

The Scenario of a Pandemic Spread of the Coronavirus SARS-CoV-2 is Based on a Statistical Fallacy

Christof Kuhbandner¹

¹University of Regensburg

April 20, 2020

Abstract

In view of the rapidly increasing numbers of reported new coronavirus infections, many speak of an upcoming pandemic. However, since the number of conducted coronavirus tests has rapidly increased over time as well, the apparent increase in infections may actually reflect increased testing, rather than a rapid spread of the coronavirus. To examine this issue, data from Austria, Belgium, France, Germany, Italy, and USA were analyzed. In all countries, the rapid increase in reported new infections was largely attributable to the rapid increase in conducted tests. Statistically controlling for the increased amount of testing revealed that the increases in reported infections dramatically overestimate the true increases in every country. According to the estimated true courses of new infections, the increases were initially much smaller, and the courses of new infections have already flattened or are even decreasing since the beginning of calendar week 13 (March 23) in almost all countries. The courses of reported new infections and deaths started to increase almost simultaneously in every country, which further confirms that the increases in reported new infections reflect effects of increased testing. These results indicate that the scenario of a coronavirus pandemic is based on a statistical fallacy.

Title: The Scenario of a Pandemic Spread of the Coronavirus SARS-CoV-2 is Based on a
Statistical Fallacy

Authors: Christof Kuhbandner^{1*}

¹Department of Human Sciences, University of Regensburg, 93053 Regensburg, Germany.

Abstract: In view of the rapidly increasing numbers of reported new coronavirus infections, many speak of an upcoming pandemic. However, since the number of conducted coronavirus tests has rapidly increased over time as well, the apparent increase in infections may actually reflect increased testing, rather than a rapid spread of the coronavirus. To examine this issue, data from Austria, Belgium, France, Germany, Italy, and USA were analyzed. In all countries, the rapid increase in reported new infections was largely attributable to the rapid increase in conducted tests. Statistically controlling for the increased amount of testing revealed that the increases in reported infections dramatically overestimate the true increases in every country. According to the estimated true courses of new infections, the increases were initially much smaller, and the courses of new infections have already flattened or are even decreasing since the beginning of calendar week 13 (March 23) in almost all countries. The courses of reported new infections and deaths started to increase almost simultaneously in every country, which further confirms that the increases in reported new infections reflect effects of increased testing. These results indicate that the scenario of a coronavirus pandemic is based on a statistical fallacy.

* Corresponding author.

Email: christof.kuhbandner@ur.de

Introduction

For weeks, people around the world have been looking at the apparently rapid spread of the coronavirus SARS-CoV-2. In view of the increasing numbers of daily new infections reported from many countries, experts, politicians, and the media speak of an upcoming pandemic with millions of infected people worldwide. In response to such horror scenarios, extreme fear is experienced at the individual level and draconian countermeasures have been adopted in many countries.

In the face of such dynamics, a fundamentally important question arises: Do the observed increases in the reported numbers of new infections really reflect what they seem to reflect at first glance – a true increase in the number of new infections? If looking more closely at the reported increasing numbers of new infections from a methodological perspective, one will notice that one important problem regarding the interpretation of such data has so far been neglected: that the number of tests carried out for the coronavirus has rapidly increased as well.

The fundamental problem is that if there are many infected people that are not detected because too few tests are conducted (i.e., unreported infections), which is assumed to be the case for coronavirus infections¹, the number of reported new infections depends on the number of conducted tests: when the number of tests is increased, the number of detected new infections will automatically increase as well because more hitherto unreported infections are detected. This introduces a potential statistical fallacy: An observed rapid increase in detected new infections may give the impression that there might be a rapid spread of a virus. However, the observed rapid increase actually may reflect the rapid increase in testing, and tell nothing about the true course of new infections, which may actually be much less steep or even decreasing.

The statistical fallacy can be illustrated by a simple example: Imagine there is a garden

where ten Easter eggs are hidden every day (i.e., the true number of new infections). On the first day, the children are allowed to search for one minute and they find one egg; on the second day, they are allowed to search for two minutes and they find two eggs; and on the third day, they are allowed to search for four minutes and they find four eggs (i.e., the number of reported new infections). The children could get the misleading impression that exponentially more Easter eggs are hidden in the garden every day because they find exponentially more eggs every day. But of course, this is a problematic interpretation because in reality there were always the same number of eggs hidden in the garden, and the increased number of eggs found is only due to the increased number of search attempts (i.e., the increase in the number of tests). As illustrated in Fig. 1 based on data from Italy and the USA²⁻⁴, regarding the reported numbers of new coronavirus infections, such problem indeed exists.

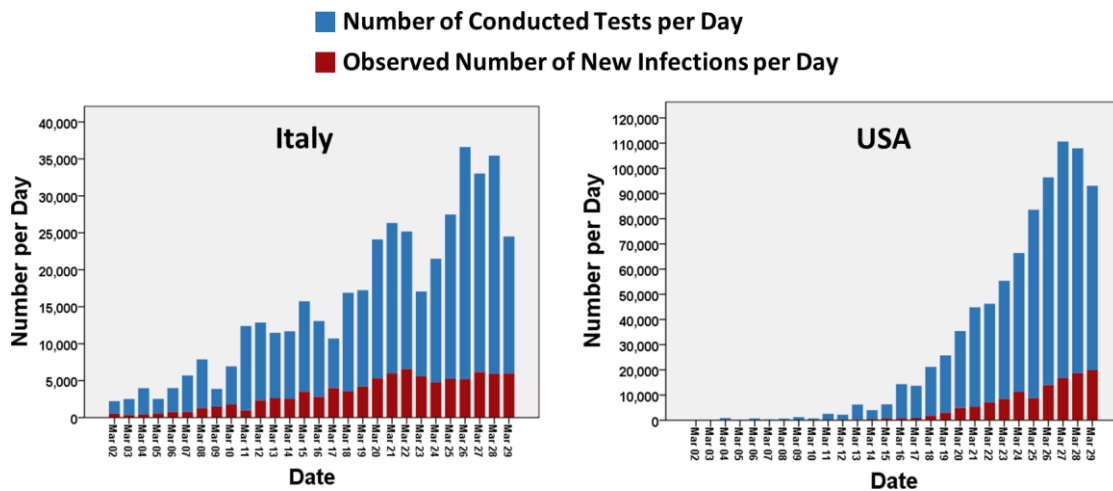


Fig. 1: Illustration of the relationship between the number of coronavirus tests and reported new coronavirus infections. The course of the number of conducted coronavirus tests (height of the blue bars) and the course of reported new coronavirus (height of the red bars) in Italy and the USA in calendar weeks 10-13 is shown (from March 2 to March 29).

As can be seen in Fig. 1, the number of reported new infections increases simultaneously

with the number of conducted tests. However, as illustrated by the Easter egg example, if there are unreported cases (in the Easter egg example: the hidden eggs that are not found due to too few search attempts), one will automatically find at least as many new infections as the number of tests has been increased (unless the true number of new infections is in reality decreasing). For example, if one runs twice as many tests, one will also find at least twice as many new infections. Consequently, if there were a true increase in new infections, one would have to find a larger increase in detected new infections than is caused solely by the increase in the number of tests. For instance, if the number of tests were doubled, one would have to find more than twice as many new infections if there were a true increase in new infections.

Thus, based on an analysis of the relationship between the increase in the number of tests and the concurrent increase in reported new infections, the question of whether the increase in reported new infections is prone to such a statistical fallacy can be examined: if the number of new infections is in reality increasing, the factor by which the reported new infections increase should be larger than the factor by which the number of tests is increased. If the number of new infections does in reality not change, the factor by which the reported new infections increase should mirror the factor by which the number of tests is increased. If the number of new infections is in reality decreasing, the factor by which the reported new infections increase should be smaller than the factor by which the number of tests is increased. The basic principle of the statistical fallacy is illustrated in Fig. 1a.

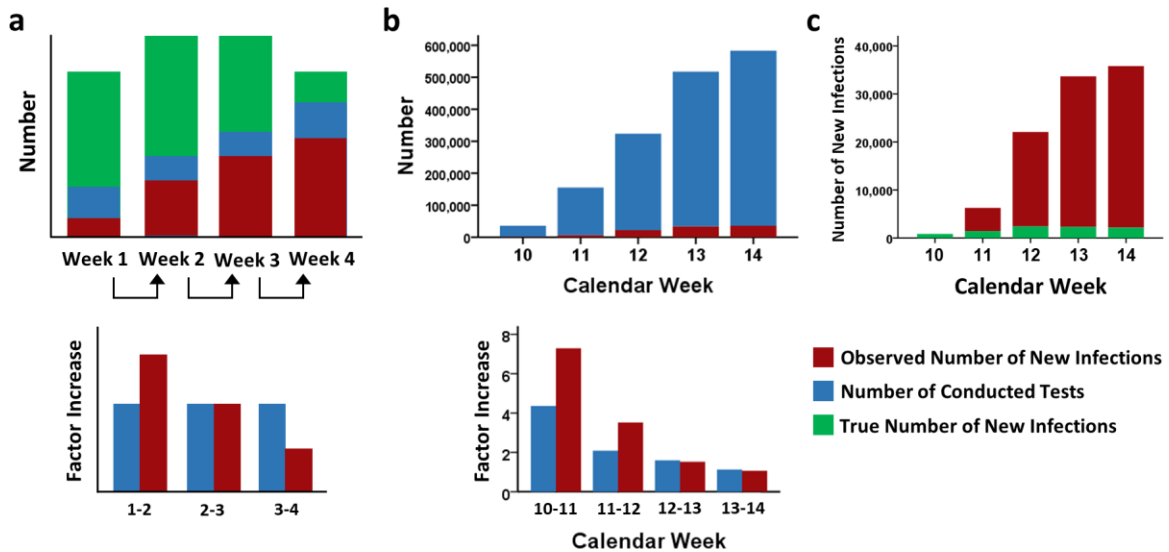


Fig. 2: Illustration of the statistical fallacy and the method of correction. As illustrated in the upper panel of (a), if there are many unreported infections (green bars), the number of reported new infections (red bars) is determined by the number of tests carried out (blue bars). If the number of tests increases over time, more new infections will be observed, although the true number of new infections may in reality be much less increasing (from week 1 to 2), not change (from week 2 to 3), or even decrease (from week 3 to 4). As shown in the lower panel of (a), whether an observed increase in reported new infections reflects a true increase beyond the test-number induced increase can be determined by a comparison of the factors by which the number of tests (blue bars) and the reported new infections (red bars) increase from week to week. As an example with real data, (b) shows the relationship between the number of conducted coronavirus tests (blue bars) and the number of reported new coronavirus infections (red bars) for Germany in calendar weeks 10 to 14 (upper panel), and the respective factors by which the numbers of conducted tests (blue bars) and reported new infections (red bars) increased from week to week (lower panel). (c) shows for the data from Germany the test-number biased course of reported new infections (red bars), and the course of new infections when statistically controlling for the increased amounts of testing (green bars). Note that for the purpose of visual comparison, the growth curves in (c) are scaled to the number of reported infections in calendar week 10. In reality, the true number of new infections is higher than the reported number of new infections due to the existence of unreported cases (see Fig. 1A).

As an example with real data, Fig. 1b shows the relationship between the number of conducted coronavirus tests and the number of reported new coronavirus infections for Germany^{5,6} in calendar weeks 10 to 14 (upper panel), and the respective factors by which the numbers of conducted tests and reported new infections increased from week to week (lower panel). As can be seen, the number of tests increased rapidly with time, indicating that large parts of the observed increase in reported new infections is attributable to increased testing. Examining the factors by which reported new infections and tests increased from week to week indicates that the number of new infections increased stronger than the number of conducted tests from calendar weeks 10 to 12, indicating that the number of new infections initially truly increased, albeit smaller than suggested by the reported number of new infections. However, from calendar week 12 on, the number of tests and the number of reported new infections increased simultaneously, indicating that the observed increases in reported new infections are fully attributable to increased testing, and tell nothing about the true course of the spreading of the virus.

In a situation where an increase in reported new infections does not necessarily tell something about the true course of new infections due to the fact that the number of tests has simultaneously increased as well, there is a simple statistical technique that can be used to estimate the true course of new infections: the observed numbers of reported new infections can be statistically controlled for the increase in conducted tests. The basic principle can be described as follows: how many new infections would have been observed if the number of tests would not have been increased across weeks? This can easily be estimated by dividing the weekly number of reported new infections by the factor by which the number of tests has been increased per week. Fig. 1c illustrates this for the data from Germany. As can be seen, the test-number biased

course of reported new infections dramatically overestimates the true course of new infections, as revealed by statistical control for the increase in test numbers. Contrary to what is suggested by the observed rapid increase in reported new infections, the number of new infections initially increased much less, and is actually decreasing since the beginning of calendar week 13.

Results

To examine whether the increases in reported new infections in other countries are prone to the same statistical fallacy as well, data from Austria, Belgium, Germany, France, Italy, and USA on the numbers of conducted coronavirus tests^{3-5,7-9} and reported new coronavirus infections^{2,6} in calendar weeks 10 to 14 (March 2 to April 5) were analyzed. To account for potential temporal variability in the timeline running up to a test being reported, both in terms of the time it takes for a symptomatic person to receive a test, and in the time for that test to get reported, and because for Germany and France only data on the number of conducted tests per week is available, data were aggregated by week.

Fig. 3 shows for each of the countries the relationship between the numbers of conducted tests and reported new infections (left panels), and the test-number biased courses of reported new coronavirus infections and the estimated true courses based on statistical control for the increased amount of testing (right panel). In all countries, the rapid increase in the number of new infections per week was largely attributable to the rapid increase in the number of conducted tests per week. Statistically controlling for the increased amount of testing consistently revealed that the observed rapid increases in reported new infections dramatically overestimate the true increases in every country. According to the estimated true course, the initial increases in new infections were much smaller, and in almost every country, the course of new infections has already flattened or is decreasing since about calendar week 13.

■ Number of Conducted Tests ■ Observed Number of New Infections ■ True Number of New Infections

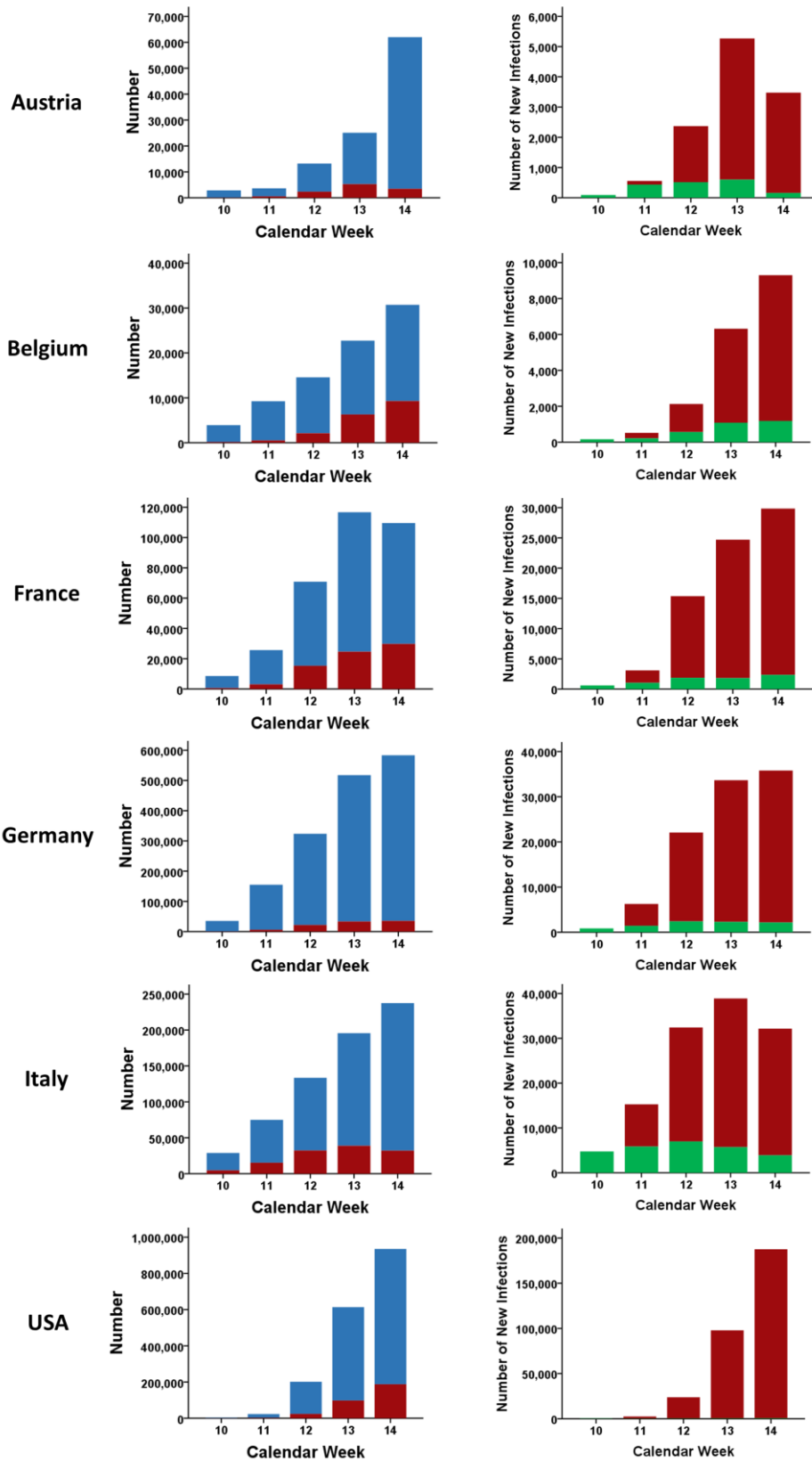


Fig. 3. Statistical fallacy in the countries Austria, Belgium, France, Germany, Italy, and USA. The left panels show the relationships between the number of conducted coronavirus tests and the number of reported new coronavirus infections in every country for calendar weeks 10 to 14 (March 2 to April 5). The right panels show for every country the test-number biased course of reported new infections, and the estimated true course based on statistical control for the increased amount of testing. Note that for the purpose of visual comparison, the growth curves are scaled to the number of reported infections in calendar week 10. In reality, the true number of new infections is higher than the reported number of new infections due to the existence of unreported cases (see Fig. 1A).

The previous analyses indicate that the observed rapid increases in new infections largely reflect the fact that the number of tests has been rapidly increased over time. To further examine this issue, the courses of reported new infections and reported deaths were compared for the six countries. To account for the much longer reporting lag for deaths (about up to two weeks in many countries, e.g.¹⁰), only data until March 28 were examined. Fig. 4 shows the courses of the daily increases in reported new infections and deaths. To enable a visual comparison, the values for new infections were scaled to the level of the number of deaths, based on the respective death rates in each country. Intriguingly, in every country, the numbers of reported new infections and deaths started to increase almost simultaneously. Correlation analyses revealed that the growth curves were highly related (Austria: $r = .83, p < .001$; Belgium: $r = .88, p < .001$; France: $r = .94, p < .001$; Germany: $r = .95, p < .001$; Italy: $r = .96, p < .001$; USA: $r = .95, p < .001$).

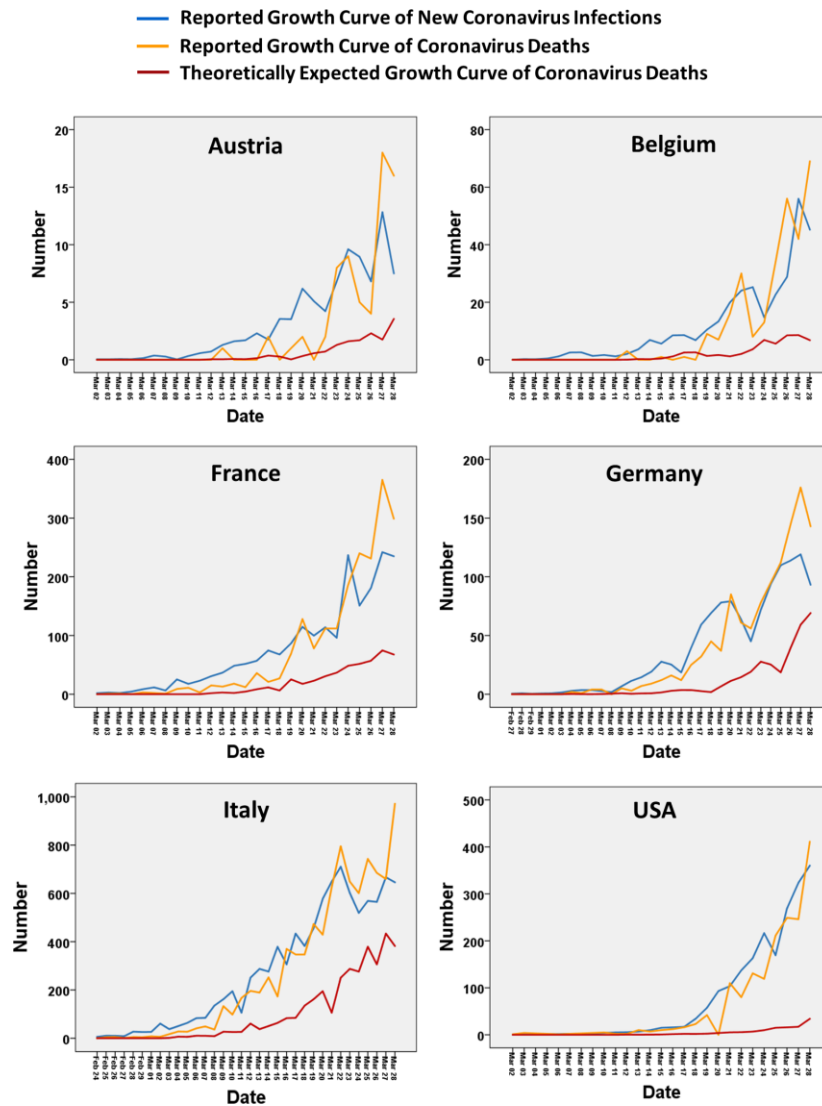


Fig. 4. Course of reported daily new coronavirus infections and deaths. The courses of the reported daily new coronavirus infections (blue lines) and deaths (yellow lines), and the theoretically expected course of the number of deaths based on an estimated temporal delay of 10 days between diagnosis and death (red lines), are shown for the countries Austria, Belgium, France, Germany, Italy, and USA. Note that for the purpose of visual comparison, the values for new infections are scaled to the level of the number of deaths based on the respective death rates in each of the countries.

Discussion

The present findings indicate that the observed increases in reported new infections dramatically overestimate the true spreading of the coronavirus in all of the examined countries. Statistically controlling for the concurrent increases in the number of tests suggest that the true increases in new infections were relatively small in every of the examined countries, and that the course of new infections has already flattened or is even decreasing in almost every country since the beginning of calendar week 13 (March 23).

The fact that the courses of reported new infections and deaths started to increase almost simultaneously in every country provides further evidence that the increases in reported new infections reflect effects of increased testing. From a biological perspective, the absence of a temporal lag between the increases in new infections and deaths is surprising since there should be a substantial temporal lag between diagnosis and death. According to findings from China, the time span between the onset of symptoms and death is about 18 days¹¹. Thus, even when conservatively assuming that individuals are tested four days after symptom onset, there should be a temporal lag between increases in new infections and deaths of 14 days. The only reasonable explanation for the absence of a temporal lag between the increases in new infections and deaths may be that that many of the deceased people were tested on the coronavirus shortly before or after death. However, if so, this implies that one of two possibilities must be true. The first possibility is that the deceased people have really died of the coronavirus. However, this would mean that if the increased testing had been started already 14 day earlier, one would have found a comparable increase in new infections. The second possibility is that the deceased people only have become infected with the virus shortly before death, but actually have died of another disease. However, this would mean that the growth curves for new infections and deaths actually depict the same thing: the increases in the number of new infection that is brought about

by the increased number of tests.

One issue that may be finally discussed is the question of how the estimated smaller increases in new infections fit with reports from several countries that intensive care units are crowded, or with pictures as the ones from Italy where coffins of died people are accumulated in churches, which has even experts led to assume that such scenarios may take place in many country if no countermeasures against the transmission of the coronavirus are taken¹². However, there is one aspect that is often overlooked. In almost any country, only a relatively small part of people tested on the coronavirus receives a positive test result. For instance, in Germany, only around seven to eight percent receive a positive coronavirus diagnosis¹³, and even in Italy where it is assumed that only people with more severe respiratory symptoms are tested for the coronavirus, only around 20 percent receive a positive coronavirus diagnosis¹⁴. Since mainly people with acute respiratory symptoms are tested, people receiving a negative test result are not healthy but suffer from other diseases, suggesting that other respiratory diseases are currently circulating that are masked by the current strong focus on the coronavirus. Thus, reports from crowded intensive care units and pictures with many coffins of died people may be partly misleading in that a relatively large part of these people may actually have suffered from other diseases, and not from the coronavirus. Indeed, this is empirically supported by data from the National Center of Health Statistics of the USA¹⁰. From the 6,427 people that have died in the USA of the coronavirus according to diagnosis in between March 22 and April 11, only 2,925 (42.2%) died of pneumonia. Within the same three weeks, however, even when excluding pneumonia deaths involving influenza, overall 10,006 people have died of pneumonia in the USA. Thus, at least in the USA, only a relatively small part of the deaths involving pneumonia were actually caused by the coronavirus.

In conclusion, the present findings indicate that the coronavirus crisis appears to be based on a statistical fallacy: at some point in time, a new virus test is developed, accompanied by a big echo in the media, leading to a rapid increase in the application of the new virus test, and thus a rapid increase in reported new virus infections and deaths, which gives the impression that we are facing a pandemic with millions of infections and deaths – although in reality the increase in new infections has been only relatively small, and the number of new infections has relatively quickly started to decrease. Becoming aware of this statistical fallacy seems to be extremely important in order to counteract the extreme fear that is induced by the fallacy-prone horror scenario that there may be soon millions of coronavirus infections and deaths.

Methods

Data. Data on the numbers of daily new coronavirus infections and deaths for the countries Austria, Belgium, France, Italy, and the USA were retrieved from the European Center for Disease Prevention and Control (ECDC), which publishes a daily updated data file on the coronavirus disease². For Germany, these data were retrieved from the NPGeo Corona Hub 2020 (Robert Koch Institute)⁹. Official data on the number of conducted coronavirus tests for Austria, Belgium, France, and Italy are provided by the respective national Institutes for Health³⁻⁶. For Germany, official data on the mean daily test capacities in Germany in calendar weeks 10-14 is provided in the daily situation report of the Robert Koch Institute on the coronavirus disease from April 8; the number of tests per week was determined by multiplying the mean daily test capacities by 5 (5-day working week)⁷. Data on the daily number of tests in the USA is provided by the CODID Tracking Project which provides data based on an aggregation of data released by individual states⁸. The raw data on which the present analyses are based can be downloaded at https://osf.io/hkaru/?view_only=830bfd6cbea14744811423308e851827.

References

1. Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **Epub ahead of print** (16 Mar 2020).
2. Data on the numbers of daily new infections and deaths for the countries Austria, Belgium, France, Italy, and the USA were retrieved from the European Center for Disease Prevention and Control (ECDC), which publishes a daily updated data file on the coronavirus disease: <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases> (data retrieved on April 13, 2020, 10:00 a.m. CET).
3. Official data on the number of conducted tests in Italy is provided by the Ministry of Health and compiled by the Department of Civil Protection on Github: <https://github.com/pcm-dpc/COVID-19>. The data is collected and made available by the webpage [ourworldindata.org](https://ourworldindata.org/covid-testing): <https://ourworldindata.org/covid-testing> (data retrieved on April 13, 2020, 10:00 a.m. CET).
4. Data on the mean daily test capacities in the USA in calendar weeks is provided by the CODID Tracking Project which provides data based on an aggregation of data released by individual states: <https://covidtracking.com/> (data retrieved on April 13, 2020, 10:00 a.m. CET).
5. Official data on the mean daily test capacities in Germany in calendar weeks 10-14 is provided in the daily situation report of the Robert Koch Institute on the coronavirus disease from April 8, 2020; the number of tests per week was determined by multiplying the mean daily test capacities by 5 (5-day working week):

https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-04-08-de.pdf?__blob=publicationFile (Accessed 14 April 2020).

6. Data for Germany on the numbers of infections and deaths were retrieved from the NPGeo Corona Hub 2020 (Robert Koch Institute): https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0 (data retrieved on April 13, 2020, 10:00 a.m. CET).
7. Official data on the number of conducted tests in Austria is provided by Austria Ministry for Health who publishes daily updates on the number of tests conducted per day: [https://www.sozialministerium.at/Informationen-zum-Coronavirus/Neuartiges-Coronavirus-\(2019-nCov\)](https://www.sozialministerium.at/Informationen-zum-Coronavirus/Neuartiges-Coronavirus-(2019-nCov)). The data is collected and made available by the webpage [ourworldindata.org](https://ourworldindata.org/covid-testing): <https://ourworldindata.org/covid-testing> (data retrieved on April 13, 2020, 10:00 a.m. CET).
8. Official data on the number of conducted tests in Belgium is provided by the Belgian Institute for Health who publishes the number of tests conducted per day: <https://epistat.wiv-isp.be/covid/>. The data is collected and made available by the webpage [ourworldindata.org](https://ourworldindata.org/covid-testing): <https://ourworldindata.org/covid-testing> (data retrieved on April 13, 2020, 10:00 a.m. CET).
9. Official data on the number of conducted tests in France is provided by the Agence Nationale de Santé Publique who has published the number of conducted tests per day until March 8, and per week starting from March 9: <https://www.santepubliquefrance.fr/recherche/#search=COVID-19%20:%20point%20epidemiologique&sort=date>. The data is collected and made available by the webpage [ourworldindata.org](https://ourworldindata.org/covid-testing): <https://ourworldindata.org/covid-testing> (data retrieved on April 13, 2020, 10:00 a.m. CET).

- 299 10. National Center of Health Statistics, USA: Provisional Death Counts for Coronavirus Disease
300 (COVID-19): <https://www.cdc.gov/nchs/nvss/vsrr/COVID19/index.htm> (data retrieved on
301 April 14, 2020, 12 a.m. CET).
- 302 11. Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., et al. Estimates of
303 the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis* **Epub**
304 **ahead of print** (30 Mar 2020).
- 305 12. Lipsitch, M. We know enough now to act decisively against Covid-19. Social distancing is a
306 good place to start. *STAT* (18 March 2020).
- 307 13. Seifried, J., Hamouda, O. Erfassung der SARS-CoV-2- Testzahlen in Deutschland. *Epid.*
308 *Bull.* **15**, 3–4 (2020).
- 309 14. Dipartimento della Protezione Civile, Aggiornamento [in Italian]: [https://github.com/pcm-](https://github.com/pcm-dpc/COVID-19/blob/master/schede-riepilogative/regioni/dpc-covid19-ita-scheda-regioni-20200404.pdf)
310 [dpc/COVID-19/blob/master/schede-riepilogative/regioni/dpc-covid19-ita-scheda-regioni-](https://github.com/pcm-dpc/COVID-19/blob/master/schede-riepilogative/regioni/dpc-covid19-ita-scheda-regioni-20200404.pdf)
311 [20200404.pdf](https://github.com/pcm-dpc/COVID-19/blob/master/schede-riepilogative/regioni/dpc-covid19-ita-scheda-regioni-20200404.pdf) (data retrieved via GitHub on April 13 2020, 10:00 p. m. CET).