Supplementary materials

Ethics approval and consent to participate

All the analyses of the data reported in this paper was done in accordance with the following criteria:

1. Article 157 of the General Data Protection Regulation (RGPD) recognises the benefit that accessS to records can provide to research on diseases, so that the results of these studies could be more solid, based on a larger population.

2. Article 89.1 of the RGPD confirms this principle as long as the appropriate measures are adopted, in particular respect for the principle of minimisation of personal data. These measures may also include the use of pseudonymised or anonymised data.

3. In turn, article 14.5 b) of the RGPD allows the data controller not to inform the interested parties when "said information is impossible or involves a disproportionate effort in particular for the treatment for purposes of public interest, scientific research purposes or historical or statistical purposes subject to the conditions and guarantees indicated in article 89, paragraph 1, or to the extent that the obligation mentioned in paragraph 1 of this article may make it impossible or seriously impede the achievement of the objectives of such treatment".

4. The seventeenth additional provision of Organic Law 3/2018, of December 5, on the Protection of Personal Data and guarantee of digital rights allows the use of health data for scientific research purposes as long as these data have been submitted to prior treatment of pseudonymisation or anonymisation.

In accordance with the exposed precepts, consent is not necessary for the following reasons:

1. Access to clinical records for scientific research purposes is lawful, as long as certain guarantees are adopted, including respect for the principle of data minimisation and that the information obtained is adequately pseudonymised or anonymised.

2. Obtaining consent would not be a legal requirement, since it would not only entail disproportionate efforts but could also hinder the development of the study due to the large amount of information that has been analysed.

3. The Murcia Health Service, in accordance with the provisions of the aforementioned regulations, has submitted this clinical information to an anonymisation process, excluding any identifying information of the patients, making their identification impossible.

Analysis performed to obtain the AORs

In order to obtain Adjusted Odds Ratios (AORs) in the study, we used logistic regression models. It is calculated as follows:

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n$$

Calling p to the probability of being a LC-19 patient, X_j any predictor and β_j It's associated weight. The following analyses were conducted:

- To obtain adjusted odds ratios for all comorbidities (diabetes mellitus, heart failure, chronic obstructive pulmonary disease, arterial hypertension, depression, ischaemic cardiomyopathy, stroke, renal insufficiency, cirrhosis, osteoporosis, osteoarthritis, arthritis, obesity, and asthma) and for sex, we employed a single logistic regression, where every comorbidity and the sex variable is adjusted by age, sex and the rest of comorbidities. In this model, each X_j is a comorbidity, age, and sex. This resulted in a model with 17 predictors, 15 of which were comorbidities.

- To obtain adjusted odds ratios for the symptomatology data (chest pain, muscle pain, cough, lack of appetite, sleep problems, nasal congestion, dyspnea, rhinorrhoea, low grade fever, dizziness, hyposmia, hypogeusia, headache, tired, eye symptoms, expectorate, fever, sore throat, stomach pain, vomit, nasal discharge, chills, rash and malaise), we also employed a single logistic regression, where every symptom is adjusted by age, sex and the rest of symptoms. In this model, each X_j is a symptom, age, and sex. This resulted in a model with 26 predictors, 24 of which were symptoms.

- When analysing the risk of suffering LC-19 if a patient is hospitalised, a logistic regression is also utilised. The present study has been conducted on the basis of disaggregation by sex and with both sex included, so that here we trained 3 logistic regression models. In the first two models (males - females), X_1 is the hospitalisation condition, and X_2 is the age variable. In the last model, X_1 represents the hospitalisation condition, and X_2 , X_3 representing the age and sex variables.

- A logistic regression model is constructed for each AOR reported in Table 3 (Table of vaccination differences for non LC-19 and LC-19 patients), with X_1 representing each row analysis, and X_2 , X_3 representing the age and sex variables. So here we trained 10 logistic regressions. Then, in these experiments, we have adjusted by age and sex.

- A logistic regression model is also constructed for each AOR reported in Fig. 4 and Supplementary Table 1. In each model, X_1 represents the vaccination schedule administered in each patient between those two that are being compared, while X_2 , X_3 representing the age and sex variables. So here we trained 210 logistic regressions. Then, in these experiments we have adjusted by age and sex.