


Monitoring the number of COVID-19 cases and deaths in Brazil at municipal and federative units level

Wesley Cota ^{*}

Departamento de Física, Universidade Federal de Viçosa, 36570-900 Viçosa, Minas Gerais, Brazil

We present a dataset containing the reported number of COVID-19 cases and deaths at municipal and federative units level in Brazil. Data is aggregated daily from official sources with the most updated numbers, providing a reliable, free and simple resource for researchers, health authorities and general public. Interactive pages in English and Portuguese are available, containing maps, graphs and tables with all the data. Data about recovered, suspected and tests made are also available for most federative units.

An outbreak of the coronavirus disease 2019 (COVID-19) was identified in Wuhan, China, in December 2019 [1], caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It has spread to provinces of mainland China and then to all six World Health Organization (WHO) regions [2]. In Brazil, the first imported case was confirmed on February 25, 2020 in the city of São Paulo/SP, and was officially announced by the Ministry of Health of Brazil – *Ministério da Saúde* (MS) – one day later [3]. Since then, the Federal Government of Brazil reports daily the total number of cases and deaths for each federative unit (UF)¹, that corresponds to the Federal District (*Distrito Federal*) and 26 states. However, Brazil has 5570 municipalities with varied population distributions and characteristics and official information at least at municipal level is essential for researchers, health authorities and the general public. In order to provide reliable data also at municipal level, we developed an online interactive page to visualize and track the reported cases of COVID-19 in Brazil by using official data from each federative unit reports. These reports contain data at municipal level for both the number of cases and deaths.

The page [<https://covid19br.wcota.me/en>] was first shared publicly March 13, and contains interactive plots of the number of confirmed, deaths, recovered and suspected cases. A full dataset is publicly available and updated daily through a GitHub repository [<https://github.com/wcota/covid19br>] in CSV format. During Mar 13–Apr 09 the data collection was done by reading the reports and updating the number of cases for each municipality and states. On April 10 a semi-automated strategy was adopted by using municipal data source via Brasil.IO [<https://brasil.io/dataset/covid19/caso/>], that started to collect data in a similar way by a task force of 40 volunteers. As before, the municipal level data is aggregated to the number officially reported by *Ministério da Saúde*, see Sec. A for details. In order to keep the dataset updated almost in real time, we monitor the cases compiled by the Twitter account @Coronavirusbra1 [<https://twitter.com/coronavirusbra1>] through a Google Sheet. Before updating the dataset, we

confirm the case numbers with the official sources cited by the data streams. Data is also available for the number of recovered, suspects and tests made in each federative unit, when available.

In Fig. 1 we present the comparison between the number of reported cases by *Ministério da Saúde* and the ones aggregated in this dataset. Official data from the Brazilian government is updated only once a day, resulting in differences in the numbers of cases reported each day. Note that while the government assumes that the first case was registered on February 26, other sources confirm that the epidemics started on February 25 in Brazil. We also present the temporal evolution for each federative unit in Fig. 2. As a result of the aggregation of federative unit data before being confirmed by the federal government, some small differences are found. In addition, we also provide georeferenced data at municipal level, allowing the visualization of the cases in a map such as the one presented in Fig. 3.

We believe our dataset can help efforts for modeling, tracking and controlling the current outbreak situation in Brazil, and we plan to continue updating this dataset.

We thank Carlos Achy (@coronavirusbra1), Álvaro Justen and all volunteers of Brasil.IO project for data contribution. We also thank Fábio Rehm, Rafael Calpena Rodrigues, Lucas Biló and Silvio C. Ferreira for useful discussions. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (Capes) – Finance Code 001.

A. Aggregation method and sources

This dataset aggregates data from at least three main sources:

1. **Ministério da Saúde** (MS) [<https://covid.saude.gov.br/>]: Official page of the Brazilian Ministry of Health that updates the number of cases per federative unit once a day. Only the number of cases and deaths for each federative unit are used.
2. **Brasil.IO** (IO) [<https://brasil.io/dataset/covid19/caso/>]: Data compiled daily from the official epidemiological bulletins of each federative

^{*} E-mail: wesley.cota@ufv.br; Homepage: <https://wcota.me/>

¹ *Unidade Federativa*, in Portuguese.

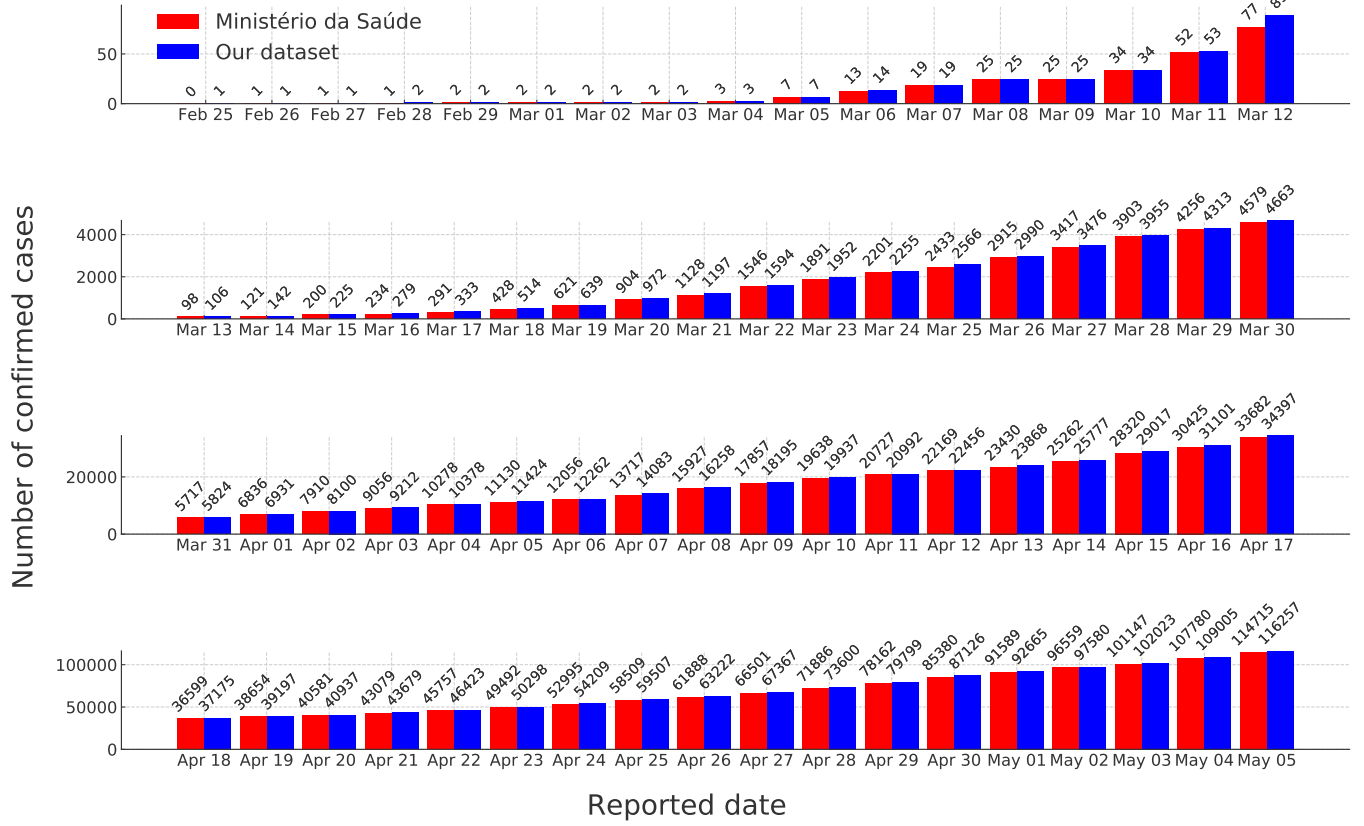


FIG. 1. Comparison between the number of cases reported officially by *Ministério da Saúde* and the ones aggregated in our dataset for each day.

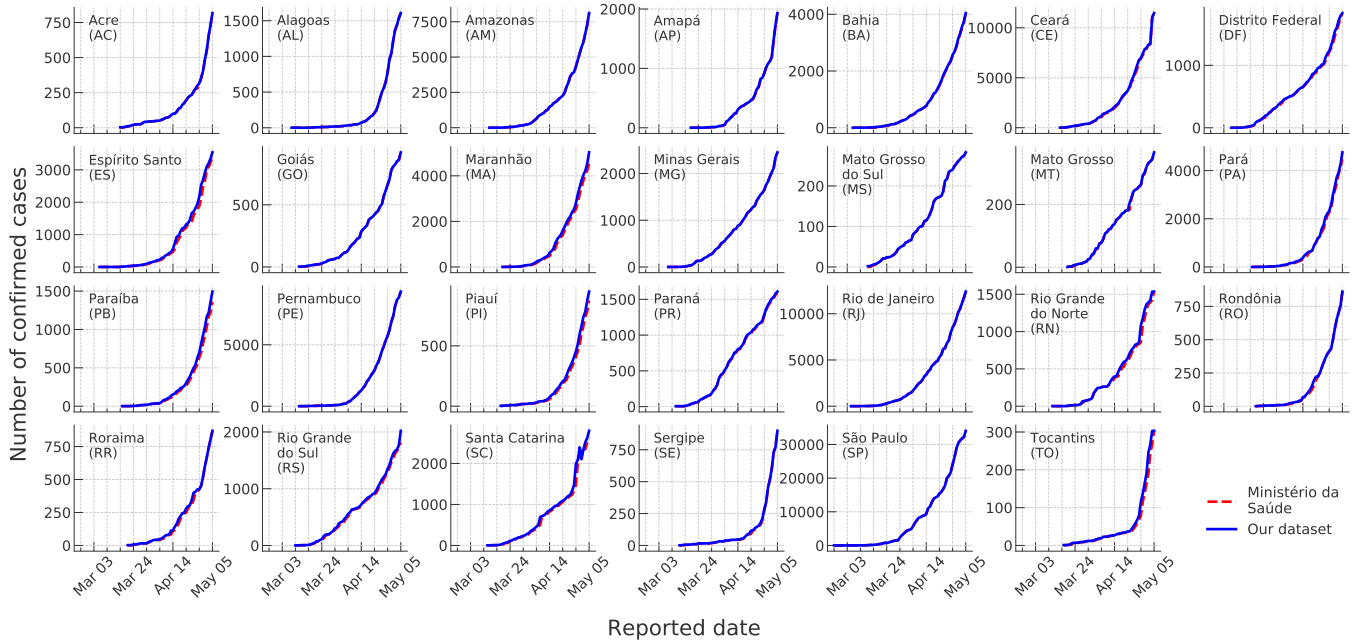


FIG. 2. Temporal evolution of the total number of confirmed cases for each Brazilian federative unit (Federal District + 26 states) for the official data by *Ministério da Saúde* and our dataset.

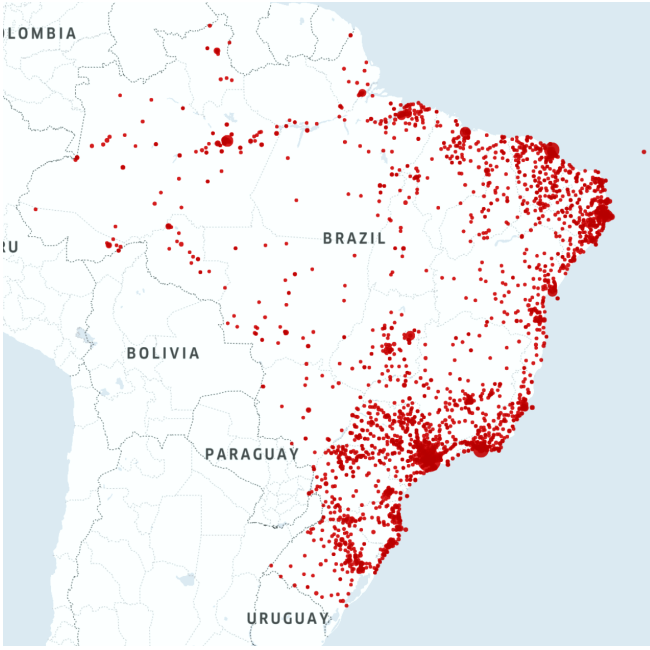


FIG. 3. Spatial distribution of the confirmed cases for each municipality in Brazil accumulated until April 28, 2020. Circles increase proportionally with the number of cases.

unit. This data is available at municipal level, containing the number of cases and deaths for each municipality.

3. **@coronavirusbra1** (TW) [<https://twitter.com/coronavirusbra1>]: Twitter account administrated by Carlos Achy that reports the most recent number of confirmed, recovered and suspected cases released by official websites, social accounts and news of each federative unit. In addition, the number of tests made is also available. The data comes from a Google Sheet [<https://docs.google.com/spreadsheets/d/1MWQE3s4ef6dxJosyqvsFaV4fDyElxnBUB6gMGvs3rEc/edit?usp=sharing>].

We denote the set of the number of cases (deaths) for each day as $\mathcal{C} = \{c_{t_1}, c_{t_2}, \dots, c_{t_N}\}$ ($\mathcal{D} = \{d_{t_1}, d_{t_2}, \dots, d_{t_N}\}$), where c_{t_i} (d_{t_i}) is the total number of cases (deaths) notified up to the day t_i . The last day of data update is denoted by t_N . First, we collect this number from the MS source for each federative unit. Then, we collect the set from the IO source for each municipality. If the number of cases (or deaths) for each federative unit by IO is greater than the ones of the MS dataset, these numbers are updated accordingly. If it is smaller, the number of the MS source prevails. The last dataset (TW) is used only for the last day t_N of data update. This ensures that the number of cases for each federative unit remains constantly updated, even if the first two sources have not yet reported those cases. For the number of recovered and suspected cases, and of tests made,

we use the TW dataset historical data, that is available for most of the federative units.

B. Description of the data

The full dataset is available in CSV format at a GitHub repository [<https://github.com/wcota/covid19br>]. It is freely available by the Creative Commons Attribution Share Alike 4.0 International License . An interactive website with the map, graphs and full tables is also available in English [<https://covid19br.wcota.me/en>] and Portuguese [<https://covid19br.wcota.me/>]. The 2019 Brazilian population estimate provided by the *Instituto Brasileiro de Pesquisas Estatísticas* (IBGE) [4] is used to compute the number of cases and deaths per 100 000 inhabitants.

1. Temporal evolution at municipal level

This data is available in the file `cases-brazil-cities-time.csv`. It contains the temporal evolution for the number of cases and deaths for each municipality with at least one reported case in Brazil. The following columns are available:

- **date**: reported date, in YYYY-MM-DD format,
- **country**: country of the data, which is always “Brazil”,
- **state**: short name of the federative unit (UF), or “TOTAL” when referring to the whole country,
- **city**: full name of the municipality in City/UF format. It can be “CASO SEM LOCALIZAÇÃO DEFINIDA/UF”², referring to those cases in the federative unit that does not have its municipality defined,
- **ibgeID**: unique ID for the municipality provided by the *Instituto Brasileiro de Pesquisas Estatísticas* (IBGE) [4],
- **newDeaths**: difference between the number of deaths of the corresponding date and the previous one,
- **deaths**: accumulated number of deaths in that date,
- **newCases**: difference between the number of cases of the corresponding date and the previous one,
- **totalCases**: accumulated number of cases in that date,
- **deaths_per_100k_inhabitants**: number of deaths per 100 000 inhabitants for that location,
- **totalCases_per_100k_inhabitants**: number of cases per 100 000 inhabitants for that location,

² Translation: case without a defined location. It also includes cases of people living outside that federative unit.

- **deaths_by_totalCases**: ratio between the number of deaths and cases.

2. Temporal evolution at federative unit level

This data is available in the file `cases-brazil-states.csv`. It contains the temporal evolution for the number of cases, deaths and recovered for each federative unit with at least one reported case in Brazil. The following columns are available:

- **date**: reported date, in YYYY-MM-DD format,
- **country**: country of the data, which is always “Brazil”,
- **state**: short name of the federative unit (UF), or “TOTAL” when referring to the whole country,
- **city**: in this file, it is always “TOTAL”,
- **newDeaths**: difference between the number of deaths of the corresponding date and the previous one,
- **deaths**: accumulated number of deaths in that date,
- **newCases**: difference between the number of cases of the corresponding date and the previous one,
- **totalCases**: accumulated number of cases in that date,
- **deathsMS**: accumulated number of deaths in that date reported by the MS source,
- **totalCasesMS**: accumulated number of cases in that date reported by the MS source,
- **deaths_per_100k_inhabitants**: number of deaths per 100 000 inhabitants for that location,
- **totalCases_per_100k_inhabitants**: number of cases per 100 000 inhabitants for that location,
- **deaths_by_totalCases**: ratio between the number of deaths and cases,
- **recovered**: number of recovered individuals reported by the TW historical data. It is empty if no data is available,
- **suspects**: number of suspect cases reported by the TW historical data. It is empty if no data is available,
- **tests**: number of tests made reported by the TW historical data. It is empty if no data is available,
- **tests_per_100k_inhabitants**: number of tests per 100 000 inhabitants for that location.

3. Accumulated number of cases at municipal level

This data is available in the file `cases-brazil-cities.csv`. It shows only the information corresponding to the most recent date reported in

the file `cases-brazil-cities-time.csv`. The columns are the same as in `cases-brazil-cities.csv`, except by **date**.

4. Accumulated number of cases at federative unit level

This data is available in the file `cases-brazil-total.csv`. It shows only the information corresponding to the most recent date reported in the file `cases-brazil-state.csv`. The columns are the same as in `cases-brazil-state.csv`, except by **date**. It also has an extra column **notConfirmedByMS** that computes the difference between **totalCases** and **totalCasesMS**.

5. Geolocated auxiliary file

This data is available in the file `cases-gps.csv`. It is an auxiliary file that provides the number of cases or deaths for each municipality, with GPS information obtained from the Nominatim API by OpenStreetMap. The columns are the following:

- **type**: it can be “0” (case without municipality information) and “1” (case with municipality information). The deaths are indicated by prepending the letter “D” to it,
- **name**: name of the municipality, or “*CASO SEM LOCALIZAÇÃO DEFINIDA/UF*” when the municipality is not known.
- **lat**: latitude information,
- **lon**: longitude information,
- **total**: total number of cases (or deaths) for that location,
- **total_per_100k_inhabitants**: the total number per 100 000 inhabitants.

An example of this data is shown in Fig. 3.

6. Sources file

This data is available in the file `sources.csv`. It contains the date of access (**date**) and URL (**URL**) for each reference used to compile the dataset.

C. Extra plots: deaths, recovered and suspects

The comparison between the reported number of deaths by MS and the ones in our dataset is shown in Fig. C.1. The number of deaths, recovered and suspects for each federative unit are shown in Figs. C.2, C.3 and C.4, respectively.

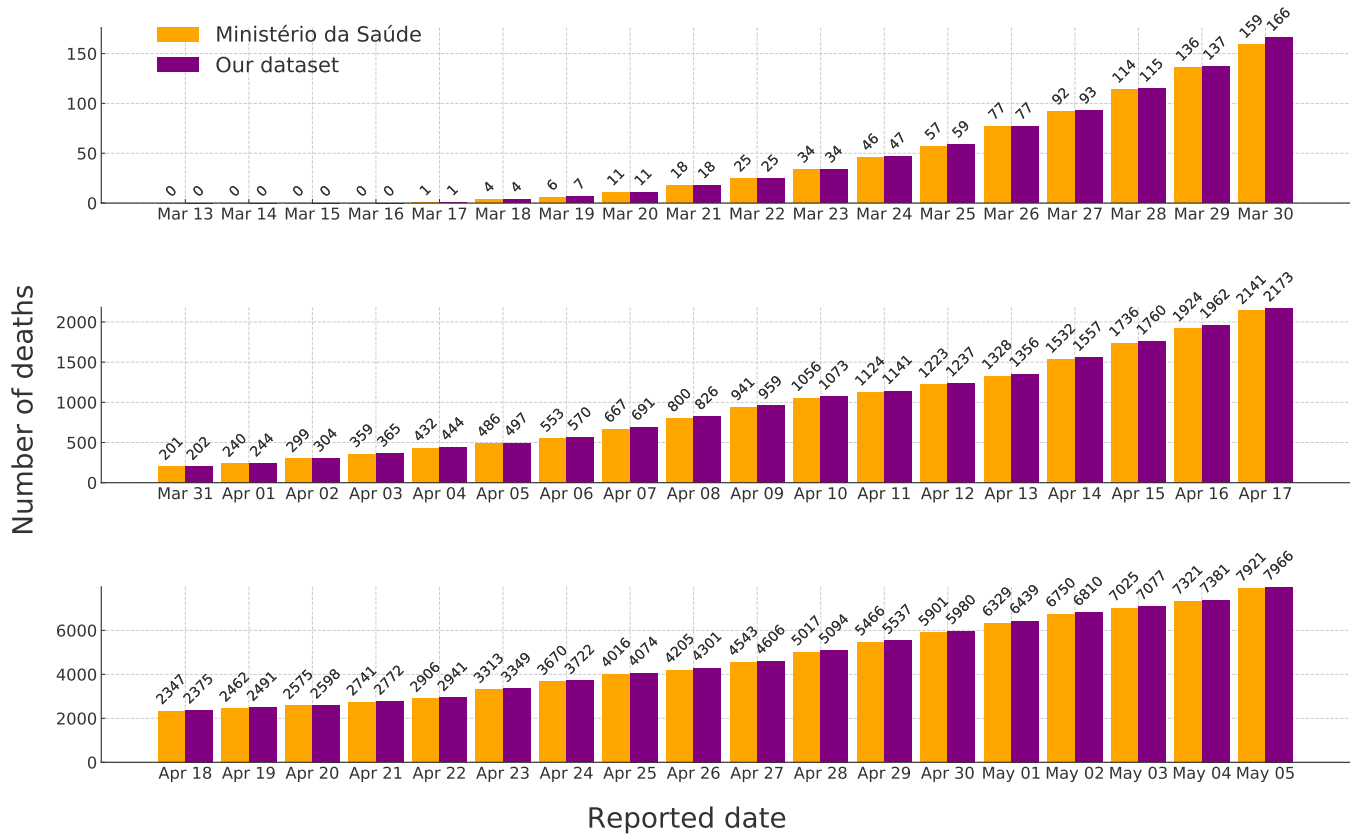


FIG. C.1. Comparison between the number of deaths reported officially by *Ministério da Saúde* and the ones aggregated in our dataset for each day.

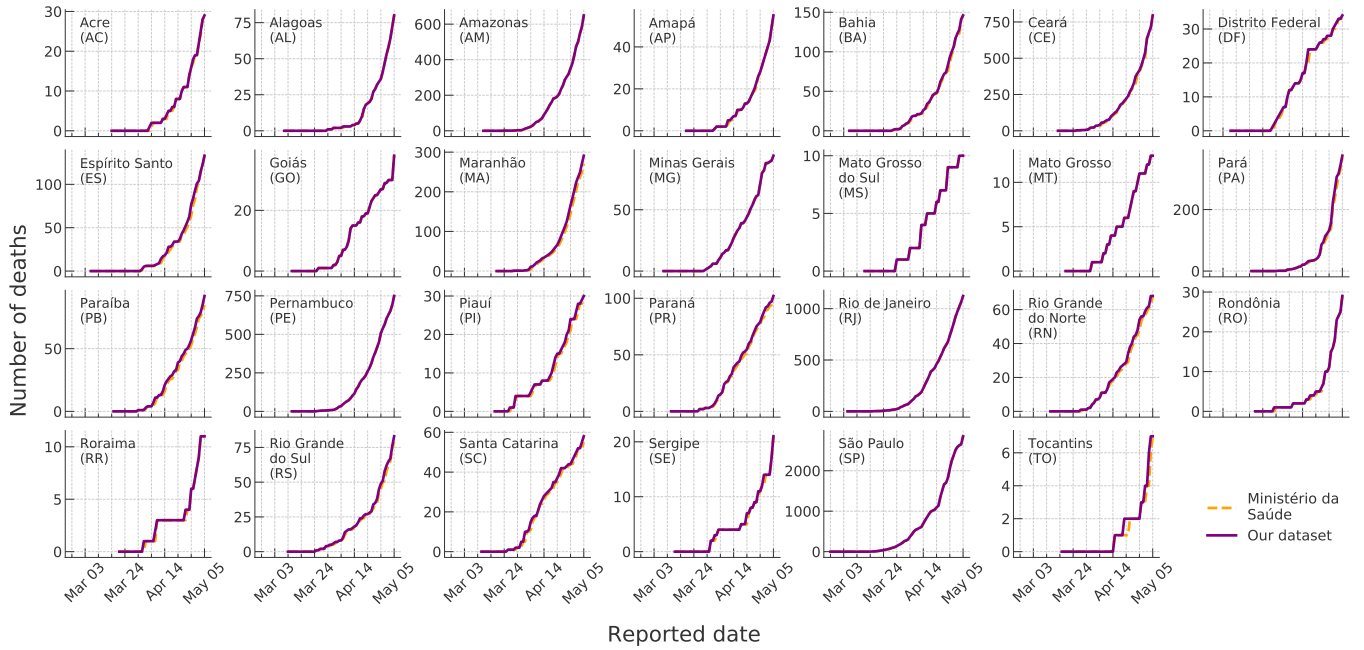


FIG. C.2. Temporal evolution of the total number of confirmed deaths for each Brazilian federative unit (Federal District + 26 states) for the official data by *Ministério da Saúde* and our dataset.

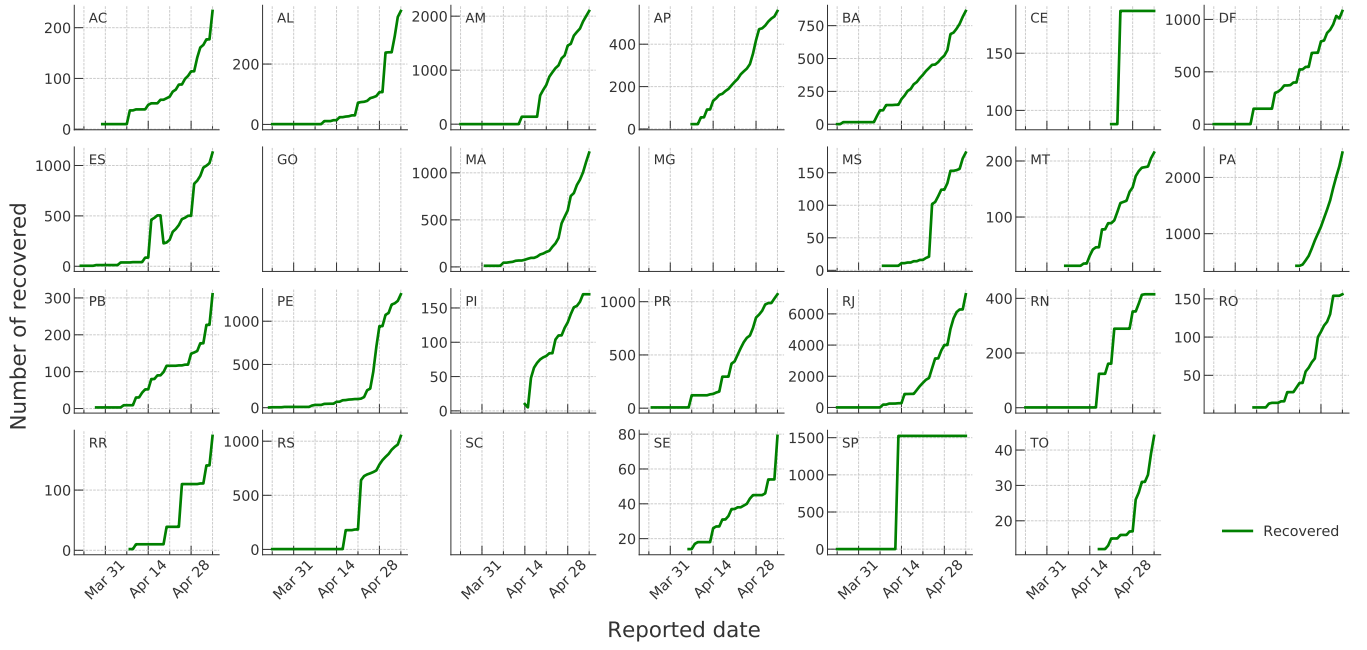


FIG. C.3. Temporal evolution of the total number of recovered cases for each Brazilian federative unit (Federal District + 26 states) for the official data by *Ministério da Saúde* and our dataset. There is no data available for GO, MG and SC, while for SP it was not recently updated.

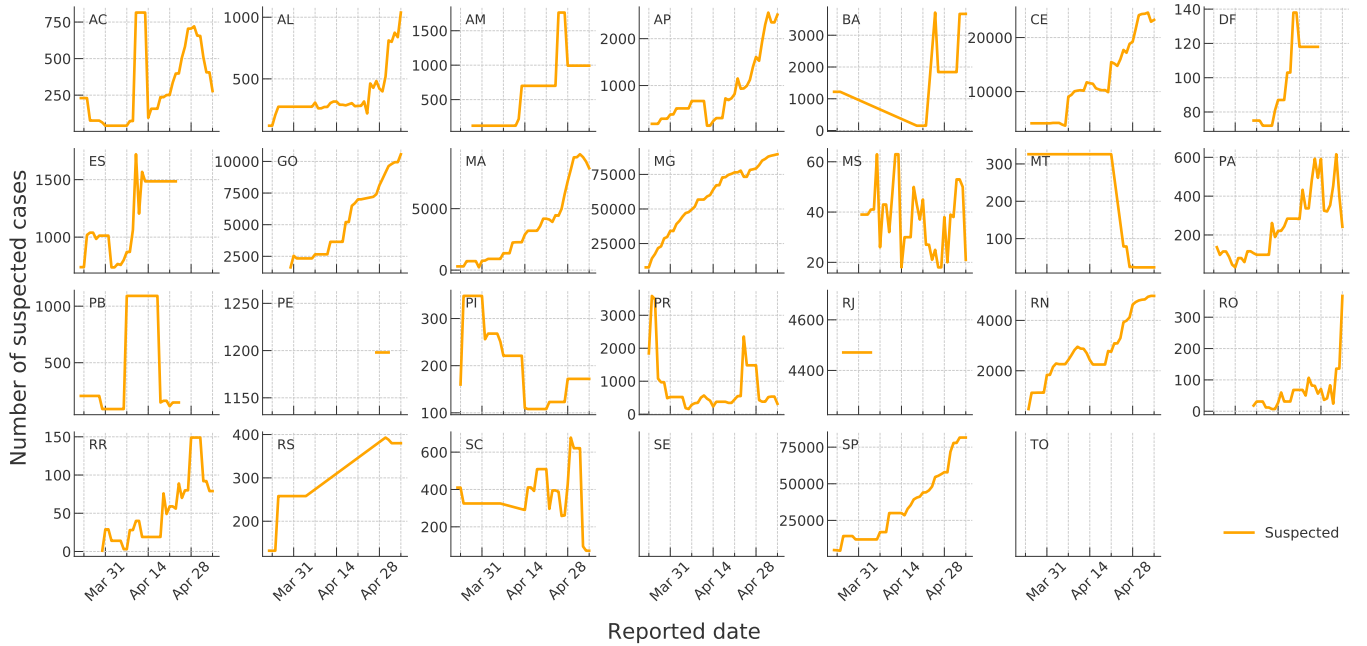


FIG. C.4. Temporal evolution of the total number of suspect cases for each Brazilian federative unit (Federal District + 26 states) for the official data by *Ministério da Saúde* and our dataset. There is no data available for SE and TO.

-
- [1] WHO, “[Novel coronavirus – China](#),” (2020), [Online; accessed 29-April-2020].
- [2] WHO, “[Coronavirus disease 2019 \(COVID-19\): Situation Report – 99](#),” (2020), [Online; accessed 29-April-2020].
- [3] Brazil, “[Brasil confirma primeiro caso do novo coronavírus](#) (*Brazil confirms the first case of the novel coronavirus*),” (2020), [Online; accessed 29-April-2020].
- [4] Instituto Brasileiro de Geografia e Estatística (IBGE), “[Estimativas da população residente para os municípios e para as unidades da federação brasileiros com data de referência em 1º de julho de 2019](#),” (2019).