Moreno *et al. International Journal of Health Geographics* (2026) 25:7
https://doi.org/10.1186/s12942-025-00448-0

International Journal of Health Geographics

RESEARCH Open Access

Understanding socioeconomic inequalities in COVID-19 vaccination: controlling endogenous selection in Cali, Colombia

Manuel A. Moreno[1,2,3,5], Francisco J. Rodríguez-Cortés[4], Marc Saez[1,3,5] and Maria A. Barceló[1,3,5*]

Abstract

Background The COVID-19 pandemic displayed notable disparities in infection and mortality rates across populations, yet socioeconomic factors remain underexplored in many analyses. This study leverages an individual-level dataset from Cali, Colombia, detailing COVID-19 cases, vaccination histories, and mortality outcomes, to examine spatiotemporal vaccination patterns and their effects on mortality.

Methods Using a Bayesian two-part model with generalized linear mixed models, the analysis accounts for endogenous selection, individual heterogeneity, and spatial-temporal dependencies.

Results The findings highlight significant socioeconomic inequalities in vaccination coverage: individuals from higher socioeconomic strata were more likely to receive full vaccination regimens and booster doses, while those from lower strata faced reduced vaccination coverage and elevated mortality risks. Employment, socioeconomic status, and ethnicity emerged as key predictors of vaccination propensity and mortality, disproportionately disadvantaging vulnerable groups.

Conclusions These results stress the need for equitable vaccine distribution and targeted interventions to address disparities and enhance public health outcomes.

Keywords Endogenous selection, Bayesian model, COVID-19 pandemic, Vaccine, Spatiotemporal analysis

*Correspondence:
Maria A. Barceló
antonia.barcelo@udg.edu
[1]Research Group on Statistics, Econometrics and Health (GRECS), University of Girona, Girona, Spain
[2]Unitat de Suport a la Recerca Gerència d'Atenció Primària i a la Comunitat de la Catalunya Central, Institut Català de la Salut, Manresa, Barcelona, Spain
[3]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, Instituto de Salud Carlos III, Madrid, Spain
[4]Departamento de Estadística, Universidad Nacional de Colombia, Medellín, Colombia
[5]Carrer de la Universitat de Girona 10, Campus de Montilivi, 17003 Girona, Spain

Introduction

The COVID-19 pandemic has highlighted pre-existing health inequalities, particularly for those subjects or groups living in more socioeconomically disadvantaged areas [1]. Many studies have found that areas with higher deprivation indices and subjects with lower socioeconomic status had a higher incidence of infection and more severe COVID-19 outcomes, including higher mortality [2–4]. These populations were also more likely to exhibit lower vaccination coverage rates [5].

In addition to socioeconomic factors, environmental exposures - particularly meteorological variables and air pollutants - have also been suggested as potential risk factors for COVID-19 outcomes such as infection, hospitalization, Intensive Care Unit (ICU) admission, and/or mortality. Many studies have analysed these associations (see systematic review [1]), but most present methodological limitations that prevent and solid conclusions from being drawn. Even studies with fewer limitations have not consistently shown a protective or attenuating effect of meteorological factors, contrary to initial hypotheses proposed during the early stages of the pandemic (see systematic reviews [6, 7] and meta-analysis [8]). In contrast, research on air pollution has yielded more consistent results, with higher exposure levels associated with increased COVID-19 incidence and mortality. However, air pollutants often act as proxies for other underlying variables such as residential mobility patterns or socioeconomic conditions - including overcrowded housing, occupations that do not permit remote work, and reliance on public transportation [9, 10].

Socioeconomic inequalities also appear to play a crucial in understanding disparities in vaccination coverage. Access barriers such as the distance to vaccination sites, inflexible work schedules, lack of internet access, or the absence of paid leave disproportionately affect low-SES populations [11, 12]. In a multinational study covering 14 countries, Arsenault et al. [13] found that healthcare utilisation, having a regular provider, and receiving preventive services (indicators of higher SES) were positively associated with vaccination uptake. Conversely, unmet healthcare needs and discrimination, more common among disadvantaged populations, reduced vaccination likelihood. Institutional trust also plays a role: Lansford et al. [14] reported that confidence in government responses correlates with higher vaccination and lower hesitancy, while Cantet et al. [15] showed that trust in institutions such as the police, military and church, as well as perceived infection risk, influenced vaccination behaviour in rural Colombia. Moreover, belief in conspiracy theories significantly contributed to vaccine refusal, as demonstrated by Caycho-Rodríguez et al. [16] across several Andean countries.

Latin America and the Caribbean accounted for roughly 25% of global COVID-19 deaths as of December 2023, despite comprising only 8% of the world's population [17]. While some countries achieved high vaccination coverage (Chile 90.80%, Brazil 86.75%, Peru 85.15%, Ecuador 80.20%), the regional median for complete vaccination stood at just 57% (Q1: 48%, Q3: 81%), with Venezuela, Bolivia, Paraguay, Guyana and Suriname lagging behind [17].

Colombia experienced one of South America's highest COVID-19 burdens, ranking third in incidence and sixth in mortality [17–19]. By June 21, 2023, it had reported 6,371,090 confirmed cases (13,204 per 100,000 inhabitants) and 142,794 deaths (296 per 100,000)25. By April 2023, complete vaccination coverage reached 71.73%, placing Colombia 11th among 28 Latin American countries [17]. This moderate coverage may be linked to social unrest in 2021 and challenges in vaccine access [18, 19].

Cali, the capital of the Valle del Cauca department, is the third-largest population in Colombia, with 2,205,615 inhabitants in 2018 (after Bogotá with 7,412,566 inhabitants and Medellín with 2,533,424), a density of 3,936.50 inhabitants per km^2 in the municipal area and 18,243.3 inhabitants/km^2 in the urban area [20]. The urban area of the city is administratively organized into 22 *'comunas'* which, in turn, are divided into neighbourhoods. In total, Cali has 332 neighbourhoods. *Comuna* 19 contains the greatest number of neighbourhoods (33 neighbourhoods) while *comuna* 1 has the fewest (1 neighbourhood).

In June 2023, although Cali had a cumulative incidence (18,442 per 100,000 inhabitants) and mortality (405 deaths per 100,000 inhabitants) that was lower than the other two largest cities in Colombia (cumulative incidence equal to 25,472 in Bogotá and 21,741 in Medellín; mortality equal to 409 deaths in both Bogotá and Medellín, all figures per 100,000 inhabitants), it was considerably higher than the figures for Colombia [21]. That said, the fatality rate in Cali was much higher (2.20%) than in Bogotá and Medellín (1.60% and 1.88%, respectively) and quite like that of Colombia as a whole (2.24%) [21]. This discrepancy is likely explained by the greater inequality to be found in Cali than in other cities in Colombia. To that effect, the urban area of Cali has the highest percentage of Afro-Colombian population (30.80% compared to 10.70% in the rest of Valle del Cauca), by far exceeding any other regions in Colombia, including the departments of Bolívar (11.50%) and Chocó (8.90%) which are traditionally areas of Afro-Colombian population [22]. On the other hand, more than half of Cali's population is concentrated in low socioeconomic neighbourhoods located mainly in the east, northeast, and west, and almost a tenth of the population below the poverty line is concentrated in the eastern neighbourhoods of the city [21].

In Cali, the pandemic initially affected higher-SES areas but later shifted to densely populated, low-income neighbourhoods with high levels of informal employment [23]. Two previous studies using SIVIGILA data analysed spatiotemporal incidence: Dong et al. [24] applied a non-stationary point process model with a neural network-based kernel, while Arango-Londoño et al. [25] used a Besag-York-Mollié (BYM) model [26] to examine neighbourhood-level disparities. Both studies focused on incidence and assumed that missing data, such as socioeconomic indicators at the *comuna* level (an administrative subdivision of a city, roughly equivalent to a district or borough), were random, potentially overlooking selection bias arising from missing addresses.

Against this backdrop, our study analyses the spatiotemporal variability of COVID-19 vaccination and mortality in Santiago de Cali, Colombia (hereinafter Cali), a city marked by pronounced socioeconomic inequalities.

Most importantly, our study makes a novel methodological contribution by integrating bias correction for endogenous selection into a Bayesian spatiotemporal modelling framework. Many individuals lacked residential address information, and this missingness is likely related to sociodemographic and socioeconomic factors that also influence COVID-19 outcomes. Failure to account for this results in inconsistent estimators [27, 28]. Unlike previous research [24, 25], we explicitly model the probability of being observed (i.e. of reporting a residential address) and use this to obtain unbiased and consistent estimates. By combining spatiotemporal modelling with bias correction, we provide more reliable and spatially nuanced estimates of how socioeconomic status, occupation and ethnicity affect COVID-19 outcomes.

The primary objective of this study is to examine the spatiotemporal distribution of COVID-19 vaccination in Cali while explicitly accounting for endogenous selection bias. A secondary objective is to evaluate the protective effect of vaccination against COVID-19-related mortality. By addressing a critical methodological gap and applying advanced spatial modelling techniques, this research offers new insights and contributes to the development of more accurate and equitable public health strategies.

Our results provide clear evidence of endogenous selection in vaccination patterns, underscoring the importance of accounting for this bias in epidemiological analyses. We observed significant socioeconomic inequalities in vaccination coverage levels, with higher socioeconomic strata demonstrating better coverage (including booster doses and complete regimens) while lower strata showed reduced coverage and greater mortality rates. After implementing corrections for selection bias, our models confirmed that occupation, socioeconomic status, and ethnicity significantly influenced both vaccination probability and mortality risk, with particularly adverse effects on vulnerable population groups.

Our study differs from previous research in several important aspects. First, it covers a more extensive observation period, from March 2020 to May 2022. Second, it examines both vaccination patterns and mortality outcomes, rather than focusing solely on incidence. Third, and most crucially, this study makes an original methodological contribution by integrating bias correction for endogenous selection into a Bayesian spatiotemporal modelling framework. By explicitly modelling the probability of observation (reporting a residential address) alongside health outcomes, we correct for non-random missingness—a key but often overlooked source of bias in health geography studies—and thus produce more reliable and spatially nuanced estimates of COVID-19 outcomes. By combining spatiotemporal modelling with bias correction, we provide more reliable estimates of how SES, occupation and ethnicity affect COVID-19 outcomes.

Methods

Study design and setting

We used a mixed longitudinal design with the unit of analysis being the individual. Specifically, our study population consisted of residents in Cali who were diagnosed with COVID-19, and officially registered, between March 11, 2020, and May 19, 2022; shortly before the Colombian government declared the end of the COVID-19 Health Emergency on July 1, 2022.

We used individual-level data from Colombia's National Public Health Surveillance System (SIVIGILA) [29], collected daily by Cali's Municipal Public Health Secretariat [30]. The dataset includes laboratory-confirmed COVID-19 cases (defined via PCR or antigen test), diagnosis dates and geographic coordinates of residence.

In Cali, health facilities, laboratories and local health authorities are legally mandated to report confirmed COVID-19 cases to the Municipal Public Health Secretariat, which in turn transmits the information to the National Institute of Health via the SIVIGILA platform. Data are collected using standardised forms and protocols that include demographic, clinical and geographic information, and undergo routine quality checks before integration into the national database.

After excluding 233 records corresponding to subjects who did not have the date of their COVID-19 diagnosis, and 830 records corresponding to subjects who declared they did not reside in Cali, we were left with a sample of 380,562 subjects [31]. This figure represented 99% of the 385,529 confirmed positive cases of COVID-19 between March 2020 and May 2022, according to the Behavioural Surveillance Group for Events of Public Health Interest of the Cali Municipal Public Health Secretariat [32].

Variables

Outcome variables

We considered two sets of outcome variables: vaccination variables and deaths. As vaccination variables we considered: (i) only one dose: if the subject received a single dose of the vaccine (except Janssen/J&J); (ii) full-schedule: if the subject received two doses of Pfizer-BioNTech, Moderna, Oxford-Astra-Zeneca and Sinovac-Coronavac and one dose of Janssen/J&J; and (iii) booster: whether the subject received a third or fourth booster dose. When modelling, we did not distinguish between the different COVID-19 vaccines administered in Colombia.

We also analysed deaths, that is, if the subject had died during the study period, but without distinguishing the actual cause of death. In this case, we excluded those subjects who died within a period of less than one month after their first vaccination.

Note that, in turn, the vaccination variables, in addition to being output, are the exposure variables for the mortality output.

Explanatory variables

The original database contained 131 variables. For the present study, we implemented a systematic selection process to identify those most relevant for analysing vaccination and mortality outcomes. We prioritised variables capturing key domains of public health relevance, including comorbidities, timing of health events, sociodemographic characteristics, and residential location. This approach ensured that the variables retained were directly aligned with our study objectives and provided a clear basis for modelling both vaccination uptake and COVID-19-related mortality.

We included explanatory variables at the individual level and at the contextual level (neighbourhood).

At the individual level:

- Sex (male – reference category-, female).
- Age (categorized as under 15 years old, 15–24 years old – under 25 years old was the reference category-, 25–34 years old, 35–44 years old, 45–54 years old, 55–64 years old, 65–74 years old and 75 years old or older).
- Comorbidities: (presence or not - reference - of any comorbidity) hypertension, diabetes, asthma, cardiovascular disease, chronic obstructive pulmonary disease (COPD), cancer, kidney failure, hypothyroidism; obesity (no -reference category-, yes); smoker (no – reference category-, yes).
- Type of insurance (public or private - reference category-, SISBEN).
- SISBEN (acronym of '*Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales*' in Spanish) [33] is a fully-subsidized public health insurance for populations in extreme or moderate poverty, and/or vulnerable populations. The public regime is a contribution-based insurance for the general population not covered by the SISBEN but partially funded by the state. The private regime is a voluntary and fully-paid insurance additional to the public coverage.
- Occupation (elementary jobs/services - reference category-, professionals, administrative assistant, service/sales workers, technicians/associated professionals, manual workers, students, armed forces/police/security/protection, retired, unemployed).
- Ethnicity (White - reference category-, Afro-Colombian Raizal/Indigenous/Gypsy).

All individual-level data were obtained from SIVIGILA [29].

At a contextual (neighbourhood) level:

- Socioeconomic strata in which the subject lives (1 Low - reference category-, 2 Middle/low, 3 Middle, 4 Middle-high, 5 High, 6 High-high).
- In Colombia, the socioeconomic stratification system is a classification method used by the government to categorize residential properties based on physical and environmental characteristics. It assigns households to one of six strata (1 to 6), where Stratum 1 represents the lowest-income and most disadvantaged households, and Stratum 6 includes the most affluent. This system is primarily used to allocate subsidies and set utility rates, ensuring that lower-income households receive financial support for basic services like water, electricity, and gas.
- The socioeconomic strata variable classifies residential properties into strata to differentially charge home public services, where those with higher economic capacity pay more for public services and thus contribute to lower strata receiving subsidies on their bills (classification in six groups, where strata 1 corresponds to the most disadvantaged and strata 6 to the least disadvantaged) [20, 34].

Information of the socioeconomic strata of the subject was obtained from SIVIGILA [29].

Variables at the neighbourhood level where the subject resided.

- Unmet basic needs index.
- The unmet basic needs index (UBN) is a multidimensional poverty measure developed and applied by DANE (*Departamento Administrativo Nacional de Estadística*) to identify households in conditions of structural poverty based on the deprivation of fundamental living conditions [34, 35]. The index classifies a household as 'poor' if it meets at least one unsatisfied basic need across five specific dimensions: Housing quality and materials; Housing overcrowding; Access to basic public services; School attendance; and Economic dependency. If a household experiences one or more of these

conditions, it is counted as having unsatisfied basic needs. The index is widely used in Colombia for the design and evaluation of social policies, particularly for targeting vulnerable populations.

- Percentage of subjects aged 20 to 59 years.
- Percentage of subjects aged 70 to 79 years.
- Percentage of subjects aged 80 or older.
- Percentage of Afro-Colombians.
- Proportion of people over 60 years of age who live in single-person households.
- Proportion of households in overcrowded bedrooms.
- Population density of the block.

All variables at the neighbourhood level where the subject resided were obtained from the microdata of the 2018 National Population and Housing Census for Colombia [20], the most recent at the time of writing this paper.

Prior to model fitting, we assessed collinearity using pairwise correlations and examined its impact on model stability. This assessment revealed severe collinearity among several contextual variables (see Table S1 and Figure S1 in the Supplementary Information), which resulted in unstable parameter estimates and, in some instances, changes in coefficient sign.

To mitigate these issues, we adopted a parsimonious modelling strategy and retained only those contextual variables that exhibited lower mutual correlation and produced stable estimates when jointly included. The final set comprised neighbourhood socioeconomic strata, the percentage of the neighbourhood population aged 20–59 years, and the percentage of the population aged 60 years or older living in single-person households.

Although unmet basic needs is a commonly used socioeconomic indicator, it is a composite measure that was highly correlated with most other contextual variables considered, capturing much of their shared variability. Its inclusion alongside those variables did not contribute additional independent information and substantially exacerbated collinearity; therefore, it was excluded from the final models.

The percentage of the neighbourhood population aged 20–59 years, and the percentage of the population aged 60 years or older living in single-person households were categorized into quartiles and we took the first quartile as the reference category. For socioeconomic strata, the reference category was 'Low'.

Bias

We did not have the residential addresses of a third of the subjects (125,294). This implies that there was a clear risk of information bias if this missingness was systematically related to socioeconomic or health factors. In fact, the lack of information explaining the unavailability of the residential data suggested reasonable suspicion that it was not missing at random. To investigate this further by using the data available, the cumulative incidence of COVID-19, per 100,000 inhabitants, rate by *comuna* from March 2020 to May 2022 (Fig. 1) was compared with that reported for the same period by Cali's Health Department in its epidemiological reports [30]. It was observed that *comuna* 21 was located in a lower quartile than reported by Cali's Health Department, while *comunas* 6 and 10 were located in higher quartiles. All three *comunas* misplaced in the observed data were in neighbourhoods mainly characterized by low economic conditions (strata 1 to 3 on the map on the right of Fig. 1). That is, the absence of residential address data among subjects could not be randomly distributed and, consequently, certain unobserved factors that might account for the lack of this specific information among subjects could be correlated with the outcomes.

To correct for this endogenous selection bias, we used a two-part model [27, 28]. Our two-part model is conceptually similar to Heckman's selection models (but within a Bayesian framework). In the first part we estimated the probability that a subject had been observed (that is, reported a residential address). These probabilities were then used as weights in the second part of the model where we estimated the probability of occurrence of the outcome variables.

That is to say,

$$weight_i = \frac{e^{\widehat{\eta}_i}}{1 + e^{\widehat{\eta}_i}}$$

where i denoted subject; and $\widehat{\eta}_i$ was the linear predictor of the first part of the model. That is, the weights are the probabilities (estimated in the first part of the model) that a subject was present in the sample (i.e. that there was information about their address).

Then, the likelihood of the second part of the model was defined as:

$$\begin{aligned}&Prob\,(y_i|\ldots)\\&= weight_i 1_{[y=0]} + (1 - weight_i)\\&xBinomial\,(y_i|y_i > 0)\end{aligned}$$

where y_i denoted the dependent variable of the second part of the model.

These two parts were estimated simultaneously in a Bayesian model [28] based on the Integrated Nested Laplace Approximation (INLA) [36–38]. The idea is to use the facilities of this approach to model a multivariate response variable composed of two processes. Each of the processes has its linear predictor that may or may

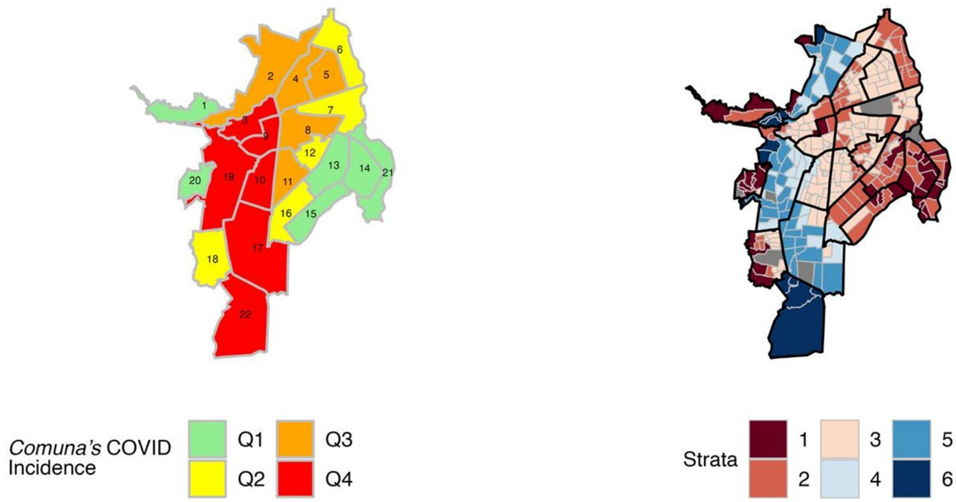

Fig. 1 Cumulative incidence of COVID-19, per 100,000 inhabitants, between March 2020 and May 2022 by *'comuna'* for subjects with available residential address (left) and strata of the neighbourhood (right). The map on the left identifies each '*comuna*' (with numbers). In the map on the left, the quartiles of the distribution of the cumulative incidence are represented in different colours (Q1 first quartile, Q2 second quartile, Q3 third quartile, Q4 fourth quartile). The socioeconomic strata, represented on the map to the right, is categorized into six groups, where strata 1 corresponds to the most disadvantaged and strata 6 to the least disadvantaged. Own construction based on data from SIVIGILA (incidence of COVID-19) and DANE (socioeconomic strata). SIVIGILA. Sistema Nacional de Vigilancia en Salud Pública [in Spanish] [Available at: https://portalsivigila.ins.gov.co, last accessed on March 28, 2024]. DANE. Departamento Administrativo Nacional de Estadística. Unmet basic needs [Available at: https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/necesidades-basicas-insatisfechas-nbi, last accessed on December 16, 2024]

not share explanatory variables, but what is important is that they share a random effect that includes those unobservable factors that would explain both probabilities (of reporting the residence and of the occurrence of the outcome variable).

Statistical methods

Maps of risk

To evaluate the existence of a geographical pattern in vaccination, we first represented the smoothed odds (sOdds) on a map of Cali. In particular, we smoothed the only-one dose odds (probability of only-one dose versus no vaccination); full-schedule (probability of complete regimen versus only-one dose); and booster (probability of booster versus that of full-schedule). To estimate the sOdds, we specified a generalized linear mixed model (GLMM) with a binomial link (equivalent to a logistic regression), without including explanatory variables, but controlling for extra variability by including various random effects. These random effects captured individual, neighbourhood and temporal dependencies and were incorporated simultaneously from the outset of the model specification, rather than being entered sequentially.

We assumed that conditional to the true risk for subject i, the occurrence of the vaccination outcome variables ($Y_i;\ 0\,no,\ 1\,yes$) was distributed as a binomial,

$$\log\left(\frac{Prob\,(Y_i=1)}{1-Prob\,(Y_i=1)}\right)$$
$$=\beta_{\,0}+\ \nu_{\,i}+S\,(neighbourhood_i)+\tau_{\,t}$$

where $\frac{Prob(Y_i=1)}{1-Prob(Y_i=1)}$ were the odds.

We included three random effects in the models.

The first random effects, $\nu_{\,i}$, were indexed on the subject. These random effects were unstructured (independent and identically distributed random effects), and captured individual heterogeneity, i.e., unobserved confounders specific to the subject and invariant in time.

We also included structured random effects to control spatial dependency, $S\,(neighbourhood_i)$, where neighborhood was the neighborhood of the subject's residence. That is to say, the fact that small areas that are close in space show more similar values of the outcome variables than areas that are not close.

To control the temporal dependency we used structured random effects (random walk of order one) indexed on time (week in which the subject was vaccinated), $\tau_{\,t}$. Following the INLA approach [36–38] when, as in this case, the random effects are indexed on a continuous variable, they can be used as smoothers to model non-linear dependency on covariates in the linear predictor.

Random effects were defined using a multivariate Gaussian distribution with a zero mean and precision matrix kΣ, where *k* was a constant, and Σ was a matrix that defined the dependence structure of the random effects [36–38]. In unstructured random effects, Σ was a diagonal matrix of 1s; and in random walk random effects, Σ was defined assuming that increments (in rw1, $\Delta u_i = u_t - u_{t-1}$) followed a Gaussian distribution with a zero mean and a constant precision *k* [38].

The spatially structured random effect S was normally distributed with a zero mean and a Matérn covariance function [39]:

$$Cov\left(S\left(x_i\right), S\left(x_{i'}\right)\right) = \frac{\sigma^2}{2^{\nu-1}\Gamma\left(\nu\right)}\left(\kappa\left\|x_i - x_{i'}\right\|\right)^{\nu} \mathrm{K}_{\nu}\left(\kappa\left\|x_i - x_{i'}\right\|\right)$$

where K_{ν} is the modified Bessel function of the second type and order $\nu > 0$. ν is a smoother parameter, σ^2 is the variance, and $\kappa > 0$ is related to the range ($\rho = \sqrt{8\nu}/\kappa$), the distance to which the spatial correlation is close to 0.1.

We represented the sOdds on the map of Cali by neighbourhoods in two different subperiods: until August 31, 2021, which corresponds to Phase 5 of the vaccination scheme in Colombia [40] (the entire population aged 12 years or older was allowed to be vaccinated); and from August 31, 2021 until the end of the period studied (on November 24, 2021, Phase 1 of booster doses began).

We also computed exceedance probabilities [41] which, in our case, were the probability that the smoothed odds were above the median of the accumulated percentage in each of the weeks. Richardson et al.[41] recommend using as a specific interpretation rule the cut-off 80% (and 20%). This cut-off can be used to help assess the existence of agglomerations of excess (deficiency) cases (i.e., clusters). The exceedance probabilities were also represented on a map of the study area.

Two-part model

We specified a Bayesian two-part model based on INLA for each outcome variable. Specifically, the two-part model was composed of two GLMMs with binomial link functions (equivalent to logistic regressions).

First part:

$$\log\left(\frac{Prob\left(residence_i = 1\right)}{1 - Prob\left(residence_i = 1\right)}\right) = \eta\_first\_part_i + \nu_i + S\left(neighbourhood_i\right)$$

Second part:

$$\log\left(Prob(Y\_i = 1)/(1 - Prob(Y\_i = 1))\right) = \eta\_second\_part\_i + \nu\_i + S(neighbourhood\_i)' + \tau\_t$$

where $residence_i$ was the indicator that we had the subject's residence (0 no, 1 yes); Y_i denoted one of the four dependent variables, the vaccination outcomes variables (only one dose; full-schedule; booster) or the death outcome variable; and the linear predictor of the first and the second part. Note that, the random effects that capture spatial dependence (the distribution of risk between neighbourhoods) did not coincide in the two parts. However, individual heterogeneity (unobserved confounding) was a shared random effect (i.e., the same for both parts).

The linear predictor of the first part included the variables at the individual level, whereas the linear predictor of the second part had variables at the individual and contextual levels. When the outcome variable was death, the linear predictor of the second part also included the vaccination variable (if the subject received any dose of the vaccine: not vaccinated - reference category -; only one-dose; only full-schedule; and booster).

We indexed the structured random effects that control the temporal dependency by the week in which the subject was vaccinated (vaccination variables) and by the week in which the subject died (death).

As we mentioned above, it should be noted that our two-part specification is implemented as a single Bayesian model rather than as two sequential steps. Both the probability of an individual being observed (reporting a residential address) and the probability of the outcome of interest (vaccination or mortality) are estimated jointly within the same Integrated Nested Laplace Approximation (INLA) framework. This joint estimation allows the two components to share random effects, capturing latent factors that influence both processes and thereby correcting for endogenous selection bias. By integrating the two parts into one coherent model, we obtain consistent and spatially nuanced estimates while fully accounting for the non-random nature of the missing residential data.

Inference

Inferences were made following a Bayesian perspective, using the INLA approach [36–38]. We used priors that penalize complexity (called PC priors). These priors are robust in the sense that they do not have an impact on the results and, in addition, they have an epidemiological interpretation [42].

All analyses were carried out using the open-source software R (version 4.3.1) [43], through the INLA package [36–38, 44] in the experimental mode [45]. The maps were represented using the *tmap* package [46].

Results

Descriptive analyses

Most contextual socioeconomic variables were highly correlated with neighbourhood strata, in particular, unmet basic needs index.

Descriptive statistics of the variables are provided in Table 1. Among the subjects, 27.40% were unvaccinated,

Table 1 Descriptive of variables selected for analysis. Cali, Colombia, March 2020-May 2022

	Missings[1]	*n* (%) (*N* = 380,562)
Vaccine schedule progress	0%	
[Unvaccinated]		104,172 (27·4%)
Only one dose		39,529 (10·4%)
Full-schedule		138,250 (36·3%)
Booster		98,611 (25·9%)
Died [No]: Yes	0%	8,599 (2·3%)
Hypertension [No]: Yes	0%	22,350 (5·9%)
Diabetes [No]: Yes	0%	10,354 (2·7%)
Obesity [No]: Yes	0%	6,559 (1·7%)
Asthma [No]: Yes	0%	5,673 (1·5%)
Hypothyroidism [No]: Yes	0%	3,638 (1·0%)
Cardiovascular disease [No]: Yes	0%	3,634 (1·0%)
COPD [No]: Yes	0%	2,304 (0·6%)
Cancer [No]: Yes	0%	2,126 (0·6%)
Kidney failure [No]: Yes	0%	1,758 (0·5%)
Smoker [No]: Yes	0%	3,300 (0·9%)
Sex [No]: Female	0%	204,542 (53·7%)
Age	0%	
[Less than 15]		21,035 (5·5%)
[15–24]		48,706 (12·8%)
25–34		88,210 (23·2%)
35–44		75,350 (19·8%)
45–54		57,236 (15·0%)
55–64		45,432 (11·9%)
65–74		24,828 (6·5%)
75 or more		19,765 (5·2%)
Socioeconomic strata[2]	34.40%	
[Low]		26,660 (10·7%)
Middle-low		60,554 (24·3%)
Middle		95,984 (38·4%)
Middle-high		29,138 (11·7%)
High		28,433 (11·4%)
High-high		8,895 (3·6%)
Occupation	47.32%	
[Elementary Jobs – Services]		72,928 (36·4%)
Professionals		37,586 (18·7%)
Administrative Assistant		23,069 (11·5%)
Service And Sales Workers		20,694 (10·3%)
Technicians		15,622 (7·8%)
Manual Workers		9,862 (4·9%)
Student		8,339 (4·2%)
Armed Forces and Police		5,545 (2·8%)
Retired		4,595 (2·3%)
Unemployed		2,245 (1·1%)
Ethnicity	31.13%	
[Non-ethnic]		257,638 (98·3%)
Afro-Colombian		2,921 (1·1%)
Raizal, Indigenous or Gypsy		1,534 (0·6%)

[1] Percentage of missing data

[2] In Colombia, the socioeconomic stratification system is a classification method used by the government to categorize residential properties based on physical and environmental characteristics. It assigns households to one of six strata (1 to 6), where Stratum 1 represents the lowest-income and most disadvantaged households, and Stratum 6 includes the most affluent. This system is primarily used to allocate subsidies and set utility rates, ensuring that lower-income households receive financial support for basic services like water, electricity, and gas

Reference categories in brackets

10.40% received only one dose, and 62.20% received full-schedule (36.30%) or booster (25.90%). The fatality rate was 2.30%. The most prevalent chronic condition was high blood pressure (5.90%), followed by diabetes (2.70%), obesity (1.70%) and asthma (1.50%). In terms of age, 61.30% were under 45 years of age, 18.30% under 25, and only 11.70% were 75 or older.

Regarding socioeconomic variables, only 52.00% of the subjects provided information about their occupation. The most frequent occupation was elementary jobs -services (36.40%), followed by professionals (18.70%), administrative assistants (11.50%) and services and sales workers (10.30%). Only 1.00% of the subjects identified themselves as Afro-Colombian and even fewer identified as Raizal, Indigenous or Gypsy. Subjects living in low strata neighbourhoods accounted for 10.70% while 24.00% lived in middle-low strata. As shown in Figure S1 of the supplementary material, when looking at the context of the neighbourhoods in each stratum by the UBN quartiles, the lower the strata, the more prominent households with high UBN index became, with the most populous being the middle strata (38.00% of all subjects), while the three better economically established strata encompassed 25.00% of the subjects.

In Fig. 2, the figure at the top shows the evolution of the onset vaccination status of subjects and how this changed through time. In mid-February 2021, the first subjects with at least one dose of the vaccine appear, and then in September 2021 subjects with booster doses. Note that, death behaviour changed through time. It is worth mentioning the observed link in spikes of unvaccinated people tested for COVID-19 and the death spikes in December 2021 and January 2022. In December 2020, vaccines were administered to priority groups of health workers and the elderly, and that large scale roll-out was launched in February 2021 (see Figure S2 in supplemental information).

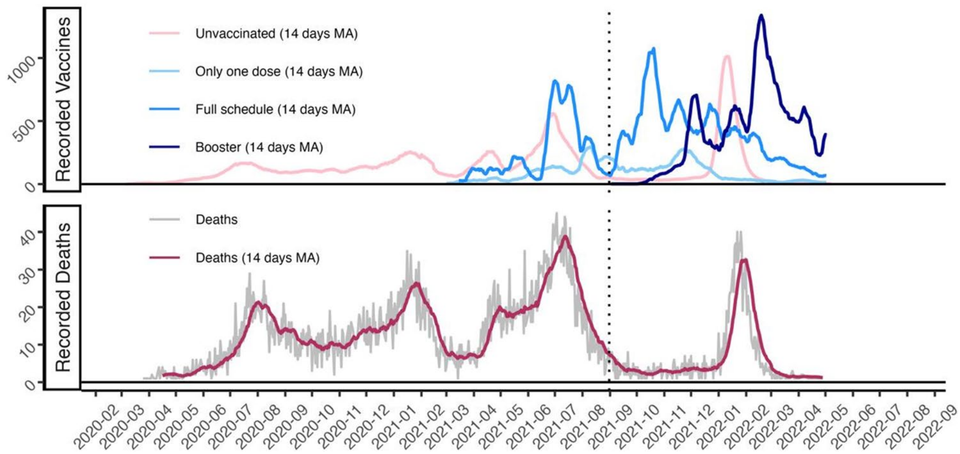

Fig. 2 Daily register of subjects' onset vaccination status and daily recorded deaths. On the x-axis, the vaccination date (for those vaccinated) and the date of the COVID-19 test (for those unvaccinated). The vertical dotted line indicates date of vaccine roll-out to the general public. "14 days MA" denotes a 14-day moving average, which is used to smooth the temporal evolution of the variable, making it easier to observe trends and patterns. Own construction based on data from SIVIGILA. SIVIGILA. Sistema Nacional de Vigilancia en Salud Pública [in Spanish] [Available at: https://portalsivigila.ins.gov.co, last accessed on March 28, 2024]

Maps of smoothed odds

Smoothed odds maps help us to highlight geographical zones where the outcome event has high and low likelihoods of occurrence. In Fig. 3, the maps for the third subperiod of interest (November 2021 – May 2022) indicate that for single-dose vaccination the odds (i.e. the vaccination likelihood) were higher in *comunas* of predominately lower strata, (numbers 4, 14 and 18), located in the north, southwest and east, respectively (see Fig. 1 right), whilst higher strata *comunas* had lower odds of single-dose status. Full-schedule vaccination displayed higher odds in middle income *comunas*, around the centre of the city, however, the odds seemed evenly distributed across the metropolitan area. For booster shots, there was a predominant spread of higher odds across the whole city, however, mostly higher income *comunas* have distinctively higher odds (numbers 2, 19, 18 and 22). The findings in the smoothed odds maps were further analysed with the exceedance probabilities maps which are displayed at the bottom of Fig. 3. The previously mentioned *comunas*, for each of their corresponding outcomes, hold neighbourhoods with probabilities above 50.00% of exceeding the odds of the median for the whole city.

Validation of model estimation results

In Figures S3 in the Supplemental Information, we provide a residual analysis of the models. When, as in our case, the models are equivalent to logistic regressions, the graphs are interpreted as follows [47]: If the left-hand graphs (fitted vs. deviance residuals) display two parallel lines without curvature, the model does not exhibit specification errors. When the right-hand graphs (fitted vs. square root of the deviance residuals in absolute value) show two intersecting lines without curvature, the model residuals will be homoscedastic.

Overall, we observe that the models perform reasonably well, although some curvature is present, particularly in the right-hand graphs, thus highlighting some heteroskedasticity issues. Heteroskedasticity is more pronounced (in relative terms) in the first-part model and in the only-one-dose model. We believe this indicates that there are other explanatory variables not captured in the model. In any case, the curvatures are not severe enough to invalidate the models.

First part of the model

The results of the estimation of the first part of the model can be seen in Table S2 in the supplemental information. As can be observed, the probability of reporting the residential address depends on almost all the variables. Women were more likely than men to report their residential address (OR = 1.03, 95%CrI: 1.01–1.06, when the outcome in the second part of the model was only one dose of the vaccine; OR = 1.05, 95%CrI: 1.03–1.07, when it was full schedule, OR = 1.06, 95%CrI: 1.03–1.09, in the case of booster and also in the case of death; in all cases the credible intervals at 95% did not contain the unity). Subjects under 25 years of age were more likely to report residential information. Suffering from any of the comorbidities recorded also significantly increased the odds of residential information records. Elementary service workers have the least odds of having the residential address compared to all other occupations. Being of an ethnicity other than white implied a greater propensity of reporting one's residential address.

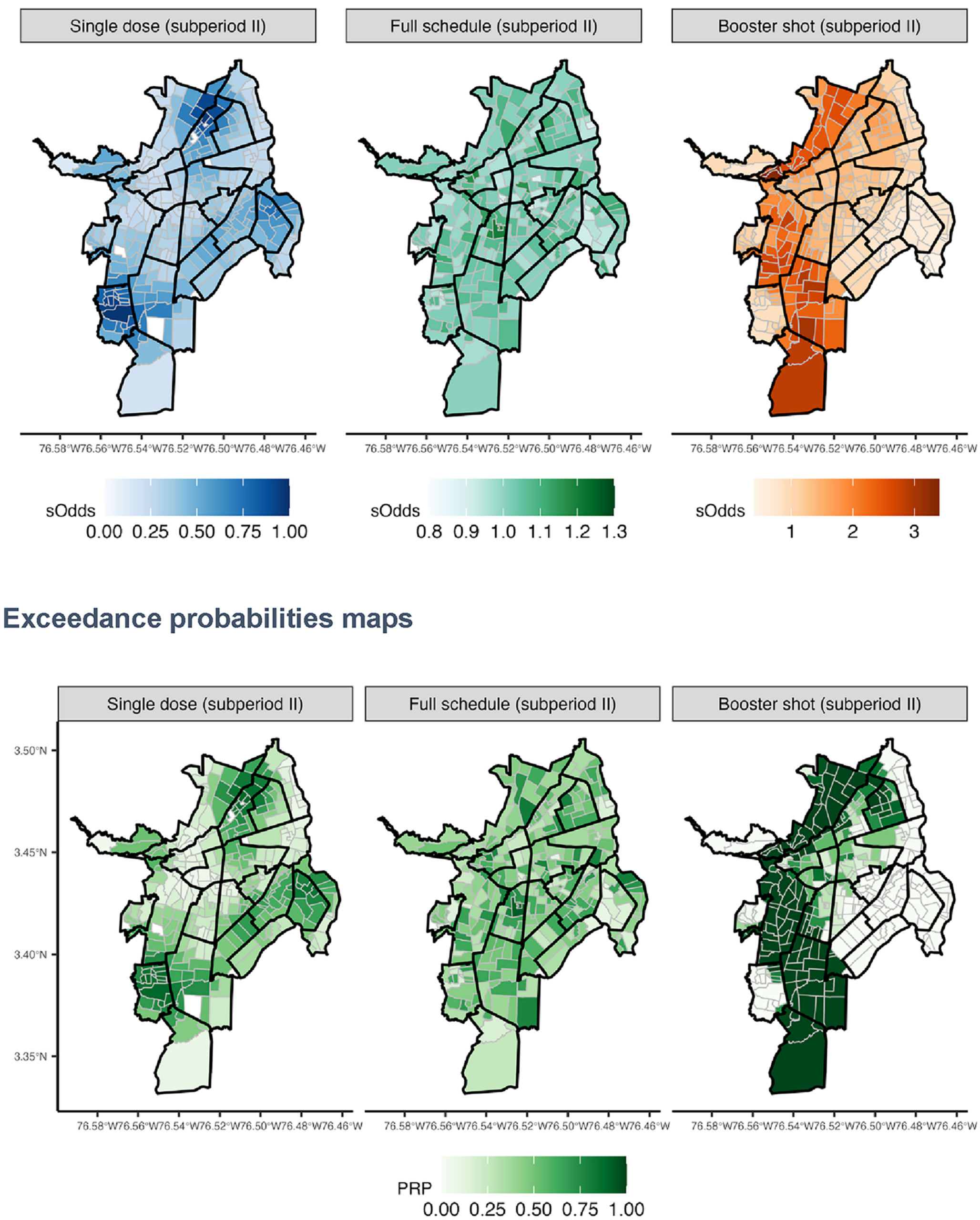

Fig. 3 Smoothed odds for vaccination outcome in Cali, Colombia, in subperiod II, August 2021 to May 2022. represents the smoothed odds of each vaccination dose, defined as follows: Only-one dose: The probability of receiving only-one dose versus no vaccination. Full-schedule: The probability of complete regimen versus only-one dose. Booster: The probability of booster dose versus full-vaccination schedule. Exceedance probabilities maps: PRP Exceedance probability. Probability that the smoothed odds were above the median of the accumulated percentage in each of the weeks. On the abscissas of maps, longitude is indicated, and on the ordinates, latitude is indicated. Own construction. The full neighbourhood-level tables underlying the maps are available from the authors upon reasonable request

Second part of the model

Vaccination outcomes

Table 2 shows the results of the estimation of the second part of the model for vaccination and death outcomes.

Being a woman implied a lower propensity to have received only one dose of vaccine (OR = 0.76, 95% credible interval, 95%CrI:0.71 – 0.18), but a greater propensity to receive only full-schedule (OR = 1.12, 95%CrI = 1.07–1.18) and booster (OR = 1.15, 95%CrI = 1.05–1.11). In both cases the reference category was those who received only dose. Again, in all cases the credible interval at 95% did not contain the unity.

Subjects 24 years of age or older had a greater propensity to have received only full-schedule and booster doses than those under 25 years of age (reference category). In the case of only one dose, only subjects aged 44 years and

Table 2 Results of the Estimation of the second part of the model for vaccination and death outcomes

Variable	Category	Only one dose OR	Full-schedule OR	Booster OR	Death OR[1]
Vaccine schedule progress [Unvaccinated]	Only one dose				0.13**
	Full schedule				0.10**
	Booster				0.03**
Sex [Male]	Female	0.76**	1.12**	1.15**	0.36**
Age [Less than 25]	25–34	0.94	1.38**	2.43**	0.13**
	35–44	1.04	1.74**	3.71**	0.16**
	45–54	1.48**	2.13**	7.88**	0.41**
	55–64	5.12**	2.69**	17.79**	0.74
	65–74	10.31**	2.60**	49.59**	1.63**
	75 or more	15.59**	2.30**	230.5**	3.19**
Hypertension [No]	Yes	1.51**	1.10**	1.14**	0.91
Diabetes [No]	Yes	1.14	0.96	1.16**	1.99**
Obesity [No]	Yes	1.42**	1.07*	1.00	1.91**
Hypothyroidism [No]	Yes	0.87	1.10*	1.22**	1.06
Asthma [No]	Yes	0.49**	1.09*	1.07	0.94
Cardiovascular disease [No]	Yes	0.98	0.97	1.30**	1.22
COPD [No]	Yes	1.37**	0.75**	0.88	1.55
Cancer [No]	Yes	1.31	0.89	1.09	1.88*
Kidney failure [No]	Yes	1.13	1.04	1.12	1.17
Smoker [No]	Yes	0.83	1.16**	0.81**	0.91
Socioeconomic strata [Low]	Middle-low	0.85	0.97	0.95	0.02**
	Middle	0.80	1.15**	1.02	0.04**
	Middle-high	0.62*	1.43**	1.32**	0.04**
	High	0.78	1.43**	1.46**	0.06**
	High-high	0.47*	1.34**	1.33*	0.02**
Occupation	Professionals	0.93	1.50**	3.27**	1.02
[Elementary Jobs – Services]	Administrative Assistant	0.87	1.20**	1.25**	0.42**
	Service And Sales Workers	1.23	1.08**	1.14**	1.06
	Technicians And Associated Professionals	0.77	1.17**	1.39**	0.41*
	Manual Workers	0.65*	1.04	0.79**	0.62
	Student	0.84	0.80**	0.61**	0.03**
	Armed Forces, Police, Security, Protection	0.30**	1.32**	3.31**	0.19**
	Retired	0.91	1.02	1.59**	1.21
	Unemployed	0.34**	0.99	1.20**	2.21**
Ethnicity [Non-ethnic]	Afro-Colombian	0.37**	0.85**	0.95	0.50
	Raizal, Indigenous or Gypsy	0.79	0.85*	0.87	0.25

** The credible interval at 95% did not contain the unity

* The credible interval at 90% did not contain the unity

Reference categories in brackets

Cali, Colombia, March 2020-May 2022

older had a higher propensity than those younger than 25 years. In all cases, the propensity increased with age; with the case of the booster dose standing out.

Having hypertension implied a greater propensity to being vaccinated, although the propensity to have only one dose was much greater. Meanwhile, being obese implied a greater propensity to have received only one dose, followed, by a long way, with only full-schedule. Subjects with COPD, while having a fairly high propensity to have received only one dose of the vaccine, were less likely to have received only full-schedule than for subjects without COPD. The opposite was the case for subjects with asthma, with a fairly high propensity to have received only full-schedule and a lower propensity to receive only one dose than subjects without asthma. Subjects with hypothyroidism (in full-schedule and in booster) and with cardiovascular disease and diabetes (these two only in booster) were more likely to have been vaccinated than subjects without these chronic conditions. Finally, subjects with cancer and kidney failure did not have a different propensity to being vaccinated regardless of the dose, than subjects without these diseases.

Smokers had a greater propensity than non-smokers to receive only full-schedule and a lower propensity in the case of the booster, while we found no differences in the case of only one dose.

Subjects whose occupation was in the armed forces, police, security or protection (OR = 0.30, 95%CrI = 0,21 – 0,39), followed by the unemployed (OR = 0.34, 95%CrI = 0.16–0.51) and manual workers (OR = 0.65, 95%CrI = 0.57–0.73) had a lower propensity to receive only one dose, than those employed in elementary jobs or services (reference category). The propensity to receive only one dose in other occupations was no different from those in the reference category. In descending order, professionals (OR = 1.50, 95%CrI = 1.45–1.54, only full-schedule and OR = 3.27 booster, 95%CrI = 3.11–3.43), armed forces, police, security or protection (OR = 1.31, 95%CrI = 1.25–1.37, only full-schedule and OR = 3.31, 95%CrI = 2.97–3.65, booster), administrative assistants (OR = 1.20, 95%CrI = 1.16–1.24, only full-schedule and OR = 1.25, 95%CrI = 1.18–1.32, booster), technicians and associated professionals (OR = 1.17, 95%CrI = 1.13–1.21, only full-schedule and OR = 1.39, 95%CrI = 1.30–1.48, booster) and service and sales workers (OR = 1.08, 95%CrI = 1.04–1.12, only full-schedule and OR = 1.14, 95%CrI = 1.08–1.20, booster), had a greater propensity than subjects employed in elementary jobs or services to have received only full-schedule or booster. The retired subjects (OR = 1.59, 95%CrI = 1.38–1.80) and the unemployed (OR = 1.20, 95%CrI = 1.02–1.38), although they had a greater propensity to receive a booster than the subjects employed in elementary jobs or services, they did not present differences in receiving only full-schedule. While students had a lower propensity to receive both only full-schedule (OR = 0.80, 95%CrI = 0.62–0.98) and booster (OR = 0.61, 95%CrI = 0.52–0.70), manual workers only had a lower propensity to receive booster (OR = 0.79, 95%CrI = 0.72–0.86).

Afro-Colombians were less likely than non-ethnic groups to receive only one dose (OR = 0.37, 95%CrI = 0.15–0.59) and only full-schedule (OR = 0.85, 95%CrI = 0.75–0.95). Raizal, Indigenous or Gypsy subjects only had a lower propensity than non-ethnic subjects in the case of only full-schedule (OR = 0.85, 95%CrI = 0.71–0.99).

In terms of socioeconomic stratum, we find different propensities from those in the most disadvantaged stratum (reference category) mainly in the middle-high and higher strata (and also the middle strata, although only in only full-schedule). We found for all these strata, a greater propension to receive only full-schedule (OR = 1.34, 95%CrI = 1.24–1.40 – high-high strata-, and OR = 1.43, 95%CrI = 1.36–1.54 – high and middle-high strata-; and OR = 1.15, 95%CrI = 1.07–1.23, in middle strata) and booster (OR = 1.32, 95%CrI = 1.30–1.34 – middle-high strata-; OR = 1.46, 95%CrI = 1.37–1.55 – high strata-; OR = 1.33, 95%CrI = 1.29–1.37 – high-high strata-;). Note that the propensities to receive only one dose were much lower than in the lowest socioeconomic stratum in the high-high (OR = 0.47, 95%CrI = 0.31–0.63) and middle-high (OR = 0.62, 95%CrI = 0.50–0.74) strata.

Death

It should be noted that the mortality outcome used in this analysis corresponds to all-cause deaths during the study period rather than exclusively confirmed COVID-19-related deaths. This distinction is explicitly stated in the Methods and further discussed in the Discussion section to ensure transparency about the scope of the outcome variable.

Being vaccinated against COVID-19, whatever the dose, was a protective factor against death (i.e., only one dose had an OR = 0.13, 95%CrI = 0.05–0.21; OR = 0.10, 95%CrI = 0.05–0.14, in only full-schedule; and OR = 0.03, 95%CrI = 0.01–0.05, in booster: being in all cases the reference category not getting vaccinated).

Being a woman (compared to men; under 55 years of age (compared to those under 25 years of age); being a student, being employed in the armed forces, police, security, or protection, administrative assistants, or technicians and associated professionals (all of them regarded as being engaged in elementary jobs or services), presented a lower risk of dying.

On the contrary, being 65 years old or older (OR = 1.63, 95%CrI = 1.21–2.05, in subjects aged 65 to 74 years and OR = 3.19, 95%CrI = 2.41–3.97, in subjects aged 75 years or older); having diabetes (OR = 1.99, 95%CrI = 1.13–2.84), obesity (OR = 1.91, 95%CrI = 1.06–2.76) or cancer (OR = 1.88, 95%CrI = 1.62–2.14); and being unemployed (OR = 2.21, 95%CrI = 1.49–2.93), had a greater risk of dying.

Furthermore, note that all socioeconomic strata had a much lower risk of dying (ORs between 0.02 -middle-low and high-high - and 0.06 – high-, 95%CrI between 0.002 and 0.04 – middle-low and high-high- and 0.01–0.11 – high-) than the lowest socioeconomic strata.

Discussion

The results from the first part of our model clearly show evidence of endogenous selection. We found that the probability of reporting a residential address was significantly related to almost all covariates included in the model (see Table 3). This proves that missing data were not random and that many variables affecting whether an address was reported were also linked to the outcomes we studied. Specifically, women were 3% more likely than men to provide their address, and people under 25

Table 3 Summary of the findings

General finding	Specific findings
Existence of endogenous selection	Women were 3% more likely than men to provide their residential address. Subjects under 25 years of age were more likely to provide residential information. Having any of the recorded comorbidities also significantly increased the likelihood of having residential information on record. Elementary service workers have the lowest odds of having their residential address recorded compared to all other occupations. Being of an ethnicity other than white was associated with a greater likelihood of reporting one's residential address.
There were socioeconomic inequalities in vaccination coverage against COVID-19	The highest odds (i.e. vaccination likelihood) and, especially, the highest exceedance probabilities for the booster dose, and to a lesser extent in only full vaccination schedule, occurred in the strata with the highest socioeconomic level. On the contrary, in the case of only one dose, these strata presented the lowest odds and the lowest exceedance probabilities.
Existence of socioeconomic inequalities in COVID-19 mortality	Most occupations showed a greater propensity to receive a booster or complete the full schedule compared to subjects employed in elementary jobs or services. Afro-Colombians were less likely than non-ethnic subjects to receive only full-schedule vaccinations. The middle-high and higher strata showed greater propensities to receive only full-schedule and booster, than the less economically advantaged strata. Being unemployed or living in the most economically disadvantaged socioeconomic strata implied a very high risk of dying.
Being vaccinated against COVID-19, whatever the dose, was a protective factor against death	Only one dose had an 87% lower risk of dying than not getting vaccinated Only full-schedule regime had a 90% lower risk (than not getting vaccinated) Booster dose had a 97% lower risk (than not getting vaccinated)

years old were also more likely to report this information. Additionally, having any chronic condition significantly increased the chances of address reporting, probably because these individuals were already registered in the healthcare system.

We observed an interesting pattern concerning ethnicity. Contrary to what might be expected, Afro-Colombian, Raizal, Indigenous and Roma populations were more likely to report their address than non-ethnic groups. While economic difficulties might suggest lower registration rates, this could be explained by these groups needing to register to access government subsidies and social benefits. The same explanation may apply to people with chronic conditions, who need to be registered for medical follow-up. On the other hand, workers in elementary services had the lowest probability of address reporting, showing their limited visibility in administrative systems. As far as we know, no previous study has specifically examined which sociodemographic factors determine whether people appear in health surveillance databases.

Our analysis also revealed socioeconomic inequalities in COVID-19 vaccination coverage, visible in both the smoothed odds maps and the second part of the model. The highest probabilities of receiving booster doses - and to a lesser extent, the full schedule - were concentrated in higher socioeconomic groups. Meanwhile, these groups showed lower probabilities of having only one dose, suggesting wealthier people were more likely to get fully vaccinated while poorer groups often remained with partial vaccination. These results agree with other studies showing that lower socioeconomic status creates barriers to vaccination through transportation problems, limited information access, and job-related constraints [5, 11, 12].

It should be recognised, however, that some of the observed differences in vaccination or mortality probabilities across occupational groups may reflect the influence of unmeasured confounders. In this sense, the comparatively high vaccination coverage seen among police officers could arise from job-related vaccination policies or from other contextual circumstances. In a similar way, differences in uptake between professional groups might be shaped by government restrictions or workplace measures rather than by the factors explicitly modelled.

Similarly, differences in vaccination uptake across people with existing health conditions could simply arise from their underlying health concerns rather than from external factors. For example, those with chronic respiratory illnesses may prioritize vaccination as a precaution against the heightened risks that COVID-19 could pose to their condition.

These inequalities also affected COVID-19 mortality. Our model showed that unemployed people and those living in the poorest areas had significantly higher risks of death. Afro-Colombians were less likely than non-ethnic individuals to complete vaccination, which might help explain their higher mortality risk. Most job categories showed better vaccination rates than elementary service workers, confirming that occupation is an important health determinant. Similarly, middle-high and high socioeconomic groups were much more likely to be fully vaccinated or receive boosters than the poorest groups (i.e. lower strata).

But, as we commented above, the comparatively low mortality recorded among police and military staff may also reflect underlying fitness levels and overall health rather than vaccination alone, suggesting the need for caution when attributing these differences to a single cause.

Although we could not find individual-level studies with directly comparable results, our findings are consistent with Barceló et al.'s ecological study in Catalonia [5], which found that poorer areas with more unemployment had lower vaccination rates and poorer COVID-19 outcomes including hospitalizations and deaths. Vaccination coverage was higher in areas with more elderly people and lower in densely populated areas with poor housing conditions - all indicators of socioeconomic disadvantage.

Regarding individual characteristics, women were significantly less likely than men to die from COVID-19 and more likely to be fully vaccinated or receive boosters. While people aged 25–54 had lower mortality risk than our reference group (under 25 years), older adults - especially those 75 or older (OR: 3.19) - faced much higher risks. These results match Reina et al.'s findings that men had higher mortality risk (HR: 1.96) and that risk increased by 8% for each year of age [48]. Their later study also found that the proportion of women among COVID-19 deaths decreased across pandemic waves while the average age at death increased from 72.6 to 77.9 years, supporting our observation of strong age effects. These results are in line with previous studies. For example, Reina et al. [48] found that men had a higher hazard of death than women (HR: 1.96, 95% CI: 1.83–2.10), and that the risk of death increased by 8% for each year of age. In their later study [49], they observed that during the second and fourth waves in Cali, the percentage of women among COVID-19 deaths was significantly lower than that of men, and that mean age at death increased from 72.6 to 77.9 years between those waves. They also reported a higher proportion of deaths among those aged over 80 in the fourth wave, which supports our finding of a age gradient in mortality risk.

Furthermore, Reina et al. [48] found a hazard ratio of 2.74 (95% CI: 2.55–2.94) for the risk of death among individuals with comorbidities. Similarly, our model shows significantly increased odds of death for individuals with diabetes (OR: 1.99), cancer (OR: 1.88), and obesity (OR: 1.91) (all $p<0.05$), reaffirming that chronic conditions amplify vulnerability to COVID-19.

Using the same database as we did, both at the individual [24] and ecological levels [25], and controlling for spatial and temporal dependence [24], two studies also find socioeconomic inequalities, although in relation to the incidence of COVID-19. Specifically, both studies found that *comunas* and neighbourhoods with higher values of the unsatisfied basic needs index and low socioeconomic strata were more likely to present a higher incidence of COVID-19.

In addition, there may be inequalities along other axes, such as age and gender. On one hand, women had a much lower risk of dying compared to men, while on the other hand, although subjects aged 25 to 54 years were less likely to die than those under 25 years (reference category), those aged 65 years or older were much more likely to die. Of particular note are subjects aged 75 years or older, who had an OR of 3.19.

Our results confirm that being vaccinated against COVID-19 was a strong protective factor against death, regardless of the dose received. Receiving only one dose was associated with an 87% reduction in mortality risk, while the full vaccination schedule reduced the risk by 90%, and receiving a booster dose reduced the risk by 97%, compared to being unvaccinated. These findings align with those of Reina et al. [48], who found a hazard ratio of 0.12 (95% CI: 0.07–0.20) for full vaccination. Likewise, in their later study, Reina et al. [49] found that the drop in mortality between the second and fourth waves was consistent with the introduction and expansion of vaccination in Cali, estimating that approximately 3,763 deaths occurred in the fourth wave alone.

Our results are also comparable to those of Gálvez et al. [50], who found a 28% reduction in infection risk for individuals who received a booster compared to those with only the full schedule, and to the national-level analysis by Rojas-Botero et al. [51], who estimated an effectiveness of 86.0% (95% CI: 85.5–86.5) for the full schedule and 83.1% (95% CI: 81.5–84.5) for the booster dose. While hazard ratios reported in those studies are slightly higher than the odds ratios observed here, their confidence intervals encompass our point estimates.

Conclusions

Our study is not without limitations. First, when interpreting the results, it is important to note that our data only include registered cases of infection. Second, it is an observational study, and therefore, we were unable to prove causation. Third, some infected individuals may have been asymptomatic, while others, though symptomatic, might not have been recorded. Therefore, both groups could be missing from our database. However, we hypothesize that such cases were few. Thus, the data consisted of a total of 381,625 records. After excluding 233 records lacking a COVID-19 diagnosis date and 830 records from individuals who reported not residing in Cali, the final sample comprised 380,562 subjects. This figure is very close to that reported by the Behavioural Surveillance Group for Events of Public Health Interest of the Cali Municipal Public Health Secretariat [32] for the same study period (385,529 confirmed positive

COVID-19 cases). While it is possible that these same individuals were also be missing from these official records, the fact that our data represent 99% of the total reported cases gives us confidence that the extent of the missing data for these reasons were minimal.

Fourth, as mentioned, we used death from any cause during the study period. While it is highly likely that COVID-19 was the cause of death in more vulnerable subjects (such as the elderly and/or those with chronic conditions), we are less certain that the increased risk identified for those under 25 years of age, compared to those aged 25 to 54 years, can be entirely attributed to COVID-19 infection.

Fifth, we did not rule out the existence of unmeasured residual confounding. In fact, we did not consider certain important confounders that could have influenced vaccination rates (see, Figures S3 in Supplemental information). These confounders can be categorised as either objective and subjective, and some operate at the individual level. Among the objective confounders, we find factors such as lack of reliable Internet access [11, 12], employment conditions [11, 12], access to information [15], and use of the healthcare system. In this regard, Arsenault et al. [13] find that in Colombia individuals who had three or four visits to the healthcare system in the previous year and received at least three other preventive health services in the last year were significantly more likely to be vaccinated against COVID-19. Conversely, having unmet health-care needs in the past year, decreased the likelihood of vaccination.

Additionally, subjective confounders may have influenced vaccination rates, such as the perceived risk of contracting COVID-19 [15], fear driven by conspiratorial beliefs about vaccines [16], trust in institutions like the police, military, and church [15], trust in the government [13], vaccine hesitancy [14, 52], or, specific to Colombia, being confident in obtaining and affording quality healthcare [23].

There could also be contextual confounders that we did not address. For example, we did not consider the challenges in accessing vaccination sites. Reina et al. [47] highlight that in Cali, the vaccination campaign faced significant disruptions due to access barriers caused by blockages in various parts of the city during Colombia's 2021 civil unrest [53]. Likewise, government responses to the COVID-19 pandemic [14] may have influenced vaccination probabilities. On the other hand, and also a contextual confounder, we had no information on variants of the SARS-CoV-2 virus that could affect the effectiveness of the vaccine.

Finally, spatial designs assume that an individual's exposure to the explanatory variables is the same as that of the area in which their residence is located. However, the individual may not have always resided in the same area and/or could have been exposed in different areas. While this measurement error is unavoidable in any spatiotemporal design, it is important to note that it is also random [10].

We believe that these limitations are at least largely offset by our strengths. First, as we noted, we had a large sample size, which resulted in high statistical power (i.e., a high probability of rejecting the null hypotheses of no association when they were false). Second, we had an individual design, thus avoiding ecological fallacy. Third, we controlled for endogenous selection by modelling the probability of being observed (i.e. of reporting the residential address) and used this to obtain unbiased and consistent estimators. None of the studies cited in this paper, including those with an individual design, control for such bias. Our study does not simply describe spatiotemporal variations or estimate associations between socioeconomic characteristics and health outcomes. Instead, it explicitly integrates a model of endogenous selection bias into a Bayesian spatiotemporal framework. By jointly modelling the probability of observation (i.e. whether an individual reported a residential address) and the health outcomes of interest, we are able to correct for non-random missingness—a source of bias rarely addressed in health geography or spatial epidemiology. This combined two-part approach provides unbiased and consistent estimates that better reflect the true underlying relationships between socioeconomic status, occupation, ethnicity and COVID-19 outcomes. In doing so, it substantially improves the reliability and interpretability of our findings compared with standard approaches that assume missing data at random or treat unobserved cases as negligible. We believe this integration of endogenous selection correction within a Bayesian spatiotemporal model represents a significant methodological advance for the field, enabling future studies to produce more accurate, policy-relevant evidence. This innovation strengthens the potential of health geography and spatial epidemiology to inform equitable interventions and to address complex public health challenges beyond the COVID-19 context.

Fourth, we were the only study among those referenced to control for socioeconomic status at both the individual level (occupation and ethnicity) and, with the exception of Arango-Londoño et al. [25], at the contextual level (socioeconomic strata). However, it should be noted that the study by Arango-Londoño et al. [25] employs an ecological design, with the limitations inherent to this type of approach, (i.e. no inferences can be drawn at the individual level, existence of unmeasured confounding bias inherent in this type of design, etc.). These covariates could capture much of the uncontrolled residual confounding. Thus, objective individual factors such as lack of reliable Internet access, poorer employment conditions, and having unmet health-care needs are associated with lower individual socioeconomic status. Likewise, access to information, use of the healthcare system, and, partially, subjective individual factors, may also be related to individual-level socioeconomic variables. Moreover, some of the omitted contextual confounders could be captured by the socioeconomic strata. Finally, in our models we explicitly control for unobserved confounders, which can capture uncontrolled subjective individual factors and, to some extent, certain objective factors (such as use of the health system). Additionally, we account for spatial dependence, which helps capture omitted contextual confounders, such as lack of access to health services and vaccination points. Temporal dependence, modelled non-linearly, could also help control for the effect of SARS-CoV-2 virus variants on vaccine effectiveness or the impact of events such as blockages.

In conclusion, our study shows how socioeconomic inequalities affected both COVID-19 vaccination and mortality. By accounting for endogenous selection bias and controlling for individual characteristics and spatial-temporal patterns, we obtained more reliable estimates. We also found complex patterns in healthcare access, with typically disadvantaged groups (ethnic minorities, chronic disease patients) showing better registration rates because they interact more with the health system.

Our results show how social factors combine with individual and geographic characteristics to influence pandemic outcomes, highlighting the need for public health measures that address social and economic barriers to healthcare, not just medical interventions. Our approach using spatial analysis while correcting for selection bias provides a useful method for identifying hidden inequalities and designing better responses to health emergencies.

This study confirms the importance of tackling socioeconomic inequalities during pandemics. We found major disparities in vaccination and death rates, with poorer groups being less vaccinated and more likely to die. After correcting for selection bias, we confirmed that all vaccine doses provided strong protection against death. These findings show that public health strategies must consider social and economic factors to ensure fair vaccine distribution and overcome the barriers that worsen health inequalities. They can inform the design of more effective interventions not only for COVID-19 but also for future health crises, emphasizing the need to combine fair vaccination programs with other protective measures.

Finally, our study reinforces the importance of integrating socioeconomic factors into public health strategies to ensure equitable vaccine distribution and address the structural barriers that exacerbate health disparities. Our results provide actionable insights for developing targeted interventions, not only for controlling ongoing pandemics but also for building resilient health systems capable of addressing future global health crises effectively. These findings offer valuable guidance for nations in the prevention and management of future pandemics, emphasizing the combined importance of equitable vaccination strategies and effective non-pharmaceutical interventions.

That said, it should be noted, however, that the associations identified in this study do not establish causality. As such, the evidence presented can only partially support the advanced policy recommendations, including fairer vaccine distribution and targeted measures to address inequalities and improve population health. Without a demonstrated causal link, it is uncertain to what extent these interventions would effectively reduce disparities or enhance outcomes on the basis of this study alone.

Abbreviations

BYM	Besag-york-mollié model
COPD	Chronic obstructive pulmonary disease
DANE	Departamento administrativo nacional de estadística
GLMM	Generalized linear mixed model
ICU	Intensive care unit
INLA	Integrated nested laplace approximation
PC	Priors penalize complexity priors
UBN	Unmet basic needs index
SES	Socioeconomic status
SISBEN	Sistema de identificación de potenciales beneficiarios de programas sociales
SIVIGILA	Colombia's national public health surveillance system
sOdds	Smoothed odds

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12942-025-00448-0.

Supplementary Material 1

Acknowledgements

This study was carried out within the 'Health Inequalities and COVID-19' subprogram of CIBER of Epidemiology and Public Health (CIBERESP). The authors extend their gratitude to the Municipal Public Health Secretary of Cali, Valle del Cauca, Colombia, for generously providing the crucial COVID-19 data utilized in this paper. We are grateful to two anonymous reviewers for their constructive comments on an earlier version of this work, which have undoubtedly helped us to improve it. The usual disclaimer applies.

Author contributions

MS, MAB and FRC had the original idea for the paper and designed the study. The bibliographic search and the writing of the introduction were carried out by MS and MAB. The methods and statistical analysis were chosen and performed by MAB, MS and MM. MM created the tables and figures. All authors wrote the results and the discussion. The writing and final editing was done by all authors. All authors reviewed and approved the manuscript.

Funding

This work was partially financed by the Call for Industrial Doctorates 2021 (DI 2021), AGAUR, Government of Catalonia; and by the Consolidated research group 'Compositional and Spatial Data Analysis (COSDA)' 2021 SGR 01197, AGAUR, Government of Catalonia. FRC has been partially supported by the Universidad Nacional de Colombia, HERMES projects, Grant/Award Number: 612113. The funding sources did not participate in the design or conduct of

the study, the collection, management, analysis, or interpretation of the data, or the preparation, review, and approval of the manuscript.

Data availability
The data reported in this study cannot be deposited in a public repository. Due to ethical and legal regulations, there are restrictions on the transfer of data to third parties and data are not publicly available. However, after approval of the research proposal plan and with a signed data access agreement, and upon reasonable request to the corresponding author, a random sample of the data (properly anonymized) will be available. Study documents, including the protocol, will be available upon request from the corresponding author.The code used is in Supplemental information and will be available at [www.researchprojects.es](http:/www.researchprojects.es).Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Declarations

Ethics approval and consent to participate
The Municipal Public Health Secretariat of Cali ensures the protection of individuals' privacy through pseudo-anonymization of the data. All data management adheres to legal requirements. Secure servers were used for data storage, ensuring compliance with these regulations. As subject data extracted from the database were irreversibly pseudonymized, written informed consent was not required.

Declaration of generative AI and AI-assisted technologies
During the writing of the article the authors have not used any type of AI and AI-assisted technologies.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 14 June 2025 / Accepted: 19 December 2025
Published online: 28 December 2025

References
1. Barceló MA, Saez M. Methodological limitations in studies assessing the effects of environmental and socioeconomic variables on the spread of COVID-19: a systematic review. Environ Sci Eur. 2021;33(1):108. https://doi.org/10.1186/s12302-021-00550-7.
2. Marmot M, Allen J, Goldblatt P, Herd E, Morrison J. Build back fairer: the COVID-19 Marmot review. London: Institute of Health Equity; 2020 [Available at: https://www.instituteofhealthequity.org/resources-reports/build-back-fairer-the-covid-19-marmot-review, last accessed on December 16, 2024].
3. Mathur R, Rentsch CT, Morton CE, Hulme WJ, Schultze A, MacKenna B, et al. Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform. Lancet. 2021;397:1711–24. https://doi.org/10.1016/S0140-6736(21)00634-6.
4. Dey T, Dominici F. COVID-19, air pollution, and Racial inequity: connecting the dots. Chem Res Toxicol. 2021;34(3):669–71. https://doi.org/10.1021/acs.chemrestox.0c00432.
5. Barceló MA, Perafita X, Saez M. Spatiotemporal variability in socioeconomic inequalities in COVID-19 vaccination in Catalonia, Spain. Public Health. 2024;227:9–15. https://doi.org/10.1016/j.puhe.2023.11.024.
6. Mecenas P, Bastos RTDRM, Vallinoto ACR, Normando D. <article-title update="added">Effects of temperature and humidity on the spread of COVID-19: a systematic review. PLoS ONE. 2020;15(9):e0238339. https://doi.org/10.1371/journal.pone.0238339.
7. Smit AJ, Fitchett JM, Engelbrecht FA, Scholes RJ, Dzhivhuho G, Sweijd NA. Winter is coming: a Southern hemisphere perspective of the environmental drivers of SARS-CoV-2 and the potential seasonality of COVID-19. Int J Environ Res Public Health. 2020;17:5634. https://doi.org/10.3390/ijerph17165634.
8. Li HL, Yang BY, Wang LJ, Liao K, Sun N, Liu YC, et al. A meta-analysis result: uneven influences of season, geo-spatial scale and latitude on relationship between meteorological factors and the COVID-19 transmission. Environ Res. 2022;12:113297. https://doi.org/10.1016/j.envres.2022.113297.
9. Wu X, Nethery RC, Sabath MB, Braun D, Dominici F. Air pollution and COVID-19 mortality in the United States: strengths and limitations of an ecological regression analysis. Sci Adv. 2020;6(45):eabd4049. https://doi.org/10.1126/sciadv.abd4049.
10. Saez M, Tobías A, Barceló MA. Effects of long-term exposure to air pollutants on the spatial spread of COVID-19 in Catalonia, Spain. Environ Res. 2020;191:110177. https://doi.org/10.1016/j.envres.2020.110177.
11. Bayati M, Noroozi R, Ghanbari-Jahromi M, Jalali FS. Inequality in the distribution of COVID-19 vaccine: a systematic review. Int J Equity Health. 2022;21(1):122. https://doi.org/10.1186/s12939-022-01729-x.
12. Guay M, Maquiling A, Chen R, Lavergne V, Baysac DJ, Racine A, et al. Measuring inequalities in COVID-19 vaccination uptake and intent: results from the Canadian community health survey 2021. BMC Public Health. 2022;22(1):1708. https://doi.org/10.1186/s12889-022-14090-z.
13. Arsenault C, Lesis TP, Kapoor NR, Okiro EA, Leslie HH, Armeni P, Jarhyan P, Doubova SV, Wright KD, Aryal A, et al. Health system quality and COVID-19 vaccination: a cross-sectional analysis in 14 countries. Lancet Glob Health. 2024;12(1):e156–65. https://doi.org/10.1016/S2214-109X(23)00490-4.
14. Lansford JE, Rothenberg A, Yotanyamaneewong S, Peña Alampay L, Al-Hassan S, Bacchini D, Bornstein MH, Chang L, Deater-Deckard K, Di Giunta L, et al. Compliance with health recommendations and vaccine hesitancy during the COVID pandemic in nine countries. Prev Sci. 2024;25(2):230–44. https://doi.org/10.1007/s11121-022-01399-9.
15. Cantet N, Ibáñez N, Muñoz-Mora JC, Quintero LM. Key determinants of COVID-19 vaccination take-up in remote rural areas: evidence from Colombia. Int J Public Health. 2024;69:1606689. https://doi.org/10.3389/ijph.2024.1606689.
16. Caycho-Rodríguez T, Ventura-León J, Valencia PD, Vilca LW, Carbajal-León C, Reyes-Bossio M, et al. Network analysis of the relationships between conspiracy beliefs towards COVID-19 vaccine and symptoms of fear of COVID-19 in a sample of latín American countries. Curr Psychol. 2022;7:1–16. https://doi.org/10.1007/s12144-022-03622-w.
17. Expansión/Datos macro.com. COVID-19. Crisis del coronavirus [in Spanish] [Available at: https://datosmacro.expansion.com/otros/coronavirus?anio=2023, last accessed on May 3, 2025].
18. Johns Hopkins University. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) functions [Available at: https://coronavirus.jhu.edu/map.html, last accessed on May 3, 2025].
19. WHO COVID-19. dashboard [Available at: https://data.who.int/dashboards/covid19/cases?m49=170%26n=c, last accessed on May 3, 2025].
20. DANE. Departamento Administrativo Nacional de Estadística. Censo Nacional de Población y Vivienda 2018 [in Spanish] [Available at: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018, last accessed on May 3 2025].
21. COVID-19 en Colombia. [in Spanish] [Available at: https://www.ins.gov.co/Noticias/Paginas/coronavirus-filtro.aspx, last accessed on May 3, 2025].
22. Urrea F, Viáfara C, Arias W, Carabalí B. *Cuantos Somos Como Vamos. Diagnóstico sociodemográfico de Cali y 10 municipios de Pacífico nariñense* Universidad del Valle, Colombia, 2009 [in Spanish] [Available at: https://www.dane.gov.co/files/censo2005/etnia/sys/cuantos_somos.pdf, last accessed on May 3, 2025].
23. Sánchez-Grajales J, Bucheli-Guerrero V, Cuartas-Arroyave D. Análisis de clustering Para entender La vulnerabilidad poblacional relacionada Con El COVID-19. Investigación E Innovación En Ingenierías. 2022;10(2):7–24. https://doi.org/10.17081/invinno.10.2.5973.
24. Dong Z, Zhu S, Yie Y, Mateu J, Rodríguez-Cortés FJ. Non-stationary spatio-temporal point process modeling for high-resolution COVID-19 data. J Royal Stat Soc Ser C Appl Stat. 2023;72:368–86. https://doi.org/10.1093/jrsssc/qlad013.
25. Arango-Londoño D, Ortega-Lenis D, Moraga P, Torres M, Rodríguez-Cortés FJ. Spatial modelling and socioeconomic inequities of COVID-19 in the urban area of the city of Cali, Colombia. Spat Spatiotemporal Epidemiol. 2023;44:100561. https://doi.org/10.1016/j.sste.2022.100561.
26. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math. 1991;43(1):1–59.
27. Heckman JJ. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann Econ Soc Meas. 1976;5(4):475–92.
28. Saez M, Barcelo MA, de Coll Tuero G. A selection-bias free method to estimate the prevalence of hypertension from an administrative primary health care

database in the Girona Health Region, Spain. Comput Methods Programs Biomed. 2009;93:228–40. https://doi.org/10.1016/j.cmpb.2008.10.010.
29. SIVIGILA. Sistema Nacional de Vigilancia en Salud Pública [in Spanish] [Available at: https://portalsivigila.ins.gov.co, last accessed on May 3, 2025].
30. Secretaría de Salud Pública. Alcaldía de Santiago de Cali, Boletín Epidemiológico No. 782 [in Spanish] [Available at: https://www.cali.gov.co/salud/publicaciones/152840/boletines-epidemiologicos/, last accessed May 3, 2025].
31. Secretaría Distrital de. Salud de Cali [Available at: https://www.cali.gov.co/salud/, last accessed on May 12, 2025].
32. Gov.co. Datos abiertos. Casos positivos confirmados por COVID-19 en Santiago de Cali [in Spanish] [Available at: https://www.datos.gov.co/en/dataset/Casos-positivos-confirmados-por-COVID-19-en-Santia/j3dw-8w54, last accessed on October 16, 2023].
33. SISBEN. Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales [Available at: https://www.sisben.gov.co/Paginas/landing.html, last accessed on May 12, 2025].
34. DANE. Departamento Administrativo Nacional de Estadística. Unmet basic needs [Available at: https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/necesidades-basicas-insatisfechas-nbi, last accessed on May 3, 2025].
35. Fresneda O. *La medida de necesidades básicas insatisfechas (NBI) como instrumento de medición de la pobreza y focalización de programas* [in Spanish]. In: CEPAL - SERIE Estudios Estadísticos y Prospectivos. p. 18, 2007.
36. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). J R Stat Soc Series B Stat Methodol. 2009;71:319–92.
37. Rue H, Riebler A, Sørbye H, Illian JB, Simpson DP, Lindgren FK. Bayesian computing with INLA: a review. Annu Rev Stat Appl. 2017;4:395–421.
38. Gómez-Rubio V. Bayesian inference with INLA. London, United Kingdom: Chapman and Hall/CRC; 2020. Chap. 3.
39. Lindgren FK, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J R Stat Soc Series B Stat Methodol. 2011;73(4):423–98.
40. Plan Nacional de Vacunación contra la COVID-19. Vacunación contra COVID-19 [in Spanish] [Available at: https://www.minsalud.gov.co/salud/publica/Vacunacion/Paginas/Vacunacion-covid-19.aspx, last accessed on May 12, 2025].
41. Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. Environ Health Perspect. 2004;112(9):1016–25. https://doi.org/10.1289/ehp.6740.
42. Simpson DP, Rue H, Martins TG, Riebler A, Sørbye SH. Penalising model component complexity: a principled, practical approach to constructing priors (with discussion). Stat Sci. 2017;32(1):1–46. https://doi.org/10.1214/16-STS576.
43. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2023 [Available at: https://www.r-project.org, last accessed on December 16, 2024].
44. R INLA project [Available at. http://www.r-inla.org/home, last accessed on December 16, 2024].
45. Van Niekerk J, Krainski E, Rustand D, Rue H. A new avenue for bayesian inference with INLA. Comput Stat Data Anal. 2022;181. https://doi.org/10.1016/j.csda.2023.107692.
46. Tennekes M. Tmap: thematic maps in R. J Stat Softw. 2018;84(6):1–39. https://doi.org/10.18637/jss.v084.i06.
47. Fox J, Weisberg S. *An R Companion to Applied Regression*. Third edition. Sage, 2019. Chaper 8. Diagnostics for Generalized Linear Models.
48. Reina C, Torres M, Muñoz E, Grillo Ardila EK, Guerrero Velasco R, Bravo LE. Mortality due to COVID-19 during the vaccination plan against the SARS-CoV-2 virus in Cali, Colombia. Colomb Med (Cali). 2021;52(4):e7005047. https://doi.org/10.25100/cm.v52i3.5047.
49. Reina C, Roa P, Garcés A, Valencia A, Torres M, Concha-Eastman A. COVID19 mortality in two waves of the pandemic in Cali, Colombia, before and during vaccination roll-out. Rev Panam Salud Publica. 2023;47:e76. https://doi.org/10.26633/RPSP.2023.76.
50. Gálvez JM, Pinzón-Rondón ÁM, Chaparro-Solano HM, Tovar-Romero HV, Ramírez-Prieto J, Ortigoza-Espitia SA, et al. Effectiveness of the booster dose in protecting against COVID-19, Colombia 2022. Vaccines. 2023;11(9):1461. https://doi.org/10.3390/vaccines11091461.
51. Rojas-Botero ML, Fernández-Niño JA, Arregocés-Castillo L, Palacios-Clavijo A, Pinto-Álvarez M, Ruiz-Gómez F. Rea-world effectiveness of COVID-19 vaccines among Colombian adults: a retrospective, population-based study of the ESPERANZA cohort. PLoS Glob Public Health. 2023;3(9):e0001845. https://doi.org/10.1371/journal.pgph.0001845.
52. Humeyra-Kadafar A, Gizem-Tekeli G, Jones KA, Stephan B, Dening T. Determinants for COVID-19 vaccine hesitancy in the general population: a systematic review of reviews. Z Gesundh Wiss 2022 Sep 19:1e17. https://doi.org/10.1007/s10389-022-01753-9}
53. Restrepo-Sanin J. Colombia 2021: between crises and hope | Colombia 2021: entre La crisis y La Esperanza. Rev Cienc Polit. 2022;42(2):255–80. https://doi.org/10.4067/s0718-090x2022005000118.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.