AIV_HZ_2021_05_COVIDHZ_AOS

Study protocol

Risk of Herpes Zoster in individuals diagnosed with SARS-CoV2 infection in the Valencia region of Spain: a retrospective cohort population-based study

Principal Investigator Alejandro Orrico-Sánchez (PharmD, PhD).

Co-Investigators Cintia Muñoz-Quiles (Biologist, PhD), Mónica López-Lacort (Mathematician, MSc) and Javier Díez-Domingo (Pediatrician, MD, PhD).

Due date Sept 2021

Description of Work Version 5.0

Contact Information Vaccine Research Department. FISABIO-Public Health

Avda Cataluña, 21
46020 Valencia, Spain
Telephone: +34 961 925 950
Alejandro.orrico@fisabio.es

Document history

Version	Date	Description	
V1.0	17 March 2021	Fisabio drafted the V1.0 of the protocol and distributed it to GSK's team	
V2.0	1 June 2021	GSK's team sent their comments and they were discussed in a meeting on June 3rd.	
V3.0	7 July 2021	Fisabio sent new version	
V 4.0	13 July 2021	Fisabio sent new version	
V 5.0	08 Sept 2021	Last version approved by GSK and Fisabio	

Table of contents

List of abbreviations

ABUCASIS	Ambulatory Medical Record
AED	Accident & Emergency Department
ATC	Anatomical Therapeutic Chemical code
BIMCV	Biobanco de Imagen Médica de la Comunidad Valenciana (Medical Imaging Databank)
DRG	Diagnosis Related Groups
FISABIO	Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana
GP	General Practitioner
ICU	Intensive care unit
ICD	International Classification of Diseases
IR	Incidence rate
MBDS	Minimum Basic Data Set at Hospital Discharge
RedMIVA	Microbiological Surveillance Network of Valencia
RR	Relative Risk

Responsible parties

Main Author(s) of the Protocol

Name	Institution	Role	Contributions
Alejandro Orrico-Sánchez	FISABIO	Principal investigator,	First draft and updates until last version
Cintia Muñoz-Quiles	FISABIO	Co-investigator	First draft and updates until last version
Monica Lopez-Lacort	FISABIO	Co-investigator and lead statistician	First draft and updates until last version
Javier Díez-Domingo	FISABIO	Medical Doctor	Supervision and medical consultant
Amit Bahvsar	GSK	Epidemiologist	Supervision and advice
Raunak Parikh	GSK	Medical expert	Supervision and advice
José Parra	GSK	Statistician	Supervision and advice

1. Rationale and background

COVID-19 is an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) first identified in Wuhan, China, in December 2019 (1). As of March 10th, 2021, COVID-19 pandemic has resulted in more than 117 million confirmed cases and above 2.6 million deaths worldwide (2). Therefore, SARS-CoV-2 outbreak has led to the largest and more severe pandemic since the influenza pandemic in 1918-19 (3). Spain ranks as the seventh most affected country with more than 3 million cases and 71,727 deaths (2). To date, the Spanish region Community of Valencia has registered 381,919 positive cases and 6,911 deaths (4).

SARS-CoV-2 infection may be asymptomatic or cause a wide spectrum of symptoms that may appear after a variable period of incubation (average time from exposure to onset is 5 days; 97.5% of people who develop symptoms do so within 11.5 days) (5). Among them, the most common symptoms derive from the primary infection of the respiratory tract and consist of dry cough, shortness of breath and fever (5). Additional symptoms include fatigue and changes to taste and smell (5, 6). However, SARS-CoV-2 infection has also been associated with alterations in the CNS and peripheral nervous system (7), liver (8), cardiovascular system (9), kidneys (10) and skin (11, 12).

Cutaneous manifestations of COVID-19 were first reported in early 2020 in China with only 2 out of 1,099 positive cases (13). However, a study carried out later in Italy found that 18 out of 88 COVID-19 positive patients developed dermatological alterations, of which 8 developed cutaneous involvement at the onset and 10 after the hospitalization. Cutaneous manifestations included erythematous rash (14 patients), widespread urticarial (3 patients) and chickenpox-like vesicles (1 patient) (14). Since then, several studies have discussed reports of patients infected with COVID-19 associated with vesicular manifestations of Herpes Zoster (HZ) (15, 16, 17, 18). A recent Turkish publication collecting data from fifteen tertiary hospitals from 13 provinces described an increase in the number of HZ diagnoses when comparing two months periods pre and post onset of the pandemic (19). These results are consistent with a latter study carried out in Brazil in which the number of HZ cases were quantified during the months of March to August from 2017 to 2019 and compared with the similar period in 2020 (20). Although the HZ cases were not assessed within COVID-19 positive patients, an extra 10.7 cases per million inhabitants during the pandemic was reported, which might suggest a correlation between both diseases.

HZ is a severe disease resulting from the reactivation of Varicella Zoster Virus (VZV), which remains latent in sensory nerve ganglia after primary infection (Varicella) (21). HZ can result in dermatomal chronic pain which is its most common complication (post-herpetic neuralgia, PHN) (22). Many patients with PHN go on to develop severe physical, occupational, and societal disabilities because of the enduring pain. Both, HZ and PHN result in reduced quality of life as well as individual and societal health care costs (23).

VZV reactivation seems to be a result of a waning of VZV-specific cell-mediated immunity as occurs with ageing or in subjects with immunosuppressive disorders (24). SARS-CoV-2 infection is associated with lymphopenia with reduced numbers of CD4 + and CD8 + T cells, B cells and natural killer (NK) cells (25). Because cell-mediated immunity is key in the protection against HZ (26), lymphopenia could predispose COVID-19 patients to HZ development. However, more scientific evidence needs to be provided to corroborate this hypothesis.

Based on the potential association between SARS-CoV-2 and HZ, we propose here a study to estimate the risk of HZ in subjects with SARS-CoV-2 infection using Real World Data (RWD). This population-based retrospective dynamic cohort study (meaning that members can leave or be added over time) will be based on eHR databases/registries from the Valencia Region (Valencia

Health System Integrated Database (VID) (27). VID allows linking socio-demographic, inpatients, outpatients, specialists, medication, and microbiology databases (among others) at individual level. The study population will consist of all population covered by the Public Health System (over 98%), representing about 5 million persons. As of March 8th, 2021, more than 381,919 confirmed cases of SARS-CoV-2 have been registered in the Valencian Community (4). Together with the VID makes the region of Valencia the ideal candidate to test the proposed hypothesis.

2. Hypothesis

We hypothesize that individuals 50 years and older previously infected with SARS-Cov-2 have higher risk of experiencing Herpes Zoster.

3. Objectives

The overall objective is to estimate the risk of HZ in subjects 18 years and older infected with SARS-CoV-2 using the Valencia health system Integrated Databases (VID).

3.1 Primary objectives

Compare the risk of HZ among individuals 50 years and older with and without laboratory confirmed SARS-CoV2 infection.

3.2 Secondary objectives

Compare the risk of HZ among individuals 18 years and older with and without laboratory confirmed SARS-CoV2 infection.

Compare the risk of HZ in overall population older than 18 years old in the pandemic period against pre-pandemic period.

Describe the likelihood of developing HZ after laboratory-confirmation of SARS-CoV2 according to severity of disease.

3.3 Exploratory objectives

Describe the changes with time in likelihood of developing HZ after laboratory-confirmation of SARS-CoV2.

4. Methods

4.1 Study design

A population based, retrospective dynamic cohort study using real-world data from the Valencia healthcare Integrated Databases (VID).

Individuals within the cohort will be followed up over time to determine the incidence of HZ. As a dynamic cohort, individuals will be recruited to or leave the cohort at different times depending on the inclusion/exclusion criteria.

4.2 Study population and period

The cohort will include all subjects older than 18 years living in Valencia Region, registered in the Regional Health System (RHS) (over 98%) from January 1st 2018 to, at least, December 31st 2020 (or date of data extraction). The inclusion and exclusion criteria will be the following:

Inclusion criteria: Subjects aged 18 years and above covered by the RHS and residing in the Valencia Region during study period.

Exclusion criteria:

- Subjects with less than 12 months of registration to the RHS and subjects with incorrect codification.
- Subjects vaccinated against HZ before the diagnosis of SARS-CoV-2
- Subjects vaccinated against SARS-CoV-2 before the diagnosis of SARS-CoV-2

Start of follow-up: The start of the follow-up is defined as the latest of start of the study (1st January 2018), first date of registration in the database or date of the 18th birthday.

End of follow-up: The end of the follow-up is defined as the earliest of the following cases:

- End of study period
- Loss of follow-up (exit of RHS)
- *Exitus*
- Vaccination with HZ vaccine
- Vaccination with COVID-19 vaccine
- Development of HZ

4.3 Case definition of HZ

Incident herpes zoster cases will be considered as the appearance of a HZ-related ICD-code (ICD-9 053.x; ICD-10 B02.x) diagnostic code for herpes zoster, excluding those with specific ICD-codes for PHN (ICD-9 053.12, 053.13 and 053.19, ICD-10 B02.2). Cases will be identified from outpatient, SIA (see section 4.8), and inpatient, MBDS, in any diagnostic position. Incident HZ will be defined as an episode of herpes zoster without any evidence of herpes zoster or PHN for at least 6 months previously.

4.4 Sars-Cov-2 exposure

Exposure status of everyone might change within the follow up period (Figure 3): Sars-Cov-2 unexposed in the pre-pandemic period and Sars-Cov-2 exposed in the pandemic period. Sars-Cov-2 exposure levels will be treated as a time-varying variable and the following states will be considered:

1) ***Sars-Cov-2 unexposed***: subject's follow up period from the start of follow-up to the declaration of the pandemic in Spain (March 14th 2020).
2) ***Sars-Cov-2 exposed***: subject's follow up period from the start of the pandemic. We will consider the following 2 comparison groups:
    - **Sars-Cov-2 free**: time from the start of pandemic (or the follow-up) to the date of the first Sars-Cov-2 confirmation (by PCR or Ag test). If no Sars-Cov-2 positive test registered, the time period will be until the end of follow-up.
    - **Sars-Cov-2 laboratory-confirmed**: time from the date of the SARS-CoV-2 positive result in RedMiva to the end of follow up.

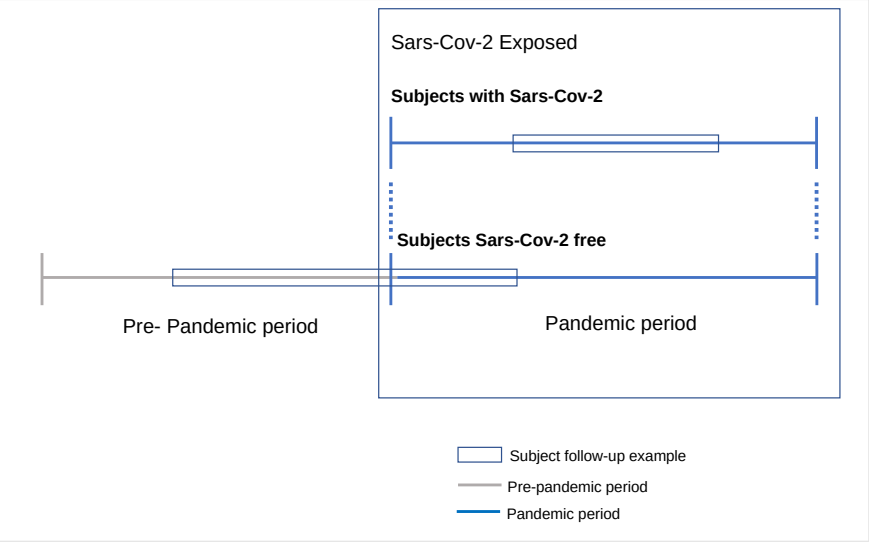


Figure 3. Example of exposure, follow-up and comparisons in the study.

* Subjects with recurrent episodes of Sars-Cov-2 will be treated separately. If the sample size is not large enough for the analysis, these individuals will be excluded for the analysis.

4.5 Other health outcomes

4.5.1 Pandemic and pre-pandemic periods

- Pre-pandemic period will be considered from January 1st 2018 to March 14th 2020
- Pandemic period will be considered from March 14th 2020 to the end of study period

4.5.2 Severity

Both hospitalizations and stays in ICU will be considered for severity analyses

4.6 Covariates

Covariate information will be collected on sociodemographic, spatial information, vaccination against Sars-Cov-2 and comorbidities. The source of the eHR databases and the variables needed to create the covariates is given in Annex 1.

4.6.1 Sociodemographic variables

Age

Age will be treated as time-dependent variable. Population follow-up will be stratified according to the following four age groups (18-49/50-64/70-79/≥80).

Sex

Female/Male

Rural residence

Rural residence (yes/no) will be classified based on the law for sustainable development of the rural environment from the Regional Government. Rural areas are classified according to population density (<100 inhabitants/km2), urban nucleus proximity, population trend, percentage of employment in primary, secondary and tertiary sectors and territorial structure.

Nationality

Spanish/Other

Social exclusion risk

At risk (yes/no). The classification is based on multiple aspects including unemployment, foreigner in irregular situation or without resources.

4.6.2 Spatial variables

The spatial variables are summarized in *Figure 2:*

Figure 2. Geographical stratification of the Valencia Region health care system: Province, Health Department, Health Care District and Municipality.

4.6.3 Comorbidities

Comorbidities that have been identified as risk indicators for HZ will be considered for the analysis. These conditions will be identified by systematic search of ICD-codes in the hospital or ambulatory databases, MBDS or SIA. Subjects will be considered to

have a given condition from the first appearance of the comorbidity-related diagnosis date until the end of the follow-up period.

Disease classification	ICD-10
Diabetes	E10-E14
COPD	J44
Ischemic heart disease	I20-I25
Other heart disease	I27-I52
Kidney disease	N18-N19
Immunocompromised	
HIV	B20-B24
Cancer	C00-D49
TOTS	Z94
IBD	K50-K51
RA	M05-M14
SLE	M32

COPD: chronic obstructive pulmonary disease, HIV: human immunodeficiency virus, TOTS: Transplanted organ and tissue status, IBD: inflammatory bowel disease, RA: rheumatoid arthritis, SLE: systemic lupus erythematous.

Presence of any comorbidity.

This is a binary variable indicating whether the patient has at least one comorbidity defined above.

Number of underlying conditions

This is a categorical variable indicating the number of comorbidities (0, 1, 2, 3+).

4.6.4 Temporal variables

Temporal variables such as calendar week, month and year will be contemplated to detect trends.

4.7 Data sources

4.7.1 The Valencia healthcare Integrated Databases (VID) - Real World Data

The Valencia Region accounts with the Valencia Health System Integrated Database (VID) [27], one of the best clinical and administrative databases to perform observational studies using RWD. VID is a set of multiple, public, population-wide electronic databases including sociodemographic and administrative data, clinical (hospital and GP setting), vaccination registry and microbiology databases (including both positive and negative results). Importantly, VID allows linking all these databases at individual level using the personal ID number of each citizen of The Region.

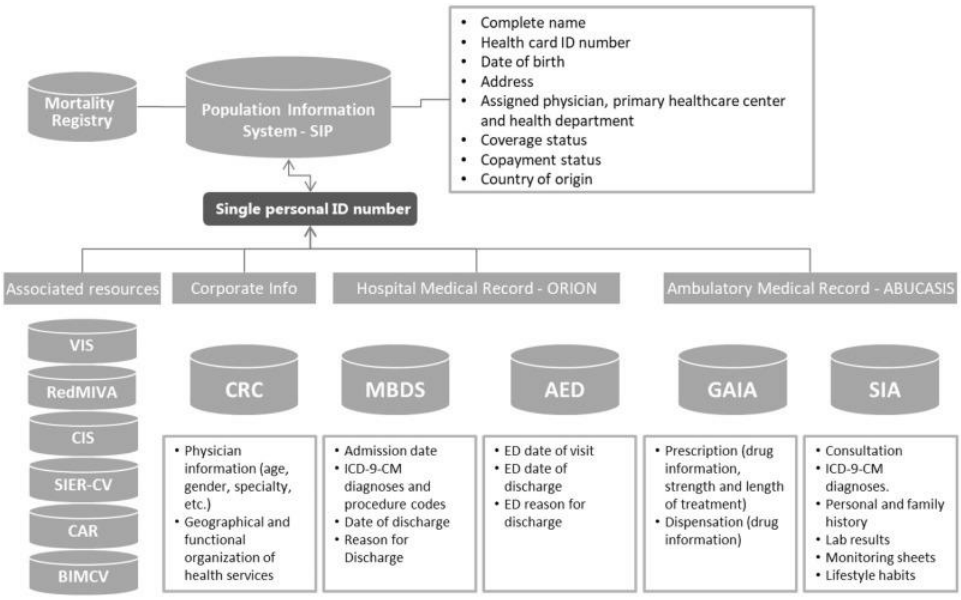


Figure1: *Scheme of the Valencia Health System Integrated Database (VID); VIS, Vaccine Information System; RedMIVA, Microbiological Surveillance System; CIS, Cancer Information System; SIER-CV, Rare Diseases Information System; CAR, Congenital Abnormalities Registry; BIMCV, Medical Image Bank; CRC, Catalogue of Corporate Resources; MBDS, Minimum Basic hospital Data Set; AED, Accident & Emergency Department record; GAIA, Pharmaceutical Module; SIA, Ambulatory Information System*

The following registries from VID will be used for the study:

Population-based administrative database (SIP)

The regional population-based administrative database, SIP (Personal Information System), collects and updates demographic data, health services assignment, and access to public health services for both residents of the Valencia Community and non-residents with access to public health services. It includes APSIG characteristic which is an identification code defined for each person at any time including: inhabitant's registration status, nationality (Spanish or not), sex, year of birth, health department assigned, health care insurance, residence status, migrations, work activity, geopolitical group, and social exclusion. Since 2005, SIP can be linked with the hospital discharge database. All other healthcare databases are able to capture the demographic data from SIP.

Primary Care Database

Abucasis SIA-GAIA (Ambulatory Information System - Care provision management) is a primary care database used across the entire Valencia healthcare system that contains medical information for each patient attended in the Primary Care (PC) setting (General Practitioners, GPs, and specialists). It was set up in 2006 and the percentage of the population included increased from 73.1% in 2007 to 88.8% in 2008 and to 95.7% in 2009. This database contains primary care diagnoses (physician coded using the International Classification of Diseases Revision, Clinical Modification (ICD-CM)) and all drug prescriptions (using Anatomical Therapeutic Chemical (ATC) Classification System). In addition, the physician or paediatrician and nurse responsible recorded text

about each episode and about the patient is included. The database is also used at Specialty Centres and is considered reliable since 2007.

Hospital Discharge Database

The Spanish hospital discharge database, MBDS (Minimum Basic Data Set), collects diagnosis and procedures as an assessment of medical activity. The coding system used is ICD-9-CM until 2016 and ICD-10-CM afterwards. The main discharge diagnosis is coded in first position, and diagnosis relevance decreases as the position number increases. Using MBDS is compulsory for all public hospitals, and over 95% of all discharges are included. According to the Spanish Ministry of Health, data are considered reliable since 2002.

Microbiological Surveillance Network

The Microbiological Surveillance Network (RedMIVA) contains all the results (positive or negative) given by all the microbiology laboratories of the Valencian Public System network. All data is transferred from the laboratories to the RedMIVA database on a daily basis. Data are considered reliable since 2008.

4.8 Data collection

The following variables (Annex 1) will be requested to the different databases for the period from July 2017 until at least December 31st 2020 (depending on the date of data extraction) for subject's ≥ 18 years. Data privacy will be protected by using dissociated data.

4.9 Sample size considerations

4.9.1 Cohort estimations

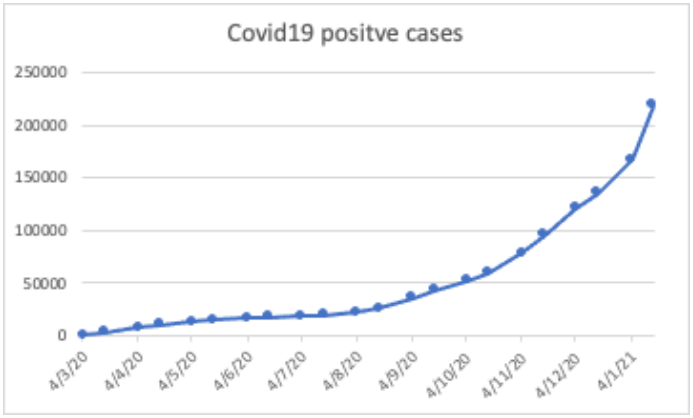
Currently, approximately 4,900,000 inhabitants of Valencia Region are covered by the PHS. Since the study is restricted to subjects aged 18 years and older and approximately 20% of the Spanish population is younger than 20 years (data from Statistics National Institute, INE), we might expect approx. 4 million subjects to fulfil the inclusion criteria.

4.9.2 HZ incidence rates estimations

Data from a previous study from our team (28), using the same health databases from Valencia Region showed an incidence rate of HZ of 5.02 (4.99–5.04), 5.3 (5.3-5.4), 7.3 (7.2-7.4), 8.3 (8.2-8.4) and 8.6 (8.5-8.8) per 1000 persons-year in subjects ≥ 18, 50-59, 60-69, 70-79 and ≥ 80 years old, respectively.

4.9.3 Expected SARS-CoV-2 cases

As of January 2021, The Valencia Region registered a total of 219.144 confirmed cases of Covid-19 in all ages. The figure below shows the evolution of the cumulative positive cases in The Region over time.



119.935 of them were confirmed in people older than 50 years of age. The table below shows the number of positives for SARS-CoV-2 by age group.

Age groups	Women	Men	Total
20-29	21273	19607	40880
30-39	22481	19960	42441
40-49	27613	25163	52776
50-59	24720	22533	47253
60-69	15886	15828	31714
70-79	10835	10182	21017
80-89	8963	6047	15010
90+	3578	1363	4941
TOTAL	63982	55953	119935

4.9.4 Number of tests performed in The Region over time

As of December 31st 2020, a total of 2.112.504 SARS-CoV-2 tests were performed. The following table shows the monthly number of positive and negative tests (both PCR and Ag) performed.

		Negative results				Positive results			
Month	Total tests	Total negatives	% of total	AG	PCR	Total	% of total	AG	PCR
feb-20	514	468	91,05058366		468	46	9,829059829		46
mar-20	39403	26394	66,98474735		26394	13009	49,28771691		13009
abr-20	80998	69783	86,15397911		69783	11215	16,07124944		11215
may-20	114973	112264	97,64379463		112264	2709	2,413062068		2709
jun-20	101371	100395	99,03719999		100395	976	0,972159968		976
jul-20	133709	128809	96,33532522		128809	4900	3,804082013		4900
ago-20	243498	216819	89,04344184		216819	26679	12,30473344		26679
sept-20	269026	240845	89,52480429	912	239933	28181	11,74536225	109	28072
oct-20	305193	255984	83,87610463	11170	244814	49209	20,10056614	4832	44377
nov-20	394144	323865	82,16920719	44718	279147	70279	25,17634078	14560	55719
dic-20	429675	348567	81,12340723	67192	281375	81108	28,82558863	20070	61038
Total 2020	**2112504**	**1824193**	**86,35216785**	**123992**	**1700201**	**288311**	**16,95746562**	**39571**	**248740**

4.9.5 Power Calculation for HZ IR comparison among SARS-Cov-2 lab-confirmed and SARS-Cov-2 free.

Considering an HZ IR of 7 cases per 1000 persons-year in SARS-Cov-2 free, and the following expected populations: 110 000 SARS-Cov-2 lab-confirmed and 2 000 000 SARS-Cov-2 free, we would be able to find statistically significant (significance level = 0.05) differences in RR for HZ (SARS-Cov-2 free vs. Lab-confirmed) of 10% (RR 1.1) with a power of ~ 80% (power = 0.760864).

Power calculation was estimated based on the Two-sample Poisson Ratio Tests (Different Sizes)from the PASSED package of R (version 4.0.2).

Statistical analysis

4.9.6 Primary objective

To develop the primary objective, we will compare the incidence of HZ among SARS-Cov-2 free and Sars-Cov-2 laboratory-confirmed subjects 50 years and older during the pandemic period. For this purpose, models using individual patient data or grouped will be implemented.

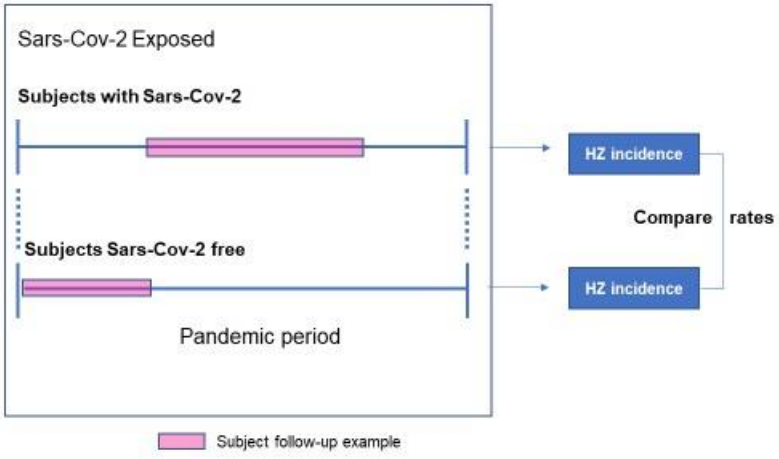


Figure 4: Cohort diagram for HZ rates comparison among subjects with and without previous episode of COVID19

The risk of HZ in subjects with Sars-cov-2 respect to subjects without Sars-Cov-2 will be estimated by a multivariate Poisson or negative binomial model according to applicability assumptions. Individual data will be grouped by SARS-Cov-2 free and Sars-Cov-2 laboratory-confirmed exposure, age, sex and comorbidities. Individual and grouped data can both be analyzed with the Poisson (or Negative binomial) distribution. Given the usual large size of the cohort, estimation when using grouped data is considerably quicker than when using individual level data and the results for both approaches are remarkably similar.

The number of HZ cases by aggregation unit will be compared among SARS-Cov-2 free and lab-confirmed subjects. Age, sex, and comorbidities will be considered for the adjustment, and the person-time-at-risk (in years) as an offset term. RR and their 95% confidence or credible intervals will be reported.

Besides, individual-level risk of HZ between SARS-Cov-2 free vs lab-confirmed will be estimated by Cox regression model for right censored data. The abovementioned variables will be also considered for the adjustment.

Both approaches will be performed and compared for the robustness of the results.

4.9.7 Demographic and baseline characteristics

Demographic characteristics of the study population will be summarized using descriptive tables including frequencies and proportions. Prevalence of comorbidities studied, and Sars-Cov-2 incidence will be assessed. Demographic data will be presented by type of comorbidity, pre-pandemic and pandemic period (including the different Sars-Cov-2 exposure levels regarding their severity).

4.9.8 HZ incidence rates (Descriptive analysis)

Incidence rates (number of patients with an incident case of HZ per 1000 persons-years) will be obtained for subjects with and without Sars-Cov-2 in the pandemic period by sex, age-group and comorbidity. HZ incident rates for the pre-pandemic period will be also described by sex, age-group and comorbidity.

The person-time-at-risk (in days) ends at the date of the first event of HZ or the end of follow-up, whichever comes first. Follow-up time will be split according to the time-varying covariates (calendar year, month, week, age and comorbidities). Persons-years will be calculated as the sum of total person-time-at-risk divided by 365.25. Rates 95% confidence intervals will be calculated by the Exact Poisson method.

4.9.9 Secondary objectives

1) A statistical analysis like the previous one will be developed to compare the risk of HZ among individuals 18 years and older between SARS-CoV2-free and lab-confirmed.
2) The risk of HZ in subjects with Sars-Cov-2 exposed (pandemic) respect to subjects unexposed (pre-pandemic) will be estimated by a multivariate Poisson or negative binomial model according to applicability assumptions. Age, sex and comorbidities will be considered for the adjustment.

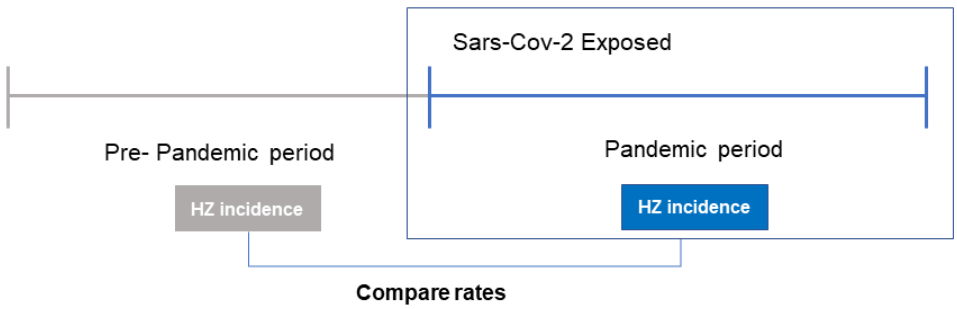


Figure 2. Visual example of HZ incidence rates comparisons among pre-pandemic and pandemic periods.

3) Describe the likelihood of developing HZ after laboratory-confirmation of SARS-CoV2 according to severity of disease.

   We will calculate and compare the predicted probability of developing incident HZ in the 1, 3, 6 months following Sars-Cov-2 (1, 3, 6 – months cumulative incidence) using logistic

regression models. We will compare the overall probability, regarding Sars-Cov-2 severity, and the probabilities obtained from models adjusting for the abovementioned covariates (See section 3.5).

4.9.10 Exploratory objectives

1) Describe the changes with time in likelihood of developing HZ after laboratory-confirmation of SARS-CoV2.

A descriptive analysis (including Means, standard deviations, quartiles and ranges) of the time to the HZ incident case after Sars-Cov-2 diagnosis will be summarized overall and stratified by age, sex, frailty category, and comorbidity and Sars-Cov-2 severity.

Analyses of the anonymised data will be carried out using R Statistical Software version 4.1.1 (Foundation for Statistical Computing, Vienna, Austria).

4.10 Limitations

Potential limitations of the VID have been described elsewhere (27). Data quality may be a strength in some databases, but also a weakness in other repositories or for certain data, such as incompleteness of early data from Accident and Emergency Department (AED) records. Also, we do not have information about people that are not in contact with the public healthcare service or who are attended to in the private sector.

Clinically-diagnosed SARS-CoV-2 or asymptomatic cases without laboratory confirmation will be considered as SARS-CoV-2 free in the study. Nevertheless, Public Health surveillance for Covid-19 team guarantees that that percentage must be insignificant.

4.11 Missing data

The Valencia Region databases have a very low percentage (0.02%) of unregistered data; subjects with missing information will be excluded for the study.

Patients with incomplete follow-up will account for the analysis as of the date of loss of follow up.

4.12 Data management

The process for data obtaining in The Valencia Region will have the next steps:1) Ethics Revision Board (ERB) approval; 2) General Director of Public Health approval for data transfer; 3) Telematic data request; 4) PROSIGA committee approval of data requested; 5) Data extraction; 6) Data checking and cleaning; 7) Potential new extraction (if errors); 8) Data analysis.

All data files will be received in plain text (.txt) format with the variables separated by pipes. Identification data will be dissociated with the same seed. The information from the different sources will be merged by the subject identification number (SIP number). All files will be passed a check list to guarantee the quality of the data and detect possible errors in the extraction.

Population will be determined by the SIP file; only data from SIA, CMBD and RedMIVA that intersect with the population will be considered. Data in long format (two or more records by subject) will be transformed in wide format having a subject in each record.

Txt files will be transformed into analytical datasets suitable for further sharing and analysis. The analytical dataset will contain anonymized data only. The data transformation files will be programmed in R.

5. Regulatory and Ethical considerations

The study will be conducted in accordance with all applicable regulatory requirements, including all applicable subject privacy requirements, the Declaration of Helsinki, and the International Ethical Guidelines for Epidemiological Investigations.

The study will be submitted for approval to the institutional review board.

For transparency purposes, the protocol will be posted into the European Network of Centres for pharmacoepidemiology and pharmacovigilance (ENCePP) website.

5.1 Informed Consent

No informed consent is necessary as this a retrospective study using deidentified/anonymized data for secondary purposes.

6. Management and reporting of adverse events

NA

7. Plans for disseminating and communicating results

Results of this study will be published as scientific papers in peer-reviewed journals. Such manuscripts will be prepared independently by the investigators and in accordance with the current guidelines of Strengthening the Reporting of Observational studies in Epidemiology (STROBE) and properly reported to improve reproducibility and facilitate validity assessment of healthcare database studies.

8. References

1. WHO statement regarding cluster of pneumonia cases in Wuhan, China. Jan 9, 2020. https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china

2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20:533–534.

3. Taubenberger JK, Morens DM. 1918 influenza: the mother of all pandemics. Emerg Infect Dis. 2006;12:15–22.

4. COVID-19 C. Valenciana: Monitoratge de la situació. http://coronavirus.san.gva.es/es/estadisticas

5. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA. 2020 Aug 25;324(8):782-793.

6. Spinato G, Fabbris C, Polesel J, et al. Alterations in smell or taste in mildly symptomatic outpatients with SARS-CoV-2 infection. JAMA. 2020;323(20):2089-2209.

7. Ellul MA, Benjamin L, Singh B, Lant S, Michael BD, Easton A, Kneen R, Defres S, Sejvar J, Solomon T. Neurological associations of COVID-19. Lancet Neurol. 2020 Sep;19(9):767-783.

8. Mao R, Qiu Y, He JS, et al. Manifestations and prognosis of gastrointestinal and liver involvement in patients with COVID-19: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2020;5(7):667-678.

9. Long B, Brady WJ, Koyfman A, Gottlieb M. Cardiovascular complications in COVID-19. Am J Emerg Med. Published online April 18, 2020.

10. Chen YT, Shao SC, Hsu CK, Wu IW, Hung MJ, Chen YC. Incidence of acute kidney injury in COVID-19 infection: a systematic review and meta-analysis. Crit Care. 2020;24(1):346.

11. Marzano AV, Cassano N, Genovese G, Moltrasio C, Vena GA. Cutaneous manifestations in patients with COVID-19: a preliminary review of an emerging issue. Br J Dermatol. 2020 Sep;183(3):431-442.

12. Jia JL, Kamceva M, Rao SA, Linos E. Cutaneous manifestations of COVID-19: A preliminary review. J Am Acad Dermatol. 2020 Aug;83(2):687-690. doi: 10.1016/j.jaad.2020.05.059.

13. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Zhong NS; China Medical Treatment Expert Group for Covid-19. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020 Apr 30;382(18):1708-1720.

14. Recalcati S. Cutaneous manifestations in COVID-19: a first perspective. J Eur Acad Dermatol Venereol. 2020 May;34(5):e212-e213.
15. Fernandez-Nieto D, Ortega-Quijano D, Suarez-Valle A, Burgos-Blasco P, Jimenez-Cauhe J, Fernandez-Guarino M. Comment on: "To consider varicella-like exanthem associated with COVID-19, virus varicella zoster and virus herpes simplex must be ruled out. Characterization of herpetic lesions in hospitalized COVID-19 patients". J Am Acad Dermatol. 2020 Sep;83(3):e257-e259.

16. Llamas-Velasco M, Rodríguez-Jiménez P, Chicharro P, De Argila D, Muñoz-Hernández P, Daudén E. Reply to "Varicella-like exanthem as a specific COVID-19-associated skin manifestation: Multicenter case series of 22 patients": To consider varicella-like exanthem associated with COVID-19, virus varicella zoster and virus herpes simplex must be ruled out. J Am Acad Dermatol. 2020 Sep;83(3):e253-e254.

17. Marzano AV, Genovese G, Fabbrocini G, Pigatto P, Monfrecola G, Piraccini BM, Veraldi S, Rubegni P, Cusini M, Caputo V, Rongioletti F, Berti E, Calzavara-Pinton P. Varicella-like exanthem as a specific COVID-19-associated skin manifestation: Multicenter case series of 22 patients. J Am Acad Dermatol. 2020 Jul;83(1):280-285.

18. Ortega-Quijano D, Jimenez-Cauhe J, Burgos-Blasco P, Jimenez-Gomez N, Fernandez-Nieto D. Reply to "Varicella-like exanthem as a specific COVID-19-associated skin manifestation: multicenter case series of 22 patients": Discussing specificity. J Am Acad Dermatol. 2020 Jul;83(1):e87.

19. Selda Pelin Kartal et al. Multicenter study evaluating the impact of COVID-19 outbreak on dermatology outpatients in Turkey Dermatol Ther. 2020 Nov 2;e14485.

20. Maia CMF, Marques NP, de Lucena EHG, de Rezende LF, Martelli DRB, Martelli-Júnior H. Increased number of Herpes Zoster cases in Brazil related to the COVID-19 pandemic. Int J Infect Dis. 2021 Feb 11;104:732-733.

21. Drolet M, Brisson M, Schmader KE, et al. The impact of herpes zoster and postherpetic neuralgia on health-related quality of life: a prospective study. CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne. 2010;182(16):1731-1736.

22. Johnson RW, Rice ASC. Postherpetic Neuralgia. N Engl J Med. 2014;371(16):1526-1533.

23. Gater A, Uhart M, McCool R, Preaud E. The humanistic, economic and societal burden of Herpes Zoster in Europe: a critical review. BMC public health. 2015;15:1514-1514.

24. Arvin AM. Humoral and cellular immunity to varicella-zoster virus: An overview. J Infect Dis. 2008;197:S58-S60.

25. Wang B, Guo S, Yao Y, Li Y, Zhang G. Dermatologists may need to pay more attention to herpes zoster during the pandemic of COVID-19. Infect Dis (Lond). 2020 Nov-Dec;52(12):917-918.

26. Cao X. COVID-19: immunopathology and its implications for therapy. Nat Rev Immunol. 2020 May;20(5):269-270.

27. Garcıa-Sempere A, et al. Data Resource Profile: The Valencia Health System Integrated Database (VID) Int J Epidemiol. 2020 Jun 1;49(3):740-741e.

28. Munoz-Quiles C, Lopez-Lacort M, Diez-Domingo J. Risk and impact of herpes zoster among COPD patients: a population-based study, 2009-2014. BMC Infect Dis. 2018;18(1):203.

1. BUDGET

Project phase	Budget	Project: Impact of COVID-19 pandemic on HZ risk: A population-based study using Real World Data	AIV
	Task	**Sub-task**	**Total cost**
Phase I	**Pre-work**		22250
		Previous meetings and emails	
	Proposal		
		Proposal's coordination	
		Brain storming's meeting	
		Literature review	
		Proposal's writing	
		Statistica analysis	
		Budget	
	Contract		
		Head of Fisabio's approval management	
		Signature's management	
		Fisabio's bureaucracy management	
	Study coordinator		
		Responsible for meeting the project scope, time and cost requirements	
		Monitor and periodically review the status of the project	
		Gantt chart	
		Meeting plan	
	Protocol		
		Design	
		Literature review	
		Writing	
		Writing support	
		Statistical analysis	
		Follow-up meetings with GSK	
	Statistical Analysis Plan		
		Thinking, preparation and writing	
		Follow-up meetings with GSK	
	Ethical requirements		
		AEMPs	
		ERB bureaucracy	
		Clarifications	
		Amendments	
		Taxes	
Phase II	**Electronic databases (5 registries)**		43170
		Request creation	
		Data providers contacts	
		Bureaucracy	
		Script	
		Data extraction	
		Data quality checking	
		Corrections	
		New extraction after corrections	
		Follow up	
	Data analysis		
		Data cleaning, tabulation and fusion	
		Mathematical model creation	
		Data analysis, adjustments	
		Results interpretation and discussion	
		Preliminary report	
		Final Report	
		Web App (Shiny) for the preentation of the results	
	Folow-up		
		Follow-up meetings with GSK	
Phase III	**Comunications**		3080
		International/national meeting	
		Scientific paper writing	
		Submission 1/2 journalts	

SUM	**68500**
Budget for 2021	**34850**
Budget for 2022	**33650**
Indirect costs (15%)	**10275**
Total cost	**78775**

2. TIMELINES

Tasks/Time (months)	ene-21	feb-21	mar-21	abr-21	may-21	jun-21	jul-21	ago-21	sept-21	oct-21	nov-21	dic-21	ene-22	feb-22	mar-22	abr-22	may-22	jun-22	jul-22	ago-22	sept-22
Proposal	***																				
Study design and protocol and SAP		**																			
IBR approval			*																		
Contract signature				*																	
Data delivery approval				*																	
Data extraction and collection (5 registries)							*			*			*								
Data cleaning and tabulation																					
Data management and analysis																					
Results interpretation and discusion														*	*						
Primary Objective																*					
Secondary Objectives																	*				
Provisional Report/Review and discusion																			**		
Final Report																				**	
Manuscript Preparation and Submission																					**

* Follow Up Meetings

3. ANNEX 1

3.1. LIST OF VARIABLES TO BE EXTRACTED FROM DATABASES

Data privacy will be protected by using anonymised data (see 2.2. section). The following variables will be requested to the different databases for the period from 1 January 2007 to 31 January 2015 for subjects ≥18 years old:

Data to be extracted from SIP:

- Identification block including SIP number, sex, date of birth and other geographical of birth, place and date of registration (excluding identifying data as surnames, ID card number, phone numbers, etc.)
- Assignation block (including municipality, healthcare district, etc)
- Regular location block that includes complete address, health map information as health department and census information among others (excluding exact data that allows the exact residence location of the subjects)
- Cessation block including cessation cause and description, cessation date and date of death (when applicable)
- APSI code
- Barthel Index

Data to be extracted from SIA

- SIP number

Morbidity database

- Date of diagnosis activation (diagnosis of interest)
- Date of diagnosis deactivation (diagnosis of interest)
- Diagnosis description (diagnosis of interest)
- ICD-codes (diagnosis of interest)

Outpatient visits database

- Date of the visit
- Diagnosis description
- ICD-code

Specialist visits database

- Date of the visit
- Diagnosis description
- ICD-code
- Professional attendance (GP/Specialist)
- Specialist service

For any diagnosed code related to HZ (see section 4.3) and comorbidities (see section 4.6.3)

Data to be extracted from MBDS:

- Anonymized Personal Identification Number (SIP)
- Birth date
- Sex
- Municipality of residence
- Postal code
- Health department
- Health care district
- Date of hospital admission
- Date of hospital discharge
- Diagnoses at discharge (main and secondary diagnoses)
- Procedures during the hospitalization
- Discharge destination (destination on discharge)
- ICU (any position)
- POA

For any diagnosed code related to HZ (see section 4.3) and comorbidities (see section 4.6.3) and SARS-Cov-2 (J12.89, J20.8, J40, J22, J98.8, J80, R65.20, R65.21, R05, R06.02, R50.9, R43.0, Z20.828, Z03.818, Z11.59 O98.5, B97.29 or U07.1 and B34.2)

Data to be extracted from RedMIVA

- SIP number
- Date of laboratory test (dd/mm/yyyy)
- Date of test result
- Sample type
- Laboratory test (PCR, antigen, etc)
- Result

For any SARS-Cov-2 test

Source	Dataset	Variable	Label	Type	Coding
SIP	Identification block	SIP	SIP number	character	Random key

SIP	Identification block	sex	Sex	character	Sex_fmt
SIP	Identification block	birth_date	Date of birth	date	yyyy-mm-dd
SIP	Identification block	birth_place	Geographical birth	date	yyyy-mm-dd
SIP	Identification block	nationality	nationality	character	Natio_fmt
SIP	Identification block	reg_status	Registration status	character	Reg_fmt
SIP	Regulation block	reg_date	Date of registration	date	yyyy-mm-dd
SIP	Regulation block	PC	Postal code	integer	Pc_fmt
SIP	Regulation block	muni	Municipality	Character	Muni_fmt
SIP	Regulation block	province	Province	Character	Prov_fmt
SIP	Assignation block	hc_district	Health care district	Character	Hc_dist_fmt
	Assignation block	h_department	Health department	Character	H_dep_fmt
SIP		APSI	APSI code	Character	
SIP	Cessation block	ces_date	Cessation date	Date	yyyy-mm-dd
SIP	Cessation block	death_date	Date of death	Date	yyyy-mm-dd
MBDS		SIP	SIP number	Character	
MBDS		hosp	Hospital	Character	
MBDS		h_department	Health department	Character	
MBDS		admission_date	Date of hospital admission	Date	yyyy-mm-dd
MBDS		dIscharge_date	Date of hospital discharge	Date	yyyy-mm-dd
MBDS		d_i (i= 1,…,30)	Diagnoses at discharge (any position)	character	ICD-9 ICD-10
MBDS		p_i (i= 1, …, 30)	Procedures during the hospitalization	character	ICD-9 ICD-10
MBDS		Des_discharge	Discharge destination (destination on discharge)	Date	yyyy-mm-dd
MBDS		Death_cause	Cause of death	character	Two diagnoses ICD coded
MBDS		ICU	ICU (any position)	integer	Days
SIA-GAIA		SIP	SIP number	character	Random key

SIA	Morbidity database	Act_date	Date of diagnosis activation (diagnosis of interest)	date	yyyy-mm-dd
SIA	Morbidity database	Dea_date	Date of diagnosis deactivation (diagnosis of interest)	date	yyyy-mm-dd
SIA	Morbidity database	Dx_desc	Diagnosis description (diagnosis of interest)	character	
SIA	Morbidity database	Dx_code	ICD-code (diagnosis of interest)	character	ICD-9 ICD-10
SIA	*Outpatient visits database*	Visit_date	Date of the visit	date	yyyy-mm-dd
SIA	*Outpatient visits database*	Dx_desc	Diagnosis description	character	
SIA	*Outpatient visits database*	Dx_code	ICD-code	character	ICD-9 ICD-10
SIA	*Specialist visits database*	Visit_date	Date of the visit	date	yyyy-mm-dd
SIA	*Specialist visits database*	Dx_desc	Diagnosis description	character	
SIA	*Specialist visits database*	Dx_code	ICD-code	character	ICD-9 ICD-10
SIA	*Specialist visits database*	Specialist	Professional attendance (GP/Specialist)	character	Profess_fmt
SIA	*Specialist visits database*	Service	Specialties Service	character	Service_fmt
RedMIVA		SIP	SIP number	character	Random key
RedMIVA		Lab_date	Date of laboratory test	date	yyyy-mm-dd
RedMIVA		micro	Microorganism/antigen	character	
RedMIVA		Res_date	Date of test result	date	yyyy-mm-dd
RedMIVA		Sample_type	Sample type	character	
RedMIVA		Lab_test	Laboratory test	character	

RedMIVA		result	Result	character	Res_fmt
RedMIVA		Antimicro_res	Antimicrobial resistance	character	
SIV		SIP	SIP number	character	Random key
SIV		Vaccine_type	Vaccine type (PCV7, PCV10, PCV13 or PPV23)	character	
SIV		Dose	Dose	integer	
SIV		Admin_date	Administration date (anti-pneumococcal vaccination)	date	yyyy-mm-dd

Format	**Categories**
Sex_fmt	Female
Sex_fmt	Male
Natio_fmt	Spanish, Other
reg_fmt	unregistered
reg_fmt	unregistered <1 month
reg_fmt	unregistered ≥ 1 month
reg_fmt	registered
Muni_fmt	542 municipalities
PC_fmt	Ex: 46112
Prov_fmt	Alicante
Prov_fmt	Castellón
Prov_fmt	Valencia
Hc_dist_fmt	241 health care districts
H_dep_fmt	24 health departments
APSI_fmt	Ex: 1SH200411261610110BA3

00	NO ASIGNADO
01	DP1 VINAROS
02	DP2 CASTELLON
03	DP3 LA PLANA
04	DP4 SAGUNTO
05	DP5 VALENCIA CLINICO
06	DP6 VALENCIA ARNAU DE VILANOVA
07	DP7 VALENCIA LA FE
08	DP8 REQUENA
09	DP9 CE JUAN LLORENS TORRENT
10	DP10 VALENCIA DR PESET
11	DP11 LA RIBERA
12	DP12 GANDIA
13	DP13 DENIA
14	DP14 DENIA
15	DP15 ALCOI
16	DP16 VILA JOIOSA
17	DP17 ALICANTE SAN JUAN
18	DP18 ELDA
19	DP19 ALICANTE
20	DP20 ELX
21	DP21 ORIHUELA
22	DP22 TORREVIEJA
23	DP23 MANISES
24	DP24 ELX CREVILLENT

	The APSIG code is made up of a set of dimensions, each representing a specific aspect that allows the segmentation of the entire integrated population in SIP. APSIG codes vary over time, 3 codes at the beginning, middle and end of the study will be requested for each subject. It is a 32-digit code that is structured according to the following table:

Desde	Hasta	Long.	Descripción	Valores
1	1	1	Situación de empadronamiento	1,2,3
2	2	1	Indicador de nacionalidad española	S,N,D
3	3	1	Sexo	H,M
4	11	8	Fecha de nacimiento	yyyymmdd
12	13	2	Departamento de asignación	1,2,3,... 24
14	15	2	D1. (Financiación cobertura)	10,20,30,40,51,52,60
16	16	1	D2. (Situación de residencia)	0,1,2,3,4
17	18	2	D3. (Migraciones)	10, 21,22,23,31,32,33,4
19	19	1	D4. (Relación con la actividad laboral)	A,B,C,D,E,O
20	21	2	D5. Grupos y subgrupos de Aseguramiento	A1,A2,A3,B1,B2,B3,B4,C
22	25	4	D6. Conjuntos geopolíticos	Pos 22 – 0,1,2,3,4,5,6,7, Pos 23 a 25 – según tabl
26	26	1	D7. Perfil de Vulnerabilidad APSI	0,1,2,3,4,5
27	32	6	D8. Unidad de residencia	Pos 27–28 1 a 15 Pos 29–30 0,1,2,3,4,5 Pos 31–32 00
33	34	2	D9. RAF-Renta(régimen de aportación Farmacia y Niveles de Renta)	Pos 33-34
35	36	2	Zona de asignación	ZZ la zona asociada a Departamento de asigna 00, 01, ...18 dependi departamento corresp

Profess_fmt	Pneumologist, …,
Service_fmt	Endocrinology , …,etc.
Res_fmt	Positive
Res_fmt	Negative
Res_fmt	Other

Signed

Dr. Alejandro Orrico Sánchez (principal investigator)

ALEJANDRO |ORRICO| SANCHEZ

Firmado digitalmente por ALEJANDRO|ORRICO|SANCHEZ
Fecha: 2021.09.13 11:17:42 +02'00'