



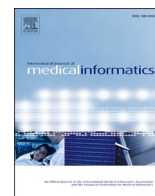
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

## International Journal of Medical Informatics

journal homepage: [www.elsevier.com/locate/ijmedinf](http://www.elsevier.com/locate/ijmedinf)

## Improving morbidity information in Portugal: Evidence from data linkage of COVID-19 cases surveillance and mortality systems

Rodrigo Feteira-Santos<sup>a,b,c,d,\*</sup>, Catarina Camarinha<sup>a,b,e</sup>, Miguel de Araújo Nobre<sup>a,b,f</sup>,  
Cecília Elias<sup>a,g</sup>, Leonor Bacelar-Nicolau<sup>a,b,c,d</sup>, Andraia Silva Costa<sup>a,d,h,i</sup>,  
Cristina Furtado<sup>a,b,d,e,j</sup>, Paulo Jorge Nogueira<sup>a,b,c,d</sup>

<sup>a</sup> EPI Task-Force FMUL, Faculdade de Medicina, Universidade de Lisboa. Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

<sup>b</sup> Instituto de Medicina Preventiva e Saúde Pública, Faculdade de Medicina, Universidade de Lisboa. Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

<sup>c</sup> Área Disciplinar Autónoma de Bioestatística (Laboratório de Biomatemática), Faculdade de Medicina, Universidade de Lisboa. Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

<sup>d</sup> Instituto de Saúde Ambiental, Faculdade de Medicina, Universidade de Lisboa. Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

<sup>e</sup> Unidade de Epidemiologia, Instituto de Medicina Preventiva e Saúde Pública, Faculdade de Medicina, Universidade de Lisboa. Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

<sup>f</sup> Clínica Universitária de Estomatologia, Faculdade de Medicina, Universidade de Lisboa. Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

<sup>g</sup> Unidade de Saúde Pública Francisco George, ACES Lisboa Norte, Administração Regional de Saúde de Lisboa e Vale do Tejo. Largo Professor Arnaldo Sampaio, 1500-559 Lisboa, Portugal

<sup>h</sup> CIDNUR - Centro de Investigação, Inovação e Desenvolvimento em Enfermagem de Lisboa, Escola Superior de Enfermagem de Lisboa. Avenida Professor Egas Moniz, 1600-190 Lisboa, Portugal

<sup>i</sup> CRC-W—Católica Research Centre for Psychological, Family and Social Wellbeing, Universidade Católica Portuguesa. Palma de Cima, 1649-023 Lisboa, Portugal

<sup>j</sup> Instituto Nacional de Saúde Doutor Ricardo Jorge. Av. Padre Cruz, 1600-560 Lisboa, Portugal

## ARTICLE INFO

## Keywords:

Health information system  
Systems interoperability  
Morbidity  
COVID-19

## ABSTRACT

**Background:** COVID-19 rapidly spread around the world, putting health systems under unprecedented pressure and continuous adaptations. Well-established health information systems (HIS) are crucial in providing data to allow evidence-based policymaking and public health interventions in the pandemic response. This study aimed to compare morbidity information between two databases for COVID-19 management in Portugal and identify potential complementarities.

**Methods:** This is an observational study using records from both COVID-19 cases surveillance (National Epidemiological Surveillance System; SINAVE) and related deaths (National e-Death Certificates Information System; SICO) systems, which were matched on sex, age, municipality of residence and date of death. After the linkage, morbidity reported in SINAVE and identified in SICO, through the application of Charlson and Elixhauser comorbidity indexes algorithms, were compared to evaluate agreement level.

**Results:** Overall, 2285 matched cases were analyzed, including 53.9% males with a median age of 84 years. According to the method of data reporting assessment, the presence of any morbidity ranged between 26.3% and 62.5%. The reporting of ten morbidities could be compared between the information reported in SINAVE and SICO databases. The proportion of simultaneous reporting in both databases ranged between 5.7% for diabetes and 0.0% for human immunodeficiency virus infection or coagulopathy. Minimal or no agreement was found when assessing the similarity of the morbidity reporting in both databases, with neoplasms showing the highest level of agreement (0.352, 95% IC: 0.277–0.428;  $p < 0.001$ ).

**Conclusion:** Different information about reported morbidity could be found in two HIS used to monitor COVID-19 cases and related deaths, as data are independently collected. These results show that the interoperability of SICO and SINAVE databases would potentially improve available HIS and improve available information to decision-making and address COVID-19 pandemic management.

\* Corresponding author at: Área Disciplinar Autónoma de Bioestatística, Faculdade de Medicina, Universidade de Lisboa. Avenida Professor Egas Moniz, 1649-028, Lisboa, Portugal.

E-mail address: [rodrigasantos@medicina.ulisboa.pt](mailto:rodrigasantos@medicina.ulisboa.pt) (R. Feteira-Santos).

<https://doi.org/10.1016/j.ijmedinf.2022.104763>

Received 7 October 2021; Received in revised form 11 March 2022; Accepted 10 April 2022

Available online 15 April 2022

1386-5056/© 2022 Elsevier B.V. All rights reserved.

## 1. Introduction

COVID-19 emerged in late 2019 as one of the greatest threats to human health over the last centuries [1]. Due to the severity and rapid dissemination of this disease across the globe, the World Health Organization (WHO) declared COVID-19 a pandemic state on March 11<sup>th</sup>, 2020 [2]. Since then, several public health measures have been adopted to “flatten the curve” of incidence and minimise the impacts of this infection [3].

Given the urgency of the phenomenon and the need for timely available data, health information systems (HIS) play an essential role in providing information to health authorities [4]. The discussion around the ability of health care systems to take advantage of exchange information and interoperability between HIS is not new [5]. Among the benefits elicited so far is the potential to generate financial savings within total health expenditures [6] and public health through improved patient care [7,8].

Although health information exchange within HIS should be promoted beyond the COVID-19 crisis [4], the relevance of a robust and interoperable HIS for public health decision-making [9] has been highlighted during this pandemic [4,10]. Throughout this health crisis, measures such as the surveillance of infections, contact tracing and the characterisation of deaths have been essential to monitor the overall epidemic impact and generate evidence to inform policymaking in adjusting or implementing other public health interventions.

The most straightforward approaches include taking advantage of available routine HIS, COVID-19-related reporting platforms, or existing surveillance systems [11]. However, the pandemic has highlighted some of the insufficiencies of health surveillance systems for COVID-19 monitoring. Consequently, adjustments were necessary to provide a timely and articulated response from the entire health system. In Portugal, for instance, human resources were strengthened with the allocation of more health professionals and contact tracing trained military personnel [12] and a specific HIS was created for doing contact tracing (e.g. platform Trace COVID-19) [13].

Nevertheless, when different platforms are created to collect similar health information independently, separated data are generated, which can constitute an obstacle to its efficient use and partly compromise a rapid, strong, and accurate response [14]. The benefits of a sound HIS have previously been described, such as their importance for the governance and management of a health system in achieving better outcomes [15,16].

Several strategies aiming to improve HIS have been studied, considering the potential of emerging technologies and automation used during data collection and analysis. One of them is the interoperability of different HIS, allowing them to communicate in a coordinated manner, and enhance available health information, particularly during a pandemic [17].

This study aimed to compare and identify the potential complementarity of the available information between two databases regarding the morbidities recorded within the surveillance of COVID-19 cases and related deaths in Portugal.

## 2. Material and methods

We conducted a registry-based observational study to analyse COVID-19-related morbidity in 2020 from two databases used in Portugal to collect information on COVID-19 infections and associated deaths. This study is reported following the STROBE statement [18] and RECORD extension [19].

### 2.1. Study population and data sources

All records, either in the National e-Death Certificates Information System [*Sistema de Informação do Certificados de Óbito*] (SICO) database or in the National Epidemiological Surveillance System [*Sistema*

*Nacional de Vigilância Epidemiológica*] (SINAVE) database, of all deceased people from COVID-19 during 2020 in Portugal were considered. These two databases include data from the most relevant HIS in Portugal. First, a subset of SICO was used to obtain all COVID-19-related deaths certificated in Portugal between March 16<sup>th</sup> and December 31<sup>st</sup>, 2020. SICO is a national and web-based information system on mortality managed by the Portuguese Directorate-General of Health (DGS), which provided the SICO dataset used in this study. The SICO database includes information on causes of death and related comorbidities, inputted by medical doctors who certify deaths, then coded by a team of trained coders allocated explicitly to this task. Thus, the presence of a “COVID-19” code - International Statistical Classification of Diseases and Related Health Problems, 10<sup>th</sup> Revision (ICD-10) [20]: U07.1 or U07.2 - as the underlying cause of death was considered to identify COVID-19 related fatalities. Second, the SINAVE database was used to obtain information about the clinical characteristics of the COVID-19 cases. The SINAVE system allows the electronic notification and epidemiological surveys of diseases which require mandatory reporting in Portugal, managed by Shared Services of the Ministry of Health (SPMS), which is the national health information technology authority, and DGS, which provided the dataset for this study. Consequently, the SINAVE database has been used in Portugal to register the surveillance data of COVID-19 cases. In addition, the SINAVE records of deceased people during 2020, with a COVID-19-related cause of death, were considered for this study. As a specialized team codes the open-ended field information in SICO, morbidity data in this database is considered to be more accurate and reliable than in the SINAVE.

### 2.2. Data linkage and records inclusion

A new database was created by matching records related to the same death in SICO or SINAVE databases. As each record was de-identified in the databases, four variables available in both databases were considered to link the same cases. Thus, the linkage of cases from each database was conducted based on age, sex, municipality of residency, and date of death. If the conjugation of those four variables occurred in more than one case, either in SICO or SINAVE, those cases were excluded. Then, information from each database was linked and included in a new database including the records from SICO and SINAVE matched on those four variables mentioned above.

In the SICO database, morbidity information is first reported through an open-ended field and subsequently coded using ICD-10. Although some inconsistencies between the open-ended field and ICD-10 coding fields were observed throughout the dataset, in some weeks, the SICO dataset presented several cases with information about morbidity in the open-ended field but no ICD-10 code in coded morbidity fields. Therefore, the authors considered using a criterion to minimize bias derived from cases dying within weeks where a substantial proportion of cases with an incomplete coding process was observed. Consequently, the completeness of the coding process for each week was assessed by identifying cases whose open-ended field information was available but was not coded according to ICD-10 to coded morbidity fields. A team consensus established that an incomplete coding process occurred during a given week if more than 5% of cases dying that week had open-ended field morbidity information but any ICD-10 code in respective fields. Thus, the analysis only included deceased people within weeks without an incomplete coding process.

### 2.3. Variables

Age (recoded into four age groups) and sex were used to describe sample characteristics. Following the International Organization for Standardization week date system (ISO 8601), the date of death was recoded into week number of the year (from 1 to 53).

Morbidities coded in the SICO database according to the ICD-10 [20] were used to calculate Elixhauser [21] (ECI) and Charlson Comorbidity

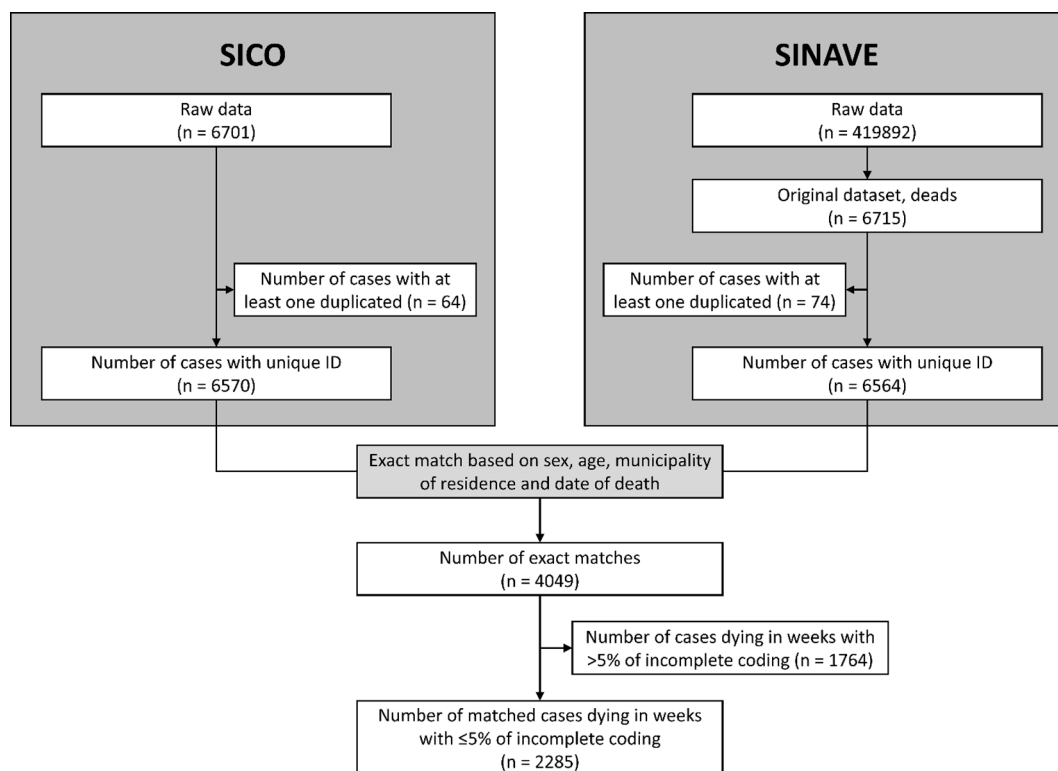


Fig. 1. Inclusion flow chart of records from SINAVE and SICO databases considered in the analysis (n = 2285) SICO, National e-Death Certificates Information System; SINAVE, National Epidemiological Surveillance System.

indexes [22] (CCI), using their ICD-10-adapted algorithms. Therefore, the occurrence of morbidities was described according to four methods: 1) the comorbidities directly reported in the SINAVE database by those who filled in the notification, which states whether the individual had or did not have comorbidities; 2) the recodification of the occurrence or not of reported morbidities in SINAVE database; 3) the identification of comorbidities in the SICO database according to the algorithm of CCI, or

4) the identification of comorbidities in the SICO database according to the algorithm of ECI.

The SINAVE system allows for recording the occurrence of 13 comorbidities, namely: neoplasia, diabetes mellitus, human immunodeficiency virus (HIV), neurologic or neuromuscular disease, asthma, chronic pulmonary disease, hepatic disease, chronic hematologic disease, chronic renal disease, chronic neurologic deficiency, acute renal

Table 1  
Number of individuals with reported comorbidities according to each method of obtaining this result.

		Total n (%)	Male n (%)	Female n (%)	p- value	0 – 69 years n (%)	70 – 79 years n (%)	80 – 89 years n (%)	greater than 90 years n (%)	p- value
<b>SINAVE Original morbidity reporting</b>	No	79 (3.5%)	46 (3.7%)	33 (3.1%)	0.812	15 (5.5%)	16 (3.4%)	33 (3.4%)	15 (2.6%)	0.460
	Yes	779 (34.1%)	439 (35.7%)	340 (32.3%)		101 (37.3%)	175 (37.3%)	323 (33.3%)	179 (31.1%)	
	Subtotal	858 (37.5%)	485 (39.4%)	373 (35.4%)		116 (42.8%)	191 (40.7%)	356 (36.7%)	194 (33.7%)	
	With no data	1427 (62.5%)	746 (60.6%)	681 (64.6%)		155 (57.2%)	278 (59.3%)	613 (63.3%)	381 (66.3%)	
<b>SINAVE Recoded morbidity reporting</b>	No	1163 (50.9%)	609 (49.5%)	554 (52.6%)	0.078	123 (45.4%)	239 (51.0%)	500 (51.6%)	301 (52.3%)	0.060
	Yes	600 (26.3%)	341 (27.7%)	259 (24.6%)		83 (30.6%)	140 (29.9%)	241 (24.9%)	136 (23.7%)	
	Subtotal	1763 (77.2%)	950 (77.2%)	813 (77.1%)		206 (76.0%)	379 (80.8%)	741 (76.5%)	437 (76.0%)	
	With no data	522 (22.8%)	281 (22.8%)	241 (22.9%)		65 (24.0%)	90 (19.2%)	228 (23.5%)	138 (24.0%)	
<b>SICO Charlson Comorbidities index</b>	No	914 (40.0%)	498 (40.5%)	416 (39.5%)	0.638	113 (41.7%)	181 (38.6%)	380 (39.2%)	239 (41.6%)	0.671
	Yes	1371 (60.0%)	733 (59.5%)	638 (60.5%)		158 (58.3%)	288 (61.4%)	589 (60.8%)	336 (58.4%)	
<b>SICO Elixhauser Comorbidities index</b>	No	857 (37.5%)	461 (37.4%)	396 (37.6%)	0.965	98 (36.2%)	163 (34.8%)	357 (36.8%)	238 (41.4%)	0.132
	Yes	1428 (62.5%)	770 (62.6%)	658 (62.4%)		173 (63.8%)	306 (65.2%)	612 (63.2%)	337 (58.6%)	

Footnotes: Fisher and Chi-Square tests were used to evaluate the association between each method of reporting of any morbidity and sex or age group. One record had missing data for age.

**Table 2**  
Number of reported comorbidities according to each method of quantification.

Number of identified morbidities	SINAVE n (%)	Charlson n (%)	Elixhauser n (%)
0	1163 (50.9%)	914 (40.0%)	857 (37.5%)
1	312 (13.7%)	713 (31.2%)	479 (21.0%)
2	185 (8.1%)	465 (20.4%)	464 (20.3%)
3	77 (3.4%)	155 (6.8%)	286 (12.5%)
4	19 (0.8%)	29 (1.3%)	134 (5.9%)
5	6 (0.3%)	8 (0.4%)	47 (2.1%)
6	1 (0.04%)	1 (0.04%)	12 (0.5%)
7	–	–	6 (0.3%)

failure, congestive heart failure and coagulopathy. A variable that indicated whether a case had or did not have comorbidities, filled in by the person who notified that given case, was also available. Two new variables were recoded through the information reported on the 13 reported comorbidities in SINAVE. One variable informs if any of the 13 comorbidities in SINAVE were reported, while the other enumerates how many were reported. Chronic renal disease and chronic renal failure, both morbidities recorded in SINAVE, were recoded into the same variable.

Thus, comorbidities reporting in both databases for the same case were compared. When possible, comorbidities identified using ECI and CCI in SICO were matched with those reported in SINAVE. The selection of what ECI and CCI morbidities from SICO would be compared with each comorbidity reported in SINAVE, within those mentioned above, was made after a consensus arising from the research team, which included two physicians, a nurse and other health professionals. Matches between comorbidities in both databases are presented in Appendix A - Table A.1.

2.4. Statistical analysis

Descriptive statistical methods were performed. Accordingly, absolute and relative frequencies were calculated for categorical variables. Fisher exact and Chi-Square tests were performed to evaluate the associations between categorical variables. Cohen’s kappa test was used to determine the level of agreement between databases. For all tests performed, the level of statistical significance was set at 0.05. Statistical analysis was conducted using IBM SPSS Statistics for Windows (version 26.0, 2019, Armonk, NY: IBM Corp).

3. Results

From a total of 6701 death records in SICO and 419892 SARS-CoV-2

**Table 3**  
Differences in the number of comorbidities reported in the SINAVE database and SICO database.

Number of comorbidities	SINAVE							Kappa	
	0	1	2	3	4	5	6		
SICO Charlson index	0	492 (70.5%)	118 (16.9%)	50 (7.2%)	25 (3.6%)	8 (1.1%)	5 (0.7%)	0 (0.0%)	0.051 (95% CI:0.024 – 0.078) p < 0.001
	1	359 (63.8%)	112 (19.9%)	64 (11.4%)	21 (3.7%)	6 (1.1%)	1 (0.2%)	0 (0.0%)	
	2	225 (64.5%)	63 (18.1%)	45 (12.9%)	16 (4.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
	3	70 (56.9%)	16 (13.0%)	21 (17.1%)	12 (9.8%)	3 (2.4%)	0 (0.0%)	1 (0.8%)	
	4	13 (54.2%)	3 (12.5%)	5 (20.8%)	2 (8.3%)	1 (4.2%)	0 (0.0%)	0 (0.0%)	
	5	3 (60.0%)	0 (0.0%)	0 (0.0%)	1 (20.0%)	1 (20.0%)	0 (0.0%)	0 (0.0%)	
	6	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
SICO Elixhauser index	0	438 (68.7%)	111 (17.4%)	49 (7.7%)	27 (4.2%)	8 (1.3%)	5 (0.8%)	0 (0.0%)	0.032 (95% CI: 0.007 – 0.057) p = 0.009
	1	239 (63.1%)	78 (20.6%)	45 (11.9%)	12 (3.2%)	4 (1.1%)	1 (0.3%)	0 (0.0%)	
	2	244 (66.3%)	63 (17.1%)	46 (12.5%)	13 (3.5%)	2 (0.5%)	0 (0.0%)	0 (0.0%)	
	3	143 (64.1%)	40 (17.9%)	25 (11.2%)	13 (5.8%)	2 (0.9%)	0 (0.0%)	0 (0.0%)	
	4	62 (62.0%)	15 (15.0%)	13 (13.0%)	8 (8.0%)	2 (2.0%)	0 (0.0%)	0 (0.0%)	
	5	28 (70.0%)	3 (7.5%)	4 (10.0%)	3 (7.5%)	1 (2.5%)	0 (0.0%)	1 (2.5%)	
	6	6 (66.7%)	1 (11.1%)	2 (22.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	
7	3 (50.0%)	1 (16.7%)	1 (16.7%)	1 (16.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		

Footnotes: CI, confidence interval. Since morbidities data were missing in SINAVE for 522 cases, these results refer to 1763 cases.

infections identified in SINAVE during 2020, of which 6715 died, the linkage approach for both databases found 4049 entries as exact matches on four variables (age, sex, municipality, and date of death). Therefore, it represents a proportion of 60.4% and 60.3% (considering only deceased patients) of matched records among the original samples, respectively. Then, after excluding cases dying within weeks where incomplete coding was verified, 2285 matches were analysed: 53.9% males and a median age of 84 years (IQR: 76–90). Fig. 1 shows the flow chart showing enrolment of included records from SINAVE and SICO databases.

According to each available method of reporting, the distribution of comorbidities occurrence (or absence) across gender and age groups is presented in Table 1. The presence of any reported morbidity for a given case varied between 26.3%, when the presentation of any morbidity was recoded through the information reported in specific morbidities variables of SINAVE, and 62.5%, when the presence of morbidities was identified using the ECI in SICO database. The reporting of any morbidity per case was similar and independent for both sex and age groups. However, the percentage of morbidities reporting was tendentially lower for people dying at age 90 years or more, as this group had the lower proportion of morbidity reported according to three out of four methods presented in Table 1. Also noteworthy is the proportion of missing values in reporting any morbidity in the SINAVE database (62.5%). However, the percentage of missing values in the recoded variable of the presentation of morbidities was lower (22.8%).

The number of reported morbidities in SINAVE or identified morbidities according to the CCI in the SICO system ranged between zero and six. Additionally, the ECI algorithm identified six cases with seven comorbidities, as presented in Table 2.

The agreement between reported morbidities number in SINAVE and each of the CCI and ECI is described in Table 3. The number of cases without any reported or identified morbidity in SINAVE and simultaneously in CCI or ECI was 492 and 438, respectively, corresponding to 27.9% and 24.8% of the total sample. Within the cases for which it was possible to match information on morbidities between both databases, a total of 662 (37.5%) and 577 (32.7%) cases had the same number of comorbidities reported in SINAVE and identified in SICO with either CCI or ECI, respectively. When a given number of comorbidities higher than zero were identified in SICO and reported in SINAVE, the higher percentage of concordance was observed with one morbidity reported in both databases, in 6.4% of the sample using CCI and in 4.4% using ECI. Those proportions of cases with one morbidity reported in SINAVE corresponded to about one-fifth of cases, also with one morbidity, identified using CCI (19.9%) or ECI (20.6%). These results revealed a lack of agreement among the number of comorbidities reported for each case between databases.

**Table 4**  
Occurrence of comorbidities reporting in each database.

Morbidity		SINAVE n (%)	SICO n (%)
Neoplasia	No	1636 (92.8%)	2076 (90.9%)
	Yes	127 (7.2%)	209 (9.1%)
Diabetes	No	1519 (86.2%)	1864 (81.6%)
	Yes	244 (13.8%)	421 (18.4%)
HIV	No	1754 (99.5%)	2281 (99.8%)
	Yes	9 (0.5%)	4 (0.2%)
Neurologic or neuromuscular disease	No	1650 (93.6%)	1765 (77.2%)
	Yes	113 (6.4%)	520 (22.8%)
Chronic pulmonary disease	No	1631 (92.5%)	2117 (92.6%)
	Yes	132 (7.5%)	168 (7.4%)
Hepatic disease	No	1739 (98.6%)	2255 (98.7%)
	Yes	24 (1.4%)	30 (1.3%)
Chronic renal disease or acute renal failure	No	1603 (90.9%)	1992 (87.2%)
	Yes	160 (9.1%)	293 (12.8%)
Congestive heart failure	No	1744 (98.9%)	1945 (85.1%)
	Yes	19 (1.1%)	340 (14.9%)
Coagulopathy	No	1763 (100.0%)	2268 (99.3%)
	Yes	0 (0.0%)	17 (0.7%)

Footnotes: information on SINAVE morbidity was missing for 522 cases. Percentages are presented only respecting to valid cases.

After a consensus between the research team, ten morbidities could be compared between SINAVE database reporting and the classification according to the CCI or ECI in the SICO database (Table A.1 in the Appendix A). The cases percentages with any of compared morbidity in each database are detailed in Table 4. The morbidity with the higher percentage of reporting in the SINAVE database was diabetes, whose information was reported for 244 cases (13.8%). In comparison, neurologic or neuromuscular disease registered the highest percentage of diseases identified with CCI or ECI in the SICO database (n = 520 cases; 22.8%). In addition, neurologic or neuromuscular diseases registered the highest difference (16.4%) between the proportion of reporting in each database, followed by congestive heart failure reporting (13.8%). Otherwise, the other differences between the morbidities reported in SINAVE and SICO databases were lower than 5%.

Table 5 shows the number of cases for which any comorbidity was reported simultaneously in both systems. The agreement percentage of morbidities reported in both SINAVE and SICO databases ranged between 0.0% for HIV and coagulopathy and 5.7% of the total sample for diabetes. When any specific comorbidity was reported in any database, the agreement percentage between this non-reporting in SICO and SINAVE ranged between 73.2% for neurologic or neuromuscular disease and 99.2% for coagulopathy. The disagreement percentages on the reporting of morbidities in both SINAVE and SICO databases ranged between 0.7% for HIV and 24.3% for neurologic and neuromuscular diseases, respectively.

Cohen's kappa was estimated to evaluate the level of similarity of reporting in both databases for each morbidity. The highest agreement level was observed for neoplasms (0.352, 95% IC:0.277–0.428; p < 0.001), followed by hepatic disease (0.282, 95% IC:0.113–0.451; p < 0.001) and either chronic renal disease or acute renal failure (0.268, 95% IC:0.204–0.333; p < 0.001), although the overall levels of agreement were rather low.

#### 4. Discussion

This study investigated the similarity in morbidity reporting between

**Table 5**  
Concordance of comorbidities reporting between SINAVE and SICO databases.

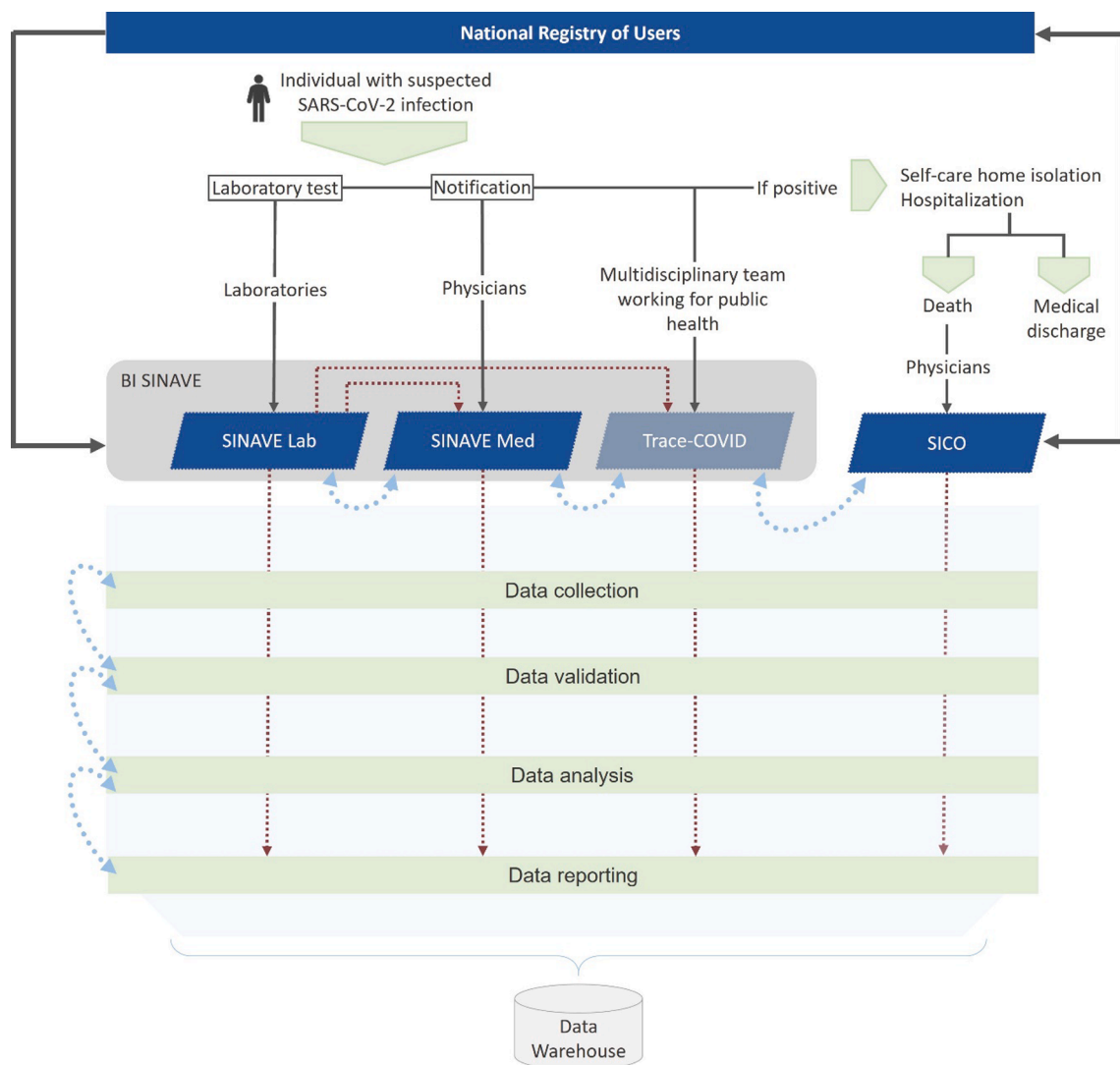
Reported on SINAVE	Neoplasms	No	Reported on SICO		Cohen's kappa
			No n (%) total	Yes n (%) total	
Neoplasms	No	1534 (87.0%)	102 (5.8%)	0.352 (95% IC: 0.277–0.428), p < 0.001	
	Yes	69 (3.9%)	58 (3.3%)		
Diabetes	No	1300 (73.7%)	219 (12.4%)	0.235 (95% IC: 0.179–0.292), p < 0.001	
	Yes	144 (8.2%)	100 (5.7%)		
HIV	No	1750 (99.3%)	4 (0.2%)	*	
	Yes	9 (0.5%)	0 (0.0%)		
Neurologic or neuromuscular disease	No	1291 (73.2%)	359 (20.4%)	0.082 (95% IC: 0.039–0.126), p < 0.001	
	Yes	68 (3.9%)	45 (2.6%)		
Chronic pulmonary disease	No	1533 (87.0%)	98 (5.6%)	0.238 (95% IC: 0.163–0.313), p < 0.001	
	Yes	92 (5.2%)	40 (2.3%)		
Hepatic disease	No	1722 (97.7%)	17 (1.0%)	0.282 (95% IC: 0.113–0.451), p < 0.001	
	Yes	17 (1.0%)	7 (0.4%)		
Chronic renal disease or acute renal failure	No	1439 (81.6%)	164 (9.3%)	0.268 (95% IC: 0.204–0.333), p < 0.001	
	Yes	92 (5.2%)	68 (3.9%)		
Congestive heart failure	No	1494 (84.7%)	250 (14.2%)	0.038 (95% IC: 0.004–0.073), p < 0.001	
	Yes	11 (0.6%)	8 (0.5%)		
Coagulopathy	No	1749 (99.2%)	14 (0.8%)	–	
	Yes	0 (0.0%)	0 (0.0%)		

Footnotes: \* observed concordance is smaller than mean-chance concordance. For Cohen's kappa interpretation, the levels reported by McHugh (2012) were used [23].

two databases used in Portugal for COVID-19 cases surveillance and deaths certification, using a set of four variables to link the information of the same case provided in each of these databases. The results showed minimal or no agreement between the information about each morbidity reported in both databases for each case. However, considering that a significant proportion of cases have morbidities reported either in SINAVE or SICO databases that were not reported in both, the results also revealed a potential to use the health information in both databases to complement each other in describing cases morbidity in individual cases.

The COVID-19 pandemic has challenged the health systems and the effectiveness of surveillance. Therefore, it was necessary to perform some adjustments in Portugal to provide a timely and articulated response from the entire health system. This work was motivated by the dynamic imposed by the COVID-19 pandemic, where ad-hoc systems had to be created to delve into collecting morbidity data for epidemiological surveillance [13]. In addition, human resources responsible for surveillance were strengthened with the allocation of more health professionals and contact tracing trained military personnel [12]. This articulation and adaptation had to be fast to appropriately respond to pandemic health needs and provide public health decision-makers with accurate information.

According to each morbidity reporting and identification method, any morbidity reported for a given case varied between 26.3% and 62.5%. This range can be explained by the differences in the number of morbidities reported, 13 in the SINAVE database, and 17 or 31, i.e., the number of different medical conditions that the CCI and ECI algorithms



**Fig. 2.** Workflow of data from SARS-CoV-2 infections and conceptual model integrated information systems.  
 Footnotes: The red dashed arrows represent the actual information flow, continuous grey arrows the actual that is also considered the ideal flow and the blue dashed arrows the ideal proposed flow. SICO, National e-Death Certificates Information System; SINAVE, National Epidemiological Surveillance System. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

respectively comprise. However, different values for the proportion of morbidity of each SINAVE case were observed when considering either the original variable (where doctors report cases' associated morbidity) or the recoded variable that quantified the presence of any of the 13 reported morbidities variables. For instance, cases were reported as not having morbidities, but some of the 13 morbidity variables were filled in. The opposite was also observed, as doctors who filled in the database reported any morbidity for some cases but did not fill in any of the specific morbidity variables. In this case, it may be because the morbidity could not be reported among the 13 available options. This difference uncovers a potential issue of reporting incompleteness, as the absence of morbidity was reported for cases where specific morbidities were simultaneously reported. Indeed, incompleteness of reporting has been frequently found when assessing the data quality of health information databases before the COVID-19 pandemic [24] and associated with the surveillance of COVID-19 cases in Portugal [25] or in other countries [26,27]. It is widely known that high-quality data is the basis for decision making guidance, especially during a crisis such as a pandemic [28]. However, the priority of public health doctors is to track and isolate their contacts, and, in this way, the surveillance data can be collected in haste. Another aspect concerns data input into the various

systems carried out by multiple sources: physicians, laboratories, or other health professionals working for the public health team during the pandemic. In this context, numerous factors can affect the quality of these data, and numerous unintentional errors can occur, potentially leading to an imprecise conclusion.

Even though the description of the morbidity in individual cases was not within the scope of this study, the results are still in line with those observed in other studies that evaluated the clinical characteristics of deceased people. No differences in the presence of any morbidity across sex or age groups were observed [29].

The CCI and the ECI were chosen to measure pre-existing death-related comorbidities in the SICO database. Those indexes have been described as valuable tools to identify comorbidity from administrative health data [30], especially when ICD codes are available [21,22].

Despite the algorithm differences in CCI and ECI, it was possible to compare the information of both databases regarding ten of the 13 morbidities reported in the SINAVE database. Though this study did not aim to validate the accuracy of the reporting of morbidities through the SINAVE, the analysis allowed us to assess the agreement in identification and reporting agreement of morbidities in both databases. Indeed, the comparison between morbidities reported in SINAVE and identified in

the SICO database showed that 37.5% and 32.7% of the total sample had the same number of morbidities identified or reported in both databases (respectively). However, minimal or no agreement was found when comparing the reported morbidity in both databases. Therefore, the analysed data did not confirm whether these differences in morbidities were due to the data quality, which the authors considered an unlikely hypothesis.

These results suggest differing information regarding the same morbidities in SINAVE and SICO systems, which would complement each other. Some differences can still be expected because different health professionals fill in each database at different disease stages and for distinct purposes. Nevertheless, the results suggest that, for any reason, different health professionals seem to identify and report distinct comorbidities for a given case at different moments, i.e., at the moment of COVID-19 case notification in SINAVE or when certifying a COVID-19 death. This lack of agreement between databases also occurs for chronic diseases that would likely be present in both moments SINAVE and SICO were filled in. Thus, the results suggest that exchanging health information reported in both databases could enhance the available data for surveillance purposes and epidemiologic characterisation of COVID-19 cases. For instance, dataset linkage strategies for COVID-19 epidemiologic research have been applied in Scotland [31] and Sweden for national-level studies [32] to better describe, analyse and model the pandemic evolution. It is expected that the health information exchange between databases describing distinct stages of the disease can allow a longitudinal medical record with information about each case [5] and better knowledge on the impact and importance of their risk factors, comorbidities and their severity [32]. The COVID-19 pandemic has elicited the need of improving the HIS as a preparedness strategy for future pandemics [4]. In addition, this information, together with data on the COVID-19 outcomes, are valuable to support decisions on how to allocate resources, allowing to anticipate hospitalization or intensive care beds occupation and its length, needed treatments, or other long-term outcomes [32].

Still, the need for greater interoperability of different HIS in Portugal through creating a data warehouse was previously identified and is viewed as strategic towards the efficiency of the National Health Service [33]. Health information integration is expected to improve the identification of health problems, population health planning, the policy-making process, or healthcare performance monitoring [33]. This study results show that a more efficient interconnected use of existing information systems can improve data availability and readiness to deal with outstanding health situations. In addition, the results are COVID-19 related but apply to any other disease of interest. Notwithstanding that a strong HIS could sustain evidence-based decisions concerning the direct impacts of COVID-19, i.e., morbidity and mortality, it can be valuable also in describing and supporting strategies to mitigate indirect impacts of the pandemic [34], suggesting how the benefits can go further the COVID-19 response.

In Portugal, the Business Intelligence tool BI SINAVE was created as a more advanced and robust system for processing data [35]. This tool allows the cross-referencing of the information in the databases: SINAVE Lab, SINAVE Med, Trace-COVID and the National Registry of Users (RNU) [Registo Nacional de Utentes], maximising the information from these various sources during pandemic response [35].

Fig. 2 demonstrates the information pathway from an individual with suspected or confirmed SARS-CoV-2 infection and its connection with the different information systems. Suspected SARS-CoV-2 infections in the Portuguese health systems are notified and followed by a laboratory test. A COVID-19 case is notified either by laboratories through SINAVE Lab or physicians through SINAVE Med. Additionally, for clinical management of individuals and contact tracing, Trace-COVID is used. Trace-COVID was developed in the pandemic context and is used by multidisciplinary teams explicitly created to act in the pandemic response to maintain epidemiological surveillance [36]. In practice, the systems are not fully integrated, which causes duplicate or

different entries of information regarding the same COVID-19 cases. Moreover, data collection and validation are carried out independently, which somewhat divides the analysis of the databases and the national data reporting. Therefore, Fig. 2 also represents a conceptual framework for better integration and enhanced health data, with blue dashed arrows representing recommended health information exchange between the distinct health HISs (please, see Fig. 2 footnotes).

Despite not being represented in Fig. 2, other available HIS could still add information and improve this approach. For instance, recent research used electronic medical records, i.e., diagnosis codes associated with hospitalizations, emergency, and outpatient visits documented within the two years before a SARS-CoV-2 infection, to provide prompt morbidity data for each COVID-19 patient [37]. This approach could still be possible in Portugal using ICD-10 diagnosis codes from the Information System for Hospital Morbidity, which gathers each patient morbidity data from all public hospitals in Portugal. In addition, it could also allow saving time throughout several stages of disease progression or during the death certification coding process.

Indeed, a fully linked information system can be the key to providing more agile and precise information to answer health emergency issues more appropriately. A connected system that allows data collection, storage, management, assurance of quality, aggregation, analysis, and continuous data updating are the essential features of an operational HIS [38]. Countries' healthcare systems are in different levels of interoperability between their HIS and can have specific requirements in terms of data policy. Therefore, assessing needs concerning their health data exchange architecture would provide the background for designing strategies for improving HIS interoperability. After perceiving the potential for improved morbidity information with both HIS herein analyzed complementing each other, it could be the next step. However, morbidity data sharing throughout the HIS recurring to a data warehouse, as presented in Fig. 2, could be a strategy to be adopted.

Notwithstanding the benefits conferred by the interoperability of different HIS, other concerns can arise regarding personal data protection. Several strategies can minimize data protection risks, such as regulation, centralization of the linkage and anonymization processes at source, separating new data from personal datasets [39]. For instance, there have been established worldwide regulations for personal data protection, such as General Data Protection Regulation in the European Union [40] or the Health Insurance Portability and Accountability Act in the United States [41].

Limitations of this study warrant discussion. First, the linkage between the databases was based on a set of variables. Therefore, it could not be completely deterministic, as this method cannot uniquely link the information of each case. However, using four variables, the fact that only exact matches were considered and the exclusion of duplicated matches maximised the accuracy of this process. Moreover, the linkage methodology used in this study was innovative and enabled the identification of potential complementarity between linked databases, contributing to achieving the proposed objective. Second, the dependency of data quality on the data entry process could compromise other, specific, and more detailed analyses. However, the potential incompleteness of databases that we found can still be a crucial alert that enhances the need to use data linkage to improve available data and improve evidence-based (COVID-19) surveillance or decision-making processes. Last, due to the criterion used to select cases for the analysis and avoid bias from incomplete coding of morbidity information, the included sample comprised 2285 cases between 4049 matched records. This more narrowed approach of not considering those patients dying in weeks with more than 5% of cases with the incomplete coding process led to the exclusion of a significant number of cases. However, it intended to address the hypothesis that the volume of work within weeks with higher mortality could affect the codification of other cases.



**Table A.1**

Comparison pairs between morbidities in SINAVE and morbidities identified in SICO through Charlson and Elixhauser comorbidities indexes.

SINAVE	SICO Charlson and Elixhauser Comorbidity indexes
Neoplasia	Cancer, for cancer (any malignancy) Metastatic solid tumour Metastatic cancer Solid tumor, without metastasis Lymphoma
Diabetes mellitus	Diabetes without complications Diabetes with complications Diabetes, uncomplicated Diabetes, complicated
HIV	AIDS/HIV AIDS/HIV
Neurologic or neuromuscular disease	Other neurological disorders Hemiplegia or paraplegia Paralysis Dementia
Asthma	NA
Chronic pulmonary disease	Chronic pulmonary disease Chronic obstructive pulmonary disease
Hepatic disease	Mild liver disease Moderate or severe liver disease Liver disease
Chronic hematologic diseases	NA
Chronic renal disease or acute renal failure	Renal disease Renal failure
Congestive heart failure	Congestive heart failure Congestive heart failure
Coagulopathy	Coagulopathy

NA, not available. Chronic neurologic disease was considered as included in “Neurologic or neuromuscular disease” category. Chronic renal disease and acute renal failure were two separated categories but were here included together.

## 5. Conclusion

The COVID-19 pandemic led to reorganizations in HIS to collect data that could provide prompt information to support policy decisions. However, the results of this study show that when data is independently collected, different health information can be found across information systems leading to low accuracy in the morbidity description of each case, due to the range observed in the agreement between the reporting of each morbidity. Therefore, the integration of the two databases would potentially increase their complementarity. However, further research is needed to confirm this hypothesis and if this integration would enhance the interoperability of HIS and the information output, making it more consistent and effective and increasing public health preparedness, as expected. Moreover, the existing communication pathways of different HIS can still be improved, potentially improving the COVID-19 crisis management decision-making, benefiting community sectors, such as the public in general, health professionals, public health researchers, and policymakers.

## Summary points

### *What was already known on the topic*

Health information system interoperability can enhance data regarding patients' morbidity.

Accurate morbidity data allow better decisions and evidence-based public health interventions.

### *What this study added to our knowledge*

Two health information systems were matched on sex, age, residence, and date of death.

Morbidity either in COVID-19 cases or related deaths information systems was compared.

The minimal agreement found suggests a potential for improvement

through data integration.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors would like to thank the Portuguese Presidency for providing the data to perform the present study. We would also like to appreciate the work of Direcção-Geral da Saúde and its professionals involved in data management and preparation.

The authors also thank António C-V Martinho for the support at the beginning of this work.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Appendix A

See [Table A.1](#).

## Appendix B. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijmedinf.2022.104763>.

## References

- [1] J. Zheng, SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat, *Int. J. Biol. Sci.* 16 (10) (2020) 1678–1685, <https://doi.org/10.7150/ijbs.45053>.
- [2] World Health Organization (WHO), Coronavirus disease 2019 (COVID-19): situation report, 51, (2020). <https://apps.who.int/iris/handle/10665/331475>.
- [3] I.W. Nader, E.L. Zeilinger, D. Jomar, C. Zauchner, Onset of effects of non-pharmaceutical interventions on COVID-19 infection rates in 176 countries, *BMC Public Heal.* 211 (21) (2021) 1–7, <https://doi.org/10.1186/S12889-021-11530-0>.
- [4] E. Negro-Calduch, N. Azzopardi-Muscat, D. Nitzan, R. Pebody, P. Jorgensen, D. Novillo-Ortiz, Health Information Systems in the COVID-19 Pandemic: A Short Survey of Experiences and Lessons Learned From the European Region, *Front. Public Heal.* 9 (2021) 1438, <https://doi.org/10.3389/FPUBH.2021.676838/BIBTEX>.
- [5] D.J. Brailer, Interoperability: the key to the future health care system, *Health Aff. (Millwood)*. *Suppl Web 24 (Suppl1)* (2005) W5-19–W5-21.
- [6] J. Walker, E. Pan, D. Johnston, J. Adler-Milstein, D.W. Bates, B. Middleton, The value of health care information exchange and interoperability, *Health Aff. (Millwood)*. *Suppl Web 24 (Suppl1)* (2005) W5-10–W5-18.
- [7] L.C. Baker, Benefits of interoperability: a closer look at the estimates, *Health Aff. (Millwood)*. *Suppl Web Exclusives 24 (Suppl1)* (2005) W5-22–W5-25.
- [8] L.M. Lee, S.B. Thacker, The cornerstone of public health practice: public health surveillance, 1961–2011, *MMWR Suppl.* 60 (4) (2011) 15–21.
- [9] C. Abouzahr, T. Boerma, Health information systems: the foundations of public health, *Bull. World Health Organ.* 83 (2005) 578–583. <https://pubmed.ncbi.nlm.nih.gov/16184276>.
- [10] A.E. Schmidt, L.A. Abboud, P. Bogaert, Making the case for strong health information systems during a pandemic and beyond, *Arch. Public Heal.* 79 (2021) 13, <https://doi.org/10.1186/s13690-021-00531-5>.
- [11] P. Setel, C. Abouzahr, E.B. Atuheire, M. Bratschi, E. Cercone, O. Chinganya, B. Clapham, S.J. Clark, C. Congdon, D. de Savigny, A. Karpati, E. Nichols, R. Jakob, J. Mwanza, W. Muhwava, P. Nahmias, E.M. Ortiz, A. Tshangela, Mortality surveillance during the COVID-19 pandemic, *Bull. World Health Organ.* 98 (6) (2020) 374, <https://doi.org/10.2471/BLT.20.263194>.
- [12] Direcção-Geral de Saúde (DGS), Norma n° 004/2020 de 23/03/2020 atualizada a 19/04/2021 Abordagem do Doente com Suspeita ou Confirmação de COVID-19, 2021.
- [13] Direcção-Geral de Saúde (DGS), Norma n° 015/2020 de 24/07/2020 atualizada a 19/02/2021 COVID-19: Rastreamento de contactos, 2021.
- [14] N. Leon, Y. Balakrishna, A. Hohlfield, W.A. Odendaal, B.-M. Schmidt, V. Zweigenthal, J.A. Watkins, K. Daniels, Routine Health Information System (RHIS) improvements for strengthened health system management, *Cochrane Database Syst. Rev.* 2020 (2020), <https://doi.org/10.1002/14651858.CD012012.PUB2>.

- [15] D. Kebede, C. Zielinski, P.E. Mbondji, I. Sanou, E. Asamoah-Odei, E.W. Soumbeiy-Alley, Improving the availability, quality and use of health information, research evidence and knowledge to strengthen health systems, *African Heal. Monit.* 12 (2010) 53–64.
- [16] M. Dixon-Woods, S. Redwood, M. Leslie, J. Minion, G.P. Martin, J.J. Coleman, Improving quality and safety of care using “technovigilance”: an ethnographic case study of secondary use of data from an electronic prescribing and decision support system, *Milbank Q.* 91 (2013) 424–454.
- [17] Pan American Health Organization, COVID-19 and the importance of strengthening Information Systems, COVID-19 Factsheets. (2020). <https://iris.paho.org/handle/10665.2/52127> (accessed August 20, 2021).
- [18] E. von Elm, D.G. Altman, M. Egger, S.J. Pocock, P.C. Gøtzsche, J. P. Vandembroucke, The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies, *Lancet* (London, England). 370 (9596) (2007) 1453–1457, [https://doi.org/10.1016/S0140-6736\(07\)61602-X](https://doi.org/10.1016/S0140-6736(07)61602-X).
- [19] E.I. Benchimol, L. Smeeth, A. Guttman, K. Harron, D. Moher, I. Petersen, H. T. Sørensen, E. von Elm, S.M. Langan, The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement, *PLOS Med.* 12 (10) (2015) e1001885, <https://doi.org/10.1371/journal.pmed.1001885>.
- [20] World Health Organization, ICD-10 : international statistical classification of diseases and related health problems : tenth revision, (2004) Spanish version, 1st edition published by PAHO.
- [21] A. Elixhauser, C. Steiner, D.R. Harris, R.M. Coffey, Comorbidity Measures for Use with Administrative Data, *Med. Care.* 36 (1) (1998) 8–27, <https://doi.org/10.1097/00005650-199801000-00004>.
- [22] M.E. Charlson, P. Pompei, K.L. Ales, C.R. MacKenzie, A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation, *J. Chronic Dis.* 40 (5) (1987) 373–383, [https://doi.org/10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8).
- [23] M.L. McHugh, Interrater reliability: the kappa statistic, *Biochem. Medica.* 22 (2012) 276, <https://doi.org/10.11613/bm.2012.031>.
- [24] H. Chen, D. Hailey, N. Wang, P. Yu, A review of data quality assessment methods for public health information systems, *Int. J. Environ. Res. Public Health.* 11 (2014) 5170–5207, <https://doi.org/10.3390/ijerph110505170>.
- [25] C. Costa-Santos, A.L. Neves, R. Correia, P. Santos, M. Monteiro-Soares, A. Freitas, I. Ribeiro-Vaz, T.S. Henriques, P.P. Rodrigues, A. Costa-Pereira, A.M. Pereira, J. A. Fonseca, COVID-19 surveillance data quality issues: a national consecutive case series, *BMJ Open.* 11 (2021) e047623, <https://doi.org/10.1136/BMJOPEN-2020-047623>.
- [26] J.A.W. Gold, J. DeCuir, J.P. Coyle, L.M. Duca, J. Adjemian, K.N. Anderson, B. N. Baack, A. Bhattarai, D. Dee, T.M. Durant, R. Ewetola, T. Finlayson, S.W. Roush, S. Yin, B.R. Jackson, K.E. Fullerton, COVID-19 Case Surveillance: Trends in Person-Level Case Data Completeness, United States, April 5–September 30, 2020, *Public Health Rep.* 136 (4) (2021) 466–474, <https://doi.org/10.1177/00333549211006973>.
- [27] T. Clare, K.A. Twohig, A.-M. O’Connell, G. Dabrera, Timeliness and completeness of laboratory-based surveillance of COVID-19 cases in England, *Public Health.* 194 (2021) 163–166.
- [28] O. Morgan, How decision makers can use quantitative approaches to guide outbreak responses, *Philos. Trans. R. Soc. B Biol. Sci.* 374 (1776) (2019) 20180365, <https://doi.org/10.1098/rstb.2018.0365>.
- [29] K. Wang, Z. Qiu, J. Liu, T. Fan, C. Liu, P. Tian, Y. Wang, Z. Ni, S. Zhang, J. Luo, D. Liu, W. Li, Analysis of the clinical characteristics of 77 COVID-19 deaths, *Sci. Rep.* 10 (2020) 16384, <https://doi.org/10.1038/s41598-020-73136-7>.
- [30] M. Yurkovich, J.A. Avina-Zubieta, J. Thomas, M. Gorenchtein, D. Lacaille, A systematic review identifies valid comorbidity indices derived from administrative health data, *J. Clin. Epidemiol.* 68 (1) (2015) 3–14.
- [31] S.J. Stock, D. McAllister, E. Vasileiou, C.R. Simpson, H.R. Stagg, U. Agrawal, C. McCowan, L. Hopkins, J. Donaghy, L. Ritchie, C. Robertson, A. Sheikh, R. Wood, COVID-19 in Pregnancy in Scotland (COPS): protocol for an observational study using linked Scottish national data, *BMJ Open.* 10 (11) (2020) e042813, <https://doi.org/10.1136/bmjopen-2020-042813>.
- [32] F. Nyberg, S. Franzén, M. Lindh, L. Vanfleteren, N. Hammar, B. Wettermark, J. Sundström, A. Santosa, S. Björck, M. Gisslén, Swedish Covid-19 Investigation for Future Insights - A Population Epidemiology Approach Using Register Linkage (SCIFI-PEARL), *Clin. Epidemiol.* 13 (2021) 649–659, <https://doi.org/10.2147/CLEP.S312742>.
- [33] C.S. Pinto, J.P. Martins, H. Martins, From Big Data to Smart Health: Putting Data to Work for the Public’s Health, Data Strategy for Next Generation, Shared Services of the Ministry of Health, Lisbon, 2019. [https://www.spms.min-saude.pt/wp-content/uploads/2020/07/Data-Strategy\\_VERSAOFINAL\\_07.01.2020.pdf](https://www.spms.min-saude.pt/wp-content/uploads/2020/07/Data-Strategy_VERSAOFINAL_07.01.2020.pdf).
- [34] C.F. Ferraro, L. Findlater, R. Morbey, H.E. Hughes, S. Harcourt, T.C. Hughes, A. J. Elliot, I. Oliver, G.E. Smith, Describing the indirect impact of COVID-19 on healthcare utilisation using syndromic surveillance systems, *BMC Public Health.* 21 (2021) 1–11, <https://doi.org/10.1186/S12889-021-12117-5/FIGURES/6>.
- [35] Direção-Geral da Saúde, BI SINAVE entrou hoje em funcionamento - COVID-19, (2020). <https://covid19.min-saude.pt/bi-sinave-entrou-hoje-em-funcionamento/> (accessed July 15, 2021).
- [36] Trace COVID-19 – SPMS, (2020). <https://www.spms.min-saude.pt/2020/07/trace-covid-19/> (accessed September 24, 2021).
- [37] B.E. Dixon, S.J. Grannis, C. Mcandrews, A.A. Broyles, W. Mikels-Carrasco, A. Wiensch, J.L. Williams, U. Tachinardi, P.J. Embi, Leveraging data visualization and a statewide health information exchange to support COVID-19 surveillance and response: Application of public health informatics, *J. Am. Med. Informatics Assoc.* 28 (2021) 1363–1373, <https://doi.org/10.1093/JAMIA/OCAB004>.
- [38] United Nations Development Programme United Nations Development Programme, Key elements of a health information system, (2021). <https://www.undp-capacitydevelopment-health.org/en/capacities/focus/health-information-systems/key-elements-of-a-health-information-system/> (accessed July 30, 2021).
- [39] World Health Organization. Regional Office for Europe, The protection of personal data in health information systems- principles and processes for public health, World Health Organization. Regional Office for Europe, 2021. <https://apps.who.int/iris/handle/10665/341374>.
- [40] European Union, General Data Protection Regulation (GDPR) Compliance Guidelines, (2022). <https://gdpr.eu/> (accessed January 26, 2022).
- [41] U.S. Department of Health & Human Services, HIPAA Home | HHS.gov, (2022). <https://www.hhs.gov/hipaa/index.html> (accessed January 26, 2022).