



Economic Commission for Europe**Conference of European Statisticians****Sixty-ninth plenary session**

Geneva, 23-25 June 2021

Item 2 (a) of the provisional agenda

Post-Covid: sustaining organisational and product innovation in national statistical offices.**Innovation in National Statistical Offices products and activities****Post-Covid: Innovation in products at the Spanish Statistics Office****Prepared by Spain***Summary*

The COVID-19 pandemic has produced changes in some statistical paradigms. One of the lessons learned has been that for analysing what was happening and have the ability to react quickly and effectively, the information had to be available as quickly as possible, close to on a continuous basis and with enough degree of granularity. Thus, what could be understood as short-term statistics in a normal situation, drastically changes its meaning and the availability of data with a very high frequency (daily, weekly) becoming necessary for policy-makers and, in general for society. NSOs had to respond to this challenge, investigating and seeking for new sources of information and innovating in processes and products to be useful to population.

This document describes some new and innovative statistical products elaborated at INE-Spain on an experimental basis that that are expected become official statistics in the near future, concerning mobility of people, daily turnover of big companies and, finally accounting population deaths on a weekly basis

The paper is presented to the Conference of European Statisticians session 'Innovation in National Statistical Offices products and activities' for discussion.



I. Some statistical demands arising during the pandemic – Innovative products

1. Mid-March 2020. In Spain, like in many other countries, the population was locked-down at home, borders were closed and the non-essential activity of the economy was stopped or reduced to a minimum. These measures, with no precedent in the recent past, intended to limit the impact of Covid-19 on population health but had serious impact on the compilation of statistics.
2. Nevertheless, the pandemic opened a window of opportunities and challenges to statisticians. On one side, they have had to innovate in processes to maintain statistical production, for instance replacing personal interviews by other methods guaranteeing a sufficient degree of quality or scraping private websites to extract public information. But, at the same time, they have had to search new sources of information in order to innovate in statistical products, making some quick and granular information available to policy-makers and to society, in general, for an efficient decision making in this context.
3. All these challenges addressed successfully by statisticians during the pandemic under the form of experimental statistics have determined new paradigms in the post-Covid statistical production. Firstly, new data sources both from public and private origin have been used extensively. Secondly, companies have understood during the pandemic that they had to collaborate with us to attend to their social responsibility and due to this, statisticians have shown that it is possible to set partnerships with private companies for using their data and compiling new statistics. And finally, statisticians have been aware of the value of making information available as quickly as possible, notably in periods of uncertainty.
4. The following paragraphs detail some examples developed at INE as experimental statistics during 2020. Because of the length of this paper, no excessive degree of detail has been included. More details can be found on INE's website: <https://www.ine.es>.

II. Mobility

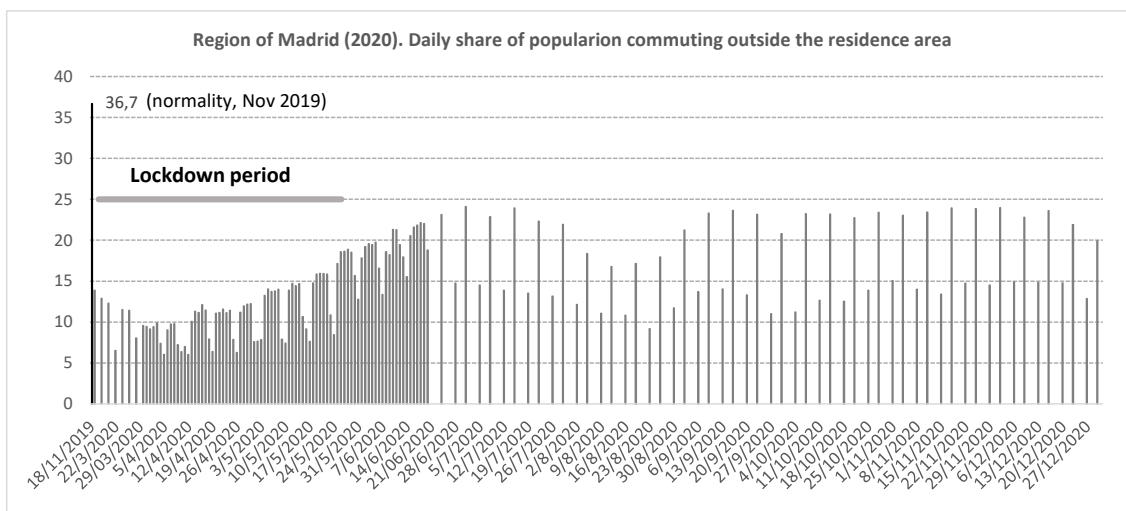
5. The first challenge in which INE innovated was addressing how the population was moving daily from resident areas to other locations (commuting) in comparison with a normal period before the pandemic, considering a high degree of granularity and publishing data very timely. For this purpose, information provided every day by the three main Mobile Network Operators (MNOs) in Spain on the location of mobile phone devices was used (80% of the total devices were counted for the statistic). Data were released with a delay of two days.
6. The study had started with reference to a normal week in November 2019, before the pandemic. Its original purpose was, as a satellite project of the 2021 Household and Population Census building commuting origin-destination matrices. Indeed, data were not on a daily basis but for this reference week only, but it was relatively not difficult to adapt the initial purpose and obtain data of mobility for pandemic management. Daily data were published from end-March to end-May, during the so-called state of alarm declared by the government. Afterwards, once data have become more stable, the series has continued providing data two days every week, one working day and one day of the weekend. More detailed information is available at the link:

https://www.ine.es/experimental/movilidad/experimental_em.htm
7. Specifically for this project Spain was divided into 3,214 territorial areas, amounting each to a population between 5,000 and 50,000 inhabitants. As extremes, some municipalities with less than 5,000 inhabitants were joined to other neighbours, in order to respect privacy and big cities were divided into districts (for instance, the city of Madrid was divided in 128 districts). The average size of each area was of 15,000 inhabitants (12,000 devices).
8. In a nutshell, the operation is as follows. To determine residency, it is analysed where the device is "sleeping", i.e. the area of residence of each mobile phone is the area where the mobile phone is found for the longest time between zero hours and 6AM every day. In the

same way, the destination area (commuting) is the one where the device is located most frequently in the frame from 10AM to 6PM.

9. The following chart represents the share of population in the Region of Madrid (around 6.5 million inhabitants) commuting daily to another area:

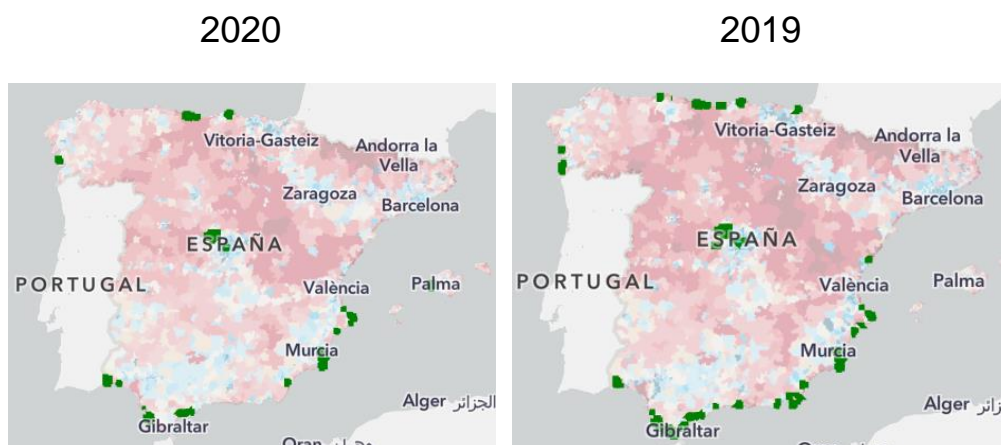
Figure 1
Daily share of population commuting outside of residence area in the region of Madrid



10. It is worth to observe that 36.7% of the population left daily its residential area for working, studying... before the pandemic, in a valley week of November 2019. This share was reduced to a 7% in the first half of April 2020 during the hard lockdown and was recovering progressively its level from springtime, but not exceeding 25% during working days in the remaining months of 2020. The measures of partial reducing of population mobility still in force as well as the use of remote work from home are behind of the causes for not have recovered the mobility to levels previous to pandemic.

11. Another result of this innovative statistic was referred, not to analyse commuting but the mobility during some strategic touristic dates, intending to answer to questions: where does the population resident in an area go on vacation? and, what is the area of residence of the population that is on vacation in a certain area?. Three strategic dates were analysed: one in July, one in August and, finally, one in December for Christmas holidays. In this case, for every mobile phone, the area of residence is defined considering the most frequent area where it is found in the night time during the last 60 days. In the following maps, we can compare the locations chosen for holidays of people resident in a centric district of Madrid for 15th August 2019 and 2020:

Figure 2
Holiday destinations of people resident in a centric district of Madrid for 15 August 2019 and 2020



12. You can get as much granular data as you want by using the web application. For each of the 3,214 areas, a list of origin/destination areas is available with the only limit of privacy. Movements of less than 15 devices among areas are not shown (not provided by NSOs).

13. This product was the first innovative product compiled at NSI during pandemic but, afterwards, the same technology has considered useful for innovating in other statistical areas, as for instance, in tourism statistics.

14. Actually, the experimental statistic describing mobility between residential and vacation areas describes a fix picture in which, people resident in an area A is detected in another area B in a concrete date, for instance August 15. But it is not known how this device/person got that destination, if it/he arrived directly from the resident area or, on the contrary, arrived there after some intermediate steps, taking part of a tour.

15. Therefore, the same technology based on MNOs data can be also used for the measurement of travels and tours carried out, not only for resident people along the Spanish territory but also for resident people in their travels abroad and for non-resident people in their travels to Spain.

16. This new product is currently under construction at INE and the first results are expected to be published, as experimental statistic, in the course of 2021.

III. Daily turnover of big retail trade companies

17. Another challenge addressed by NSI-Spain during pandemic concerned the provision of data for assessing the impact of lockdown on economic outturns. To investigate the way of providing policy makers with daily data was considered as crucial.

18. In this sense, the Spanish Tax Agency facilitated NSI daily data of sales of big companies, obtained from the system called SII (Immediate Information Supply by its initials in Spanish) system by which big companies submit invoices on a daily basis to the Tax Agency. Data were used to compile data of companies with activity in retail trade obtaining paths of daily volume of sales. More detailed information is available at the link: <https://www.ine.es/experimental/cdmge/>

19. The Spanish Retail Trade Index (RTI) is a monthly indicator compiled at INE following the EU legislation in force, based on a sample of different size companies (around 12,000) that releases monthly estimates of turnover and employment with a delay of around 25 days.

20. In this activity, big companies (amounting sales of more than 6 million euro) represent a bit more of 50% of the total sales of sector and in some cases, as in non-specialized establishments (like big surfaces and chains) the share is very significant as sales amount more than 80%.

21. Data provided by the SII are available on a daily basis. The behaviour of the series is really particular and highly dependent of the way in which companies usually do invoicing. We can observe not big differences in data along working days (not holidays), an increasing on Saturday and a high reduction of sales on Sunday. At the same time, we can observe a peak in the last working day of each month.

22. In addition, the provision of an invoice to the SII cannot be immediate. It can take several days, so the actual data of a day is more or less stable after 10 days. Initially, INE is releasing the daily indicator twice a month. First release covers the first 15 days of the month and is published usually around 26th – 30th of the same month and second release covers the complete month and is released usually around 11th – 15th of the following month.

23. As it can be observed in the following charts, if we compare the sales of big retailers in April 2020 with the ones of the previous year (before pandemic) the difference between the two lines is increasing day after day, being more than 20% less at the end of the month (year on year). Similarly, comparing April with the previous month, we can see that the highest difference in the accumulated sales is produced around mid-month. This was caused by the increasing of the sales at the beginning of the lockdown, in which, households

collected a significant amount of goods because of the uncertainty of the supply of goods during pandemic.

Figure 3

Retail daily sales. Total April 2019=100

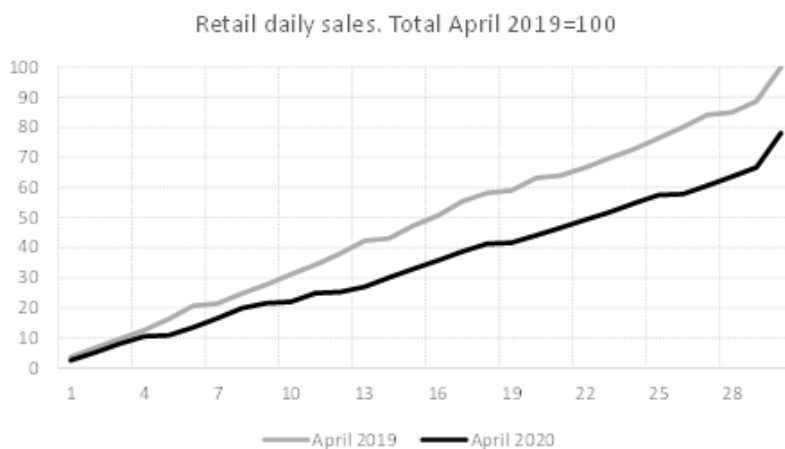
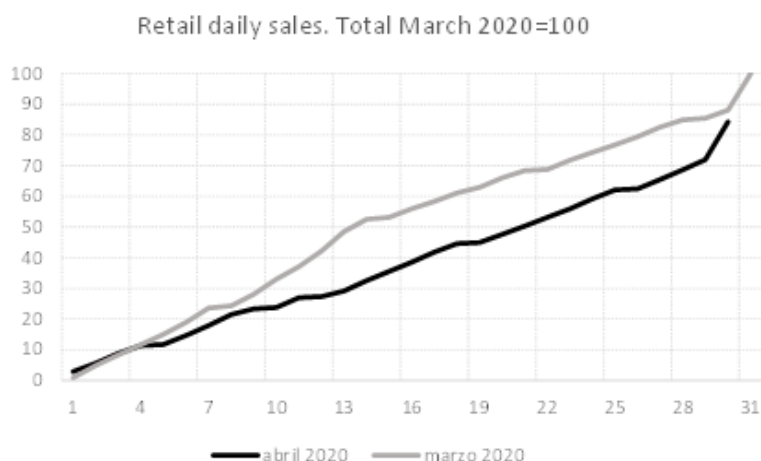


Figure 4

Retail daily sales. Total March 2020=100



24. This information was extremely useful for policy makers in order to react as quickly as possible in the decision making.

25. In the current version of the indicator released by INE, data of the so-called Common Fiscal Territory (not covering the regions of Navarre and the Basque Country) have been included. The addition of data from these territories will be explored in the future by INE.

26. In the same way, this innovative product will be extended for compiling similar information on companies of other activities, both in services and also manufacturing.

27. It is clear that with these type of innovative products using new sources of fast data, what could be understood by short-term statistics is changing its meaning. The possibilities daily data offer to analysts, even with no complete coerture (data of credit card sales is a good example) open a second tier of information, quickly available, supplementing monthly/quarterly statistics, usually assumed as short-term data, allowing a faster decision taking, notably in uncertain periods. INE is sensible to the new paradigm, increasing its statistical basis of high frequency. The indicator here shown is only an example of the new era.

IV. Weekly population deaths

28. One of the statistical needs that arose during the pandemic concerned a timely and prompt availability of population deaths. The current practice of INE regarding the measurement of this phenomenon before the pandemic was publishing data twice a year (monthly data) with six months of delay, in the context of population statistics (births, deaths, migration...). The increasing demand for quicker and timelier data for managing the pandemic and the need of comparable data among countries in the context of Covid-19 deaths was taken by INE, innovating in a new product, given as a result a series of weekly deaths using data basically provided by Civil Registries of the Ministry of Justice. These statistics are one of the most visited in the INE's website during the last months.

29. The basic information for compiling weekly estimates of deaths come from the Civil Registries and, more concretely, from an application named Inforeg, which runs in municipalities with computerized Registries, by which, approximately, a 93% of the total deaths are covered. Inforeg offers the major advantage of being always up-to-date: on any given day, deaths recorded the week before are available. No more delays are usually produced. These data recorded by Inforeg combined with historical information on vital statistics are the basis to estimate weekly deaths.

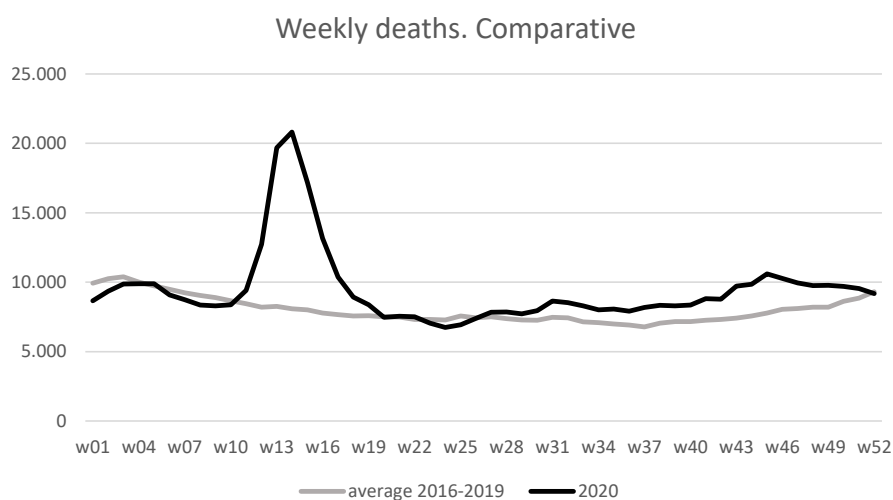
30. The experimental statistic was released by INE since June 2020 and offer weekly data by regions and provinces, twice a month with a delay around 10 days. More details can be obtained at the link below:

https://www.ine.es/en/experimental/defunciones/experimental_defunciones.htm?L=1

31. Data are classified not by the place of residence of the person deceased but the location of where the death occur, i.e. the place where the registration is made. Data are broken by gender and by 5-year age groups.

32. From an operational point of view, data from Inforeg are expanded to the total population using factors to correct the under coverage of Inforeg (Registries not computerized) in each territory.

Figure 5
Comparative of weekly deaths



33. In the chart, we can observe the difference of deaths in 2020 compared with an average of the previous 4 years. Obviously, although it is highly related with, the difference does not mean the mortality by COVID-19 because the pandemic has determined changes in the mortality rates of the population due to different causes. The statistic of Deaths by Cause is the official statistic providing the information of deaths by COVID and by other causes. This operation was advanced by INE in December 2020 for the period from January to May. In any case, the graph is very explanatory as regards the two waves of the pandemic in 2020,

One in early spring, short and very incisive and another one during the fall, much longer but less incisive.

V. Conclusions

34. The examples described in the paper show that other ways to compile statistics different from the traditional ones are possible. In this vein, the extensive use of administrative information leading to more timely statistics, increasing the frequency of data availability and making accessible very detailed and granular data to users should be one of the drivers in elaborating post-COVID statistics. In addition, other drivers should be the more extensive use of privately held data.

35. All these new innovative products, considered initially as experiments, could become official statistics in the short-term once analysed sufficiently. This should produce, as a consequence, first, the increasing of the statistical basis but, also, the revision of the designs of some current official statistics to reduce their dependence on samples, aiming to lighten the burden on respondents.
