of simplicity, let us distinguish between two different temporal phases of the COVID-19 epidemic. The first phase is predominated by health crisis and lockdown. This phase affects the population. The second phase starts when the health crisis is under control. Once the lockdown has been revoked, and the country is facing a new socio-economic situation that was completely unpredictable in January 2020. This last phase is full of challenges and opportunities for national statistics offices.

During lockdown, citizens were reading or listening to the figures of new COVID-19 cases and daily deaths. These numbers were unreliable, not harmonized and scarce. They did not provide a secure starting point to understand the situation, nor did they help take proper decisions. What drove the creation of these figures - cases and deaths of COVID-19 – was a medical objective. These figures were necessary to diagnose and to take care of patients. Nevertheless, we also need figures to better understand the spread of COVID-19. We need figures to implement the lockdown only when and where lockdowns are necessary. After the health crisis, we will need statistics to help us fully understand the new socio-economic situation and the details of the economic recession. We need reliable figures to plan for the future.

**1. Official Statistics during the Health Crisis**

Under the threat of COVID-19, the priority of national statistics offices is the health of its workers. The second priority is to describe the country’s situation based on statistical methodology and data collection. Policymakers and citizens need official statistics to make informed choices to manage the health crisis.

In the media, two figures are predominant: number of deaths and cases of COVID-19. From the point of view of a national statistics office, this data is produced by the health system and it is administrative data. We can call these figures “medical figures”. We can consider the production of this data as a side effect of the excellent work conducted by doctors and nurses who cure COVID-19 patients. The aim of medical figures is not to describe the spread of COVID-19 or the overall socio-economic implications of the outbreak. Unfortunately, figures produced by health authorities cannot give crucial information. We do need to know how, when and why the COVID-19 spread among the population. For instance, we need to know how many asymptomatic COVID-19 people have been in our country. Only a scientific random sample could give us this information (see Ioannidis 2020, Di Gennaro 2020 and Alleva et al. 2020).

Moreover, since medical figures - the number of deaths and cases of COVID-19 - were not produced by or on behalf of the national statistics office, the national statistics office cannot assess the quality of this data. While not referring to COVID-19, Radermacher (2020) has proposed that the statistical system could take on the assessment, management and certification of data. For instance, the German Federal Statistical Office\* declared that it does not collect real-time data on the outbreak. When looking for data on COVID-19, citizens are redirected to a different webpage. Interestingly, COVID-19 is not present in the webpages of several national statistics offices (Figure 2 in Misra et al. 2020).

An important issue about the number of deaths of COVID-19 is the interpretation of the main cause of death. Under this perspective, figures of COVID-19 deaths are problematic. To make an analogy, if you have the flu and a train runs you over, you would have died because of the train not because of the flu. The number of COVID-19 deaths should only indicate the number of people’s deaths caused by COVID-19 as main cause of death. Instead, different countries apply different definitions and different criteria. Often various regions within a same country apply different criteria. Normally, the national statistical systems work on metadata and harmonized definitions. Italy, Ecuador and the United States of America (CDC 2020) report as COVID-19 deaths in which the main cause of death is not COVID-19. The media compares deaths in 2020 with previous years’ deaths. This is unreliable data, as demographers well know, to have a comprehensive analysis of this data we should include an analysis of age groups.

If you do not take a blood test, then you do not have high cholesterol. The cases of COVID-19 depend on how many people were tested in each country and also how they were tested. Availability of testing and different types of tests should reflect in the statistics. Likewise, data should include people in hospitals, asymptomatic cases and people who came in contact with infected people. The number of COVID-19 cases we have today is not a reliable measure of how many people are infected by the disease. These numbers are useless both for comparison among countries and over time. The figures of COVID-19 cases we have are an underestimation of realistic figures.

Because of this underestimation of COVID-19 cases, in the media, we read systematic overestimations of the Case Fatality Rate (CFR). Since the majority of countries do not test all of the population, figures are greatly biased. If you cannot measure the main variable of the health crisis, you cannot manage the crisis and its implications.

Due to their data-collection capacity, infrastructure and experience, national statistics offices could support and help health authorities by producing reliable data. Statistical systems should provide data-collection support, quality control of figures and appropriate communication of statistics. The national statistics offices could be matching COVID-19 cases with socioeconomic aspects (like gender, age, income, etc.), previous medical problems, address (GPS), etc. Having those multivariate data could give the possibility to use more sophisticate statistical models. Statistical systems are key to implement real-time standardized reporting of the results and disaggregated data, and thus help assess the implications of COVID-19.

How can we acquire information on COVID-19? Up to the best knowledge we have, there are only two possibilities to obtain this information: either a census of the population or a random sample representative of the population. In most countries, a census is not practicable. We need a random sample representation of the population. Different sample designs and different possibilities can be implemented. Every person in the sample who is tested for COVID-19 also needs to answer a questionnaire. The questionnaire would include questions about the clinical evaluation, socio-demographic characteristics, personal characteristics, housing characteristics, and lifestyle of the individual.

Who can reach more information on COVID-19? Normally, national statistics offices have some health statistics experts, but they are not specialized in health statistics, and they should not improvise. At the same time, national statistics offices are expert institutions on population, sampling and data collection. They are in charge of census, employment survey as well as many other surveys. The collaboration between national statistics offices and National Health Systems at the national level could guarantee the necessary expertise to implement a random sample of the population under the threat of COVID-19. Due to the preparation of the Population and Housing Census, every national statistics office is ready to investigate its own population. Between the years 2020 and 2021, almost all national statistics offices in the world will implement the Census of Population and Housing. It is likely that almost all national statistics offices already have the master sample ready to prepare the census. This could serve as a starting point to implement a random sample. Between 2005 and 2014, more than six billion people around the world - more than 90% of the world’s population - were enumerated by population censuses. Only 21 countries did not conduct a census (United Nations Fund for Population Activities 2016).

…

To recapitulate, the national statistics offices during the crisis could:

* Help with the data collection regarding the crisis.
* Explain, manage and/or certify the data.
* Support a random sample representative of the population to identify the spread of the outbreak.

**2. Official Statistics after the Health Crisis**

# ….

# References

Alleva, G, Arbia, G., Falorsi, P. D.and Zuliani, A. 2020. “A sample approach to the estimation of the critical parameters of the SARS-CoV-2 epidemics: an operational design with a focus on the Italian health system”. Research Gate. Available at: <https://www.researchgate.net/publication/340514422_A_sample_approach_to_the_estimation_of_the_critical_parameters_of_the_SARS-CoV-2_epidemics_an_operational_design_with_a_focus_on_the_Italian_health_system>

Bruengger, Heinrich, 2008. “How Should a Modern National System of Official Statistics Look?”. UNECE, Statistical Division.

Cheung, Paul. 2020a. “Impact of COVID-19 on Official Statistics (2) - Is Official Statistics Non-Essential Service”. Available at: <https://www.linkedin.com/pulse/impact-covid-19-official-statistics-2-non-essential-paul-cheung/?trackingId=AWBHyvCKQqWOsV4f98T22Q%3D%3D>

Cheung, Paul. 2020b. “Impact of COVID-19 on Official Statistics.” Available at: <https://www.linkedin.com/pulse/impact-covid-19-official-statistics-paul-cheung/>

Di Gennaro Splendore, Luca. 2020. “Random testing, quality of data and lack of information: COVID-19”. Available at: <https://medium.com/data-policy/random-testing-quality-of-data-and-lack-of-information-covid-19-a6e09a398d1d>

Eurostat. 2020. Website. *COVID-19: support for statisticians*.Available at: [*https://ec.europa.eu/eurostat/data/metadata/covid-19-support-for-statisticians*](https://ec.europa.eu/eurostat/data/metadata/covid-19-support-for-statisticians)

IAOS, International Association for Official Statistics. (2020). *Official Statistics in the context of the COVID-19 crisis*. Website. Available at: <https://officialstatistics.com/news-blog/crises-politics-and-statistics>

Ioannidis, J. 2020, “A fiasco in the making? As the coronavirus pandemic takes hold, we are making decisions without reliable data”. Available at: <https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/>

Jeskanen-Sundström, Heli (2007). “Independence of Official Statistics, a Finnish Experience”. Seminar on Evolution of National Statistical Systems.

Misra, A., Schmidt, J. and Harrison, L. 2020. *Combating COVID-19 with Data: What Role for National Statistical Systems?* PARIS 21 - New Policy Brief.

Radermacher, Walter J. 2020. *Official Statistics 4.0*. Springer.

Tam, S. and Kim, J. 2018. “Big Data ethics and selection-bias: An official statistician’s perspective”. Statistical Journal of the IAOS, 34(4), pp.577-588.

United Nations. 1994. *Nations Fundamental Principles of Official Statistics*.

United Nations Fund for Population Activities (2016). *Annual Report*.

United Nations Statistics Division. 2020. Website. *COVID-19 Response – Resources for Official Statisticians*. Available at: <https://covid-19-response.unstatshub.org/>