Letter to the Editors

Defeating the Pandemic Requires High Quality and Ethical Official Statistics

Knowing the truth about the pandemic means that we know, for example, how many are currently infected, how many have been infected in the past and recovered, and how many have died from COVID-19. This information is essential for addressing the pandemic and its social and economic effects with appropriate government policies, rigorous work by the scientific community, and proper actions by the population, civil society and the business sector of any country. It is also essential for effective international cooperation, which is absolutely key to addressing the pandemic and its effects.

Getting as close as possible to the truth about the pandemic requires objective and impartial provision of data on COVID-19 that follows the highest quality principles in the collection, processing and dissemination of these data. A priori, there is no better candidate than official statistics to provide such data, and show us all the most accurate possible, unadulterated picture of the pandemic.

Official statistics are statistics that are developed, produced and disseminated as a public good by the National Statistics Office (NSO) and other agencies clearly identified in law as part of the National Statistical System and that aim to serve the statistical information needs of the entire society. Official statistics are provided regularly to the various branches of government, the state administration, the citizenry, the general public, the markets, the research community and the international partners of the country, as well as the international organizations of which the country is a member. Official statistics are to comply with international and national compilation standards and statistical ethics. These statistics should be our main source of data on the current pandemic, as well as other epidemics and health conditions.

This preference for official statistics as the source of truth about the pandemic is present because official statistics are a *global* public good (Georgiou 2017), providing the same information simultaneously to all users anywhere in the world. This global public good (official statistics) is needed to assist in the production of another global public good – the fight against epidemic diseases, whose urgency has been in bold relief in 2020. In addition, and very importantly, official statistics are supposed to be produced and disseminated on the basis of international *statistical principles and ethics*. These are enshrined in codes of practice and lists of fundamental principles adopted by the United Nations, and form part of national and regional legal frameworks. If – and this is an important ‘if’ – official statistics on the pandemic are produced and disseminated on the basis of international statistical principles and ethics that ensure the integrity and high quality of these statistics, it will bring us as close as possible to the truth about the pandemic.

Acknowledgments: The author is grateful to Hallgrímur Snorrason, Jean-Michel Durr and Gerry O Hanlon for comments on an early draft. The views expressed herein and any errors are those of the author.

So, official statistics have, *in principle*, a deontological, legal and institutional ‘head-start’ as a candidate in the production of statistics on the pandemic, that is, statistics on the incidence, prevalence and mortality of the pandemic. To ensure that they *in practice* end up living up to expectations about quality and ethics in this important task, such official statistics *must* adhere to the following principles (which, *inter alia*, draw on codifications of principles of official statistics, such as the European Statistics Code of Practice (Eurostat 2017) and the UN Fundamental Principles of Official Statistics (UNSD 2013)):

- There must be international standards and guidelines on the production and dissemination of pandemic statistics and these must be widely publicized and readily accessible to the public. The international health and epidemiology *statistics* community coordinated by World Health Organization *statisticians* must see to that,
- National official statistics on COVID-19 should strictly conform to existing international statistical standards and guidelines, by being subject to procedures to ensure that existing standard concepts, definitions, and classifications are consistently applied,
- Timeliness and periodicity of official statistics on COVID-19 should meet appropriately ambitious international release standards,
- Detailed information (metadata) about the statistical processes and primary data sources used, as well as about the statistical results on COVID-19, should be produced and disseminated and easily available to the public,
- All statistical results on COVID-19, as well as their source data and statistical processes, should be regularly assessed and validated by dedicated procedures in the official statistics agency producing them, and their quality should be regularly and openly reported to the public according to established quality criteria,
- The statistics departments of international and regional organizations dedicated to the fight against pandemics (or, alternatively, international and regional statistical organizations using relevant expertise) should regularly and rigorously assess the quality of national official statistics on COVID-19 and publish these assessments,
- The statistics departments of international and regional organizations dedicated to the fight against pandemics should reproduce, in their various publications, the national official statistics on COVID-19 (only if they can do so) with full metadata and quality information,
- COVID-19 statistics should be comparable over time and across countries. These statistics and the corresponding metadata should be presented in a form that facilitates proper interpretation and meaningful comparisons across time and across countries and regions,
- All errors discovered in published COVID-19 statistics should be corrected at the earliest possible date and publicized,
- Advance and widely publicized notice should be given regarding major revisions or changes in the methodology for producing and releasing COVID-19 official statistics. Revisions of COVID-19 data should follow standard and transparent procedures,
- Individual data (i.e., of statistical units) collected by official statistics agencies on cases of COVID-19 should be kept strictly confidential and used *exclusively* for statistical purposes, with no exceptions,

- COVID-19 official statistics should be compiled on an objective basis determined *solely* by statistical considerations. So, for example, choices of data sources and statistical methods, as well as decisions on the dissemination of these statistics, should be based solely on statistical considerations,
- All users should have equal access, at the same time, to statistical releases on COVID-19, and statistical releases and statements made in press conferences should be objective and non-partisan, and
- Finally, official statistics producers should independently—but within the confines of international standards and guidelines—decide on, and be accountable for, the methods, standards and procedures to be used in the production of COVID-19 statistics and on the time and content of statistical releases on COVID-19.

We do not aim here to assess whether any official statistics on the pandemic that may already be produced conform to the above principles. If they are going to be official statistics compiled with high quality and unimpeachable ethics, in our view they will *need* to conform to these principles. Similarly, we do not intend to pass judgment on whether international, regional and national agencies responsible for COVID-19 statistics – regarding common standards and guidelines, production, dissemination – have performed according to what is appropriate, as outlined above. They *ought* to live up to these principles. We invite observers to examine and question any available official statistics that are currently produced or may be produced in the future as to whether they meet the above list of principles and make up their own mind about the trust to be accorded to the statistics. Here, we offer only the cautionary remark that there may be room for improvement in various cases.

Political leaders and parties in office (at national or local levels) facing reelection, governments trying to attract foreign visitors, and authoritarian regimes in power needing to legitimize their hold on power might all be interested in COVID-19 statistics showing that they have ‘conquered’ the pandemic in their own jurisdictions with their epidemic mitigation efforts. At the same time, opposition politicians might also want to feed their own political narratives of alleged incompetence and negligence of the political party in power in handling the pandemic. Powerful business interests might also have ‘preferred’ narratives to be potentially supported by COVID-19 statistics. Even governments and pharmaceutical companies associated with certain therapeutic drug and vaccination initiatives may have pictures of reality they would prefer to see emerge from COVID-19 statistics. Thus, all or some of the above actors may be tempted to try to influence the production of statistics regarding the prevalence and mortality of the pandemic in their countries’ and jurisdictions’ populations. Or, at least, there may be a perception among the public that such influence is taking place. Thus, it is truly imperative that official statistics compiled on COVID-19 meet the above principles regarding ethics and quality. In this way, being of unimpeachable integrity and high reliability, these trusted statistics can serve as a bulwark against politically serving ‘fake news’ and ‘alternative facts’.

It should be noted that to respond adequately to the pandemic, the statistics on the spread and mortality of the disease itself referred to above would have to be supplemented by other statistics on the many different effects of the pandemic on social, economic, environmental and other health conditions in society. These statistics, whether well-

established statistics or new flash estimates and experimental statistics, should adhere to similar principles as those outlined above.

The COVID-19 pandemic brings forth the need to think in different ways about some statistical methods and processes in identifying cases of infection with the virus. Up to now, the cases identified in each country and jurisdiction are those identified (confirmed) by medical professionals when they encounter a patient whom they assess one way or another to have the COVID-19 infection. These medical front line personnel are then supposed to report these findings up through the administrative (usually health) apparatus of the country or other administrative jurisdiction. This makes sense if the statistics to be compiled are reporting ‘identified cases of COVID-19’, but it does not make sense if one wants to know the total number of infected persons in a country or jurisdiction within it. In this sense, we agree with the recent letter to the JOS editors (Di Gennaro Splendore 2020, 230), which states that “figures produced by the health authorities cannot provide crucial information”.

Certainly, some would argue that the present practices are good enough. We believe this is not true. To see that there is a need for official statistics on the pandemic to be produced in a different way, one can consider the following: the number of cases of COVID-19 identified (confirmed) by health authorities could give a hint about the magnitude of the problem of infection, but undoubtedly that number is a function of the number of examinations/tests that are carried out by the health system of the country or jurisdiction. One can think of situations in which test availability is intentionally or unintentionally limited, or the medical system for various reasons does not report all identified COVID-19 cases, or the population and especially certain groups in it are discouraged by culture, economics or politics, or even physical incapacity, from seeking a test or examination by a health professional when they have COVID-19-like symptoms. In addition, it is well-known that many individuals infected with COVID-19 are fully asymptomatic and are likely to never present themselves for testing or examination. In addition, testing for COVID-19 most often takes place at the initiative of individuals or businesses for their own information, or of authorities targeting certain localities/demographics. Thus, results of tests and examinations for COVID-19, as they are carried out and reported now, extrapolated to the entire population, would likely give rise to *biased* estimates of the overall incidence of COVID-19 infection in the population, as well as in certain demographic groups. Thus, the number of identified COVID-19 cases and its change over time does not provide reliable information from which to adequately estimate either the true number of currently infected individuals in the population (and the true number of new infections in a specific time period) or the true number of those that have had COVID-19 and have since recovered.

Persisting with identified (confirmed) cases of infection as the main indicator of the spread of the pandemic also sets the statistics up for criticism of credibility and for exploitation in the successful peddling of narratives that may be driven by political and economic interests, rather than by a desire to find the truth and take action on that basis.

What is needed, instead, is *an expansion of the usual tools of official statistics to include testing of biological samples from respondents in the context of a statistical survey aimed at inferring the prevalence of COVID-19 in the population*. Specifically, there is a need for proper stratified multistage sampling, on a very regularly recurring basis, of the population

of the country, so that it could be tested for current COVID-19 infection (via testing for the virus) and for past infection (via testing for antibodies for the virus). Then, given appropriate size sampling and thoughtful sample design, statistical estimates of (inferences about) current and past infection of the entire population of a country and for various subpopulations can be generated. Subpopulations could consist of subnational jurisdictions (states, provinces, counties, etc.) and of specific demographic groups (by sex, age, ethnic background, etc.). The sample design of the statistical survey would have to be carefully configured to collect samples of what might be a relatively small – compared to the total – population of currently or previously infected individuals occurring in clusters.

The essential feature in what we propose is that the official statistics survey is not conducted simply by asking the respondents questions about their health status and medically related behavior and social conditions, although that should also be part of the survey, but that the respondents provide to appropriately trained survey field interviewers biomarker samples from their respiratory system (e.g., via swabbing) and blood for testing. The use of biomarkers in surveys is not new. For example, the Demographic and Health Surveys (DHS) Program has been collecting biomarker samples from interviewees in a number of surveys around the world; DHS surveys in over 50 countries have included about 20 biomarkers, relating to “a wide range of health conditions, including infectious and sexually transmitted infections, chronic illnesses (such as diabetes, micronutrient deficiencies), and exposure to environmental toxins” ([DHS Program 2020](#)).

The official statistics survey on COVID-19 we are advocating incorporates the feature of biomarker testing in appropriately sampling the population. It is important to highlight that, with this approach, there is no need to test everyone in the population to have an unbiased estimate with adequate precision (assuming an appropriate size sample) of the total number of people currently infected and the total number of people who have been infected so far. Such a survey would not be very cheap, especially as it would need to be undertaken on a very regular basis. However, its cost would be microscopic compared with the unnecessary costs – in terms of human lives lost and economic and social costs – of a pandemic, on account of not knowing the basic facts the proposed statistics would divulge to us. Policy makers should not hesitate to provide the resources needed to official statistics producers.

The proposal for such a methodology for producing COVID-19 statistics may strike some as somewhat intrusive and may raise the prospect of risks to privacy. This is actually true. This is an additional reason why statistics about COVID-19 should be compiled by official statistics producers fully committed to statistical principles and ethics and, specifically, to the principle of statistical confidentiality. As the UN Fundamental Principles of Official Statistics require: “Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes” (UNSD 2013). This means that if a respondent in the survey is found to be positive for COVID-19, the information about the specific person will definitely *not* be made available to health officials, the police, administrators involved in tracking cases or enforcing isolation, or any other public agency or private entity. (The information would be made available by the official statistics producer only to the person undertaking the testing/survey.) This approach would help ensure an adequate response rate to the survey and its success in generating unbiased

estimates, not only the first time it is carried out but also when it is carried out periodically, as it should be.

Statistical estimates of the total number of people currently infected by COVID-19 and the total number of people who have been infected so far would have to be supplemented by reliable and timely vital statistics. Vital statistics about deaths from COVID-19 infection would be particularly important. In general, it would be important for statistics about deaths (from any cause) to be highly reliable and timely, as they would offer a reasonable indirect measure of the toll of the pandemic in the form of ‘excess deaths’ measures. Such vital statistics should also be produced as official statistics (notwithstanding that the primary source would be administrative records) adhering to the principles discussed earlier in this letter. These pieces of information would then allow health scientists to have in different points in time reliable estimates of epidemiological variables, such as the susceptible part of the population, the infected part of the population and the removed part of the population (recovered or diseased). (A description of such components of the SIR (susceptibles, infected, and removed) epidemiological model can be found in [Ball 2020](#).) This information would help estimate the degree of infectiousness and the infected fatality rate, which are important components of epidemiological models. Thus, reliable official statistics would greatly help health scientists calibrate (repeatedly over time) their models of the pandemic and, thus, improve their projections about the development of the pandemic.

The idea of the need for random sampling to understand what is happening in the population regarding the pandemic has also been advanced in a previous letter to the JOS editors ([Di Gennaro Splendore 2020](#)) and by a number of other authors elsewhere (for example, [Alleva et al. 2020](#); [Cochran 2020](#); [Cook and Gray 2020](#); [Trewin 2020](#)). In this letter, our general argument for sampling the population using biomarker testing to generate statistical estimates of current and past infection of the entire population is in congruence with other authors’ ideas.

In this letter, we also specifically and explicitly argue that in order for any (such) statistical estimates on the pandemic to be of adequate quality, they have to be produced as official statistics and not, in any sense, as byproducts of public health policy making. This means they can be produced by the NSO or another official statistics producer that operates in full conformity of statistical principles such as the ones outlined earlier, but they should not be produced by an institution that does not meet those principles, nor should they be produced by some mixed team of institutions/officials, of which some are not official statistics producers.

In this sense, we would argue against any proposals or statements that could be interpreted as accepting of a merging of the efforts of national statistical offices and of public health policy institutions. In such usually well-meaning visions of official statisticians ‘working together’ with or under officials from the public health policy/national health system side lurk significant risks to the integrity and quality of the resulting official statistics ([Georgiou 2019](#)), as some important principles, from the list provided in the beginning of this letter, would be at risk of not being met.

This does not mean that official statistics producers of COVID-19 statistics would not utilize knowledge and expertise that may exist outside their institutions; they ought to, but it would have to be done on the terms of the official statistics producer acting in

independence and fully observing the statistical principles and ethics noted in this letter. In addition, our proposed approach does not mean that official statistics producers would not communicate with users of their statistics about their findings; they would, but do so fully adhering to the statistical principles above, and in particular those concerning impartiality and objectivity, as well as statistical confidentiality. There should be an exchange of knowledge and communication between official statisticians on the one hand and health scientists, providers and policy officials on the other hand, but their roles would have to be kept clear and distinct.

Thus, our proposed approach explicitly and unambiguously *disengages* the health provision, public health monitoring and policy response work on the pandemic from official statistical work on the pandemic. In order to ensure that statistics on the COVID-19 pandemic – and any future epidemic – do not fall prey to actual or suspected *conflicts of interest* or engender risks to *privacy*, the production of these statistics should be left to independent official statistics producers without the involvement of other interested parties.

Generating statistics on the pandemic in the way described above is, we believe, the best way to serve the work of health providers, public health officials, and policy makers alike. It is also the best way to serve the interests of society as a whole and of the global community.

Such official statistics can provide reliable information on the extent and the rate of the spread of the virus in a country (as well as in its individual geographic or administrative subdivisions and in its demographic groups), the true mortality rate, the extent to which herd immunity might be developing, and more. This is, of course, indispensable evidence on which to responsibly and rationally base all decisions, not only on public health policy, but also on economic policy, a multitude of social policies, national security and foreign policy, as well as international cooperation and coordination. And the same holds for decisions that have to be made by scientific researchers, families and individuals, businesses and the markets. Finally, this evidence is necessary to keep the policy makers and leaders in charge accountable during the challenging time of the pandemic. How else would one judge whether the policy responses are appropriate and successful?

In conclusion, official statistics need to ‘step up to the plate’ all around the world to quickly defeat the pandemic for good and minimize loss of life and economic and social distress. To do so, official statistics on COVID-19 must be produced with high quality, which requires innovation as well as the application of official statistics principles. Such approaches should be adopted not only for the current pandemic, but be part of the ‘armor’ in place to protect against similar attacks in the future.

References

- Alleva, G., G. Arbia, P.D. Falorsi, G. Pellegrini, and A. Zuliani. 2020. “A sample design for reliable estimates of the SARS-CoV-2 epidemic’s parameters. Calling for a protocol using panel data.” Available at: <https://web.uniroma1.it/memotef/sites/default/files/Proposal.pdf> (accessed October 2020).
- Ball, P. 2020. “How do epidemiologists know how many people will get Covid-19?” *Significance*. Available at: <https://www.significancemagazine.com/science/648-how->

- do-epidemiologists-know-how-many-people-will-get-covid-19 (accessed October 2020).
- Cochran, J. 2020. “Why we need more coronavirus tests then we think we need.” *Significance*. Available at: <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/1740-9713.01398> (accessed October 2020).
- Cook, L., and A. Gray. 2020. “Official statistics in the search for solutions for living with COVID-19 and its consequences.” *Statistical Journal of the IAOS* 36 (2): 253–278. DOI: 10.3233/SJI-200671. Available at: <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji200671> (accessed October 2020).
- DHS Program. 2020. Website. Available at: <https://dhsprogram.com/What-We-Do/Bio-markers.cfm> (accessed October 2020).
- Di Gennaro Splendore, L. 2020. “COVID-19: Unprecedented Situation, Unprecedented Official Statistics.” *Journal of Official Statistics* 36 (2): 229–235. DOI: <https://doi.org/10.2478/jos-2020-0012..>
- Eurostat. 2017. “Eurostat European Statistics Code of Practice”. Website. Available at: <https://ec.europa.eu/eurostat/web/products-catalogues/-/KS-02-18-142> (accessed October 2020).
- Georgiou, A. 2017. “Towards a global system of monitoring the implementation of UN fundamental principles in national official statistics.” *Statistical Journal of the IAOS* 33 (2): 387–397. DOI: <https://doi.org/10.3233/sji-160335>.
- Georgiou, A. 2019. “Extracting statistical offices from policy-making bodies to buttress official statistical production.” *Journal of Official Statistics* 35 (1): 1–8. DOI: <https://doi.org/10.2478/jos-2019-0001>.
- Trewin, D. 2020. “Using Random Sampling to Learn About Covid-19 Known Unknowns.” Comment in the discussion: Official Statistics in the context of the COVID-19 crisis. *Statistical Journal of the IAOS*, discussion platform. Available at: <https://officialstatistics.com/news-blog/crises-politics-and-statistics> (accessed 29 October 2020).
- UNSD. 2013. “Fundamental Principles of Official Statistics”. United Nations Statistics Division (UNSD) website. Available at: <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx> (accessed October 2020).

Andreas V Georgiou

Visiting Lecturer and Visiting Scholar, Amherst College,
Former President (2010–2015),
Hellenic Statistical Authority, Greece
Amherst College, Converse Hall,
220 South Pleasant Street, Amherst, U.S.A.
Email: avgeorgiou83@amherst.edu