Research protocol

- Project title:

Survival analysis of SARS-CoV-2 infection/COVID-19 cases to assess the effect of socioeconomic factors on time to death.

- Project summary:

To identify the effects of some socioeconomic factors of inequality on time to death by COVID-19, we will model their joint effects by a multi-predictor Cox Regression, using the whole set of confirmed cases available from the national surveillance system until October 26th.

- Project description:
  - Rationale

    The 11 Essential Public Health Functions (EPHF) summarize the health authority's expected performance in critical areas of public health. *Public Health Research*, EPHF 10, which complements *Monitoring, evaluation, and analysis of health status* (EPHF 1) and *Surveillance, research, and control of the risks and threats to public health* (EPHF 2), has been relevant to guide decisions during the COVID-19 pandemic in Colombia. In this way, deeper research complements the regular and routine assessment of COVID-19 the ministry of health and related institutions make, to answer more complex questions.

    Routine univariate and bivariate analysis of surveillance flagged some possible patterns of unequal distribution of COVID-19 across society, according to socioeconomic status and ethnicity. This triggered decision-makers' concern related to how socioeconomic inequalities could shape the extent and distribution of COVID-19 pandemics in the country. As an answer, we planned this study which deepens analysis with multi-predictor techniques that control for covariates included in the model.

  - Objectives

    Principal Objective

    Identify the effect of some socioeconomic inequality-related factors on COVID-19 mortality during the first eight months of the epidemic in Colombia.

    Secondary objectives

    - Search for other countries' evidence about socioeconomic variables' effect on COVID-19 distribution, severity and death outcome.
    - Develop a Cox model with multiple predictors to assess the joint effect on survival/time to death for COVID-19 confirmed cases.

  - Methodology

    - Research design: Observational cohort study to evaluate the effect of socioeconomic variables on time to death for SARS-CoV-2 infection/COVID-19.

    - Research subjects or participants: All persons with SARS-CoV-2 infection/COVID-19 detected and confirmed by the Colombian national surveillance system. No additional selection or exclusion criteria will be applied.

- Data collection follows the national surveillance questionnaire available on the INS Website. (1) The National Institute of Health, (INS), a Ministry of health attached agency, integrates individual data that local health care providers and public health authorities collect across all country regions. After an initial process of cleaning and verification, the INS publishes open data anonymized and deidentified individual records including several variables, available on the open data government portal. (2)

- Because its observational character, the study will not add any particular intervention. During the study, all customary decisions of the public health system will remain the same. However, results will inform decision-makers to improve public health response to COVID-19 pandemics.

- We will analyze the full data set from March 2nd to October 26th. Before the analysis, preprocessing and exploratory data analysis will assess data quality for completeness and to preprocess the variables of interest by recoding to the right formats. For categorical predictors, we will keep the original groupings, as also we will keep uncertainty groups of cases in specific categories. For symptomatic cases, time to death will rely on differences between dates of symptom onset and death or recovery. For asymptomatic cases, as there is no date of symptom onset, the first follow-up date is the clinical appointment. Time to event of symptomatic or asymptomatic censored cases will contribute to the model from dates of symptom onset or consultation, respectively, until 26th October.

- We will fit a multi-predictor Cox regression with time to death, to assess hazard ratios for Sex, Socioeconomic Status, Age groups according to life course, Ethnicity, Rural or urban area, and Health care insurance. The modeling process includes verification of assumptions of proportional hazards to introduce time-dependent variables if needed. (3)(4)

- Data management and analysis

  The national public health surveillance routinely collects data about different events including SARS-CoV-2 infection/COVID-19 (5), according to standard procedures, protocols and instruments the National Institute of Health (INS) operates. As the government and the ministry encourage the use of data, the Open Data web portal provides access to anonymized and deidentified registries. The Ministry of Health has developed guidelines for COVID-19 data and information basic, intermediate and advanced analyses, which complement multiple dashboards and resources that easily gather several dimensions of the pandemic available to local health authorities and decision-makers, researchers and the citizens.

- Ethical considerations

  Individual data and information are one main input to achieve the Essential Public Health Functions 1, 2 and 10, which provide evidence for decision-making (EPHF 4, 6 and 8) and actions (EPHF 3, 5, 7 and 11). For COVID-19 surveillance, public health assessments begin at the individual level, from the person's health records and the epidemiological case definition, the follow-up and contact tracing for COVID-19, and imply decision-making and interventions targeting also individuals, such as quarantine or isolation, and hopefully, a future vaccination. In this case, as for other events of public health interest, the Colombian public health authorities warrant the protection of the patients' confidentiality by following the national regulations of Habeas data and are responsible for using data to improve population health. (6)

Under habeas data regulations, Open individual data from COVID-19 surveillance is available after procedures to make it safe by anonymization and deidentification, which include removal of any direct or indirect identification. In this way, transparent and equal access to safe data for research allows health authorities and the full society to be aware, follow and conclude about the extent of the pandemic in the whole country. This study used this type of data.

- Gender issues: Data collection equally addresses all people inside the country, nationals or not nationals.

- References

1. INS. Ficha 346 IRA virus nuevo 2020 [Internet]. SISTEMA NACIONAL DE VIGILANCIA EN SALUD PÚBLICA Subsistema de información SIVIGILA Ficha de notificación individual. [cited 2020 Nov 23]. Available from: https://www.ins.gov.co/Noticias/Coronavirus/346_IRA_virus_nuevo_2020.pdf

2. Gobierno de Colombia. Casos positivos de COVID-19 en Colombia | Datos Abiertos Colombia [Internet]. [cited 2020 Nov 23]. Available from: https://www.datos.gov.co/en/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr

3. Vittinghoff E, Glidden DV, Shiboski S, McCulloch CE. Regression Methods in Biostatistics. Linear, Logistic, Survival and Repeated Measures Model. Second edition. New York, NY.: Springer; (Statistics for Biology and Health).

4. IBM. IBM SPSS Advanced Statistics 26. IBM;

5. INS. Fichas y Protocolos SIVIGILA [Internet]. Fichas y Protocolos Sistema Nacional de Vigilancia en Salud Pública. [cited 2020 Nov 23]. Available from: https://www.ins.gov.co/buscador/Paginas/Fichas-y-Protocolos.aspx

6. Lee LM, Heilig CM, White A. Ethical Justification for Conducting Public Health Surveillance Witout Patient Consent. American Journal of Public Health. 2012 Jan;102(1).