



Evidence of SARS-CoV-2 reinfection: analysis of 35,000 subjects and overview of systematic reviews

Valentina Pecoraro¹ · Tommaso Pirotti¹ · Tommaso Trenti¹

Received: 14 April 2022 / Accepted: 11 October 2022 / Published online: 26 October 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract

Background Reinfection by SARS-CoV-2 is a rare but possible event. We evaluated the prevalence of reinfections in the Province of Modena and performed an overview of systematic reviews to summarize the current knowledge.

Methods We applied big data analysis and retrospectively analysed the results of oro- or naso-pharyngeal swab results tested for molecular research of viral RNA of SARS-CoV-2 between 1 January 2021 and 30 June 2021 at a single center. We selected individuals with samples sequence of positive, negative and then positive results. Between first and second positive result we considered a time interval of 90 days to be sure of a reinfection. We also performed a search for and evaluation of systematic reviews reporting SARS-CoV-2 reinfection rates. Main information was collected and the methodological quality of each review was assessed, according to A Measurement Tool to Assess systematic Reviews (AMSTAR).

Results Initial positive results were revealed in more than 35,000 (20%) subjects; most (28%) were aged 30–49 years old. Reinfection was reported in 1,258 (3.5%); most (33%) were aged 30–49 years old. Reinfection rates according to vaccinated or non-vaccinated subjects were 0.6% vs 1.1% ($p < 0.0001$). Nine systematic reviews were identified and confirmed that SARS-CoV-2 reinfection rate is a rare event. AMSTAR revealed very low-moderate levels of quality among selected systematic reviews.

Conclusions There is a real, albeit rare risk of SARS-CoV-2 reinfection. Big data analysis enabled accurate estimates of the reinfection rates. Nevertheless, a standardized approach to identify and report reinfection cases should be developed.

Keywords Reinfection · SARS-CoV-2 · Big data analysis

Background

One of the most interesting aspects of the COVID-19 pandemic is that a variable percentage of patients (from 2 to 69%) could have a repeated positivity following hospital discharge or even several weeks after clinical recovery [1]. There are multiple reasons why a positive result to SARS-CoV-2, usually ascertained by reverse transcriptase-polymerase chain reaction (RT-PCR), may be detected again,

including reinfection, disease reactivation, prolonged viral shedding or false positive results [1–3].

The Centers for Disease Control and Prevention (CDC) suggest that an epidemiological and clinical reinfection case may be suspected in a person with at least one detection of SARS-Cov-2 RNA more than 90 days after the first detection (with or without symptoms) or in persons with COVID-19-like symptoms and detection of SARS-CoV-2 RNA between 45 and 89 days from the first SARS-CoV-2 infection, with evidence of close-contacts with a confirmed case and without evidence of another cause of infection [4]. For confirmation of reinfection, viral genotype assays of the first and second specimens are required [5]. However, the lack of protective immunity due to scarce development or a rapid decay of antibodies could lead to new infection with the same species and strain (recurrence) [6]. The Collaborative COvid REcurrences (COCOREC) study proposed that a COVID-19 recurrence may be considered if the second episode occurs within 21 days following a symptom-free

✉ Valentina Pecoraro
v.pecoraro@ausl.mo.it

Tommaso Pirotti
t.pirotti@ausl.mo.it

Tommaso Trenti
t.trenti@ausl.mo.it

¹ Department of Laboratory Medicine and Pathology, Azienda USL of Modena, Modena, Italy

period and where alternative etiologies are excluded [7]. However, the possibility of reactivation of a latent infection or relapse has been considered as a potential consequence of the presence of non-replicative viral traces up to a maximum of 6 weeks after the onset of symptoms [8, 9]. Similarly, the presence of viral RNA in the nasopharyngeal swabs for a prolonged period after infection has been observed in numerous patients (persistent positivity) [10].

Since the beginning of the pandemic, several authors have reported the possibility of reinfection by SARS-CoV-2 or reactivation of a latent infection [8]. These studies highlighted possible reinfection and reactivation of SARS-CoV-2, calling for urgent attention from researchers, as well as public health policymakers.

Big Data analysis has been widely used for pandemic monitoring; from predicting new trends in the spread of infection, to periodically updating the epidemiological situation to assist institutional governance in allocation of health resources decision-making [11]. The term Big Data not only indicates a large amount of data but defines a more complicated concept, described by the 5 Vs concept: (i) velocity (i.e. the speed of data acquisition, processing and manipulation), (ii) volume (i.e. large amount of information available), (iii) variety (data from different sources and in different formats), (iv) veracity (quality of data, free of errors); and (v) value (possibly bringing benefits and producing knowledge) [12]. The availability of large amounts of data, in association with adequate informatic tools, enables the application of Big Data analyses to laboratory data, thereby producing useful information for the study, control and monitoring of the SARS-CoV-2 infection rates [11]. Therefore, this approach of managing and analysing a large database was chosen for the current study.

Thousands of naso- and oro- pharyngeal swabs have been performed at the Department of Laboratory Medicine of the Azienda USL of Modena since the early stages of the pandemic, and a large amount of data has been collected. We applied big data analysis to estimate the real incidence of reinfection by SARS-CoV-2 in the population of the Province of Modena, to understand the weight of protective immunity in this pandemic. Also, we performed a review of current systematic reviews reporting reinfection with SARS-CoV-2, to summarize the current knowledge/findings about reinfection.

Methods

Analysis of laboratory results

The laboratory of the Department of Laboratory Medicine of the Azienda USL of Modena has a database which collects about 15 million results per year. These are tests carried

out in all the laboratories in the same province, serving a population of about 700,000 residents. The database is continuously updated every day, enabling users to know at any moment how many exams are booked, how many are in progress and how many have been reported. Users can also check this same information for the previous days and months.

For this study, we retrospectively analyzed the results of the oro- or naso-pharyngeal swabs performed for the determination of SARS-CoV-2 RNA in samples collected over a period of 18 months (1st January 2021–30th June 2021), in subjects for which the molecular RNA research was requested for the diagnosis of SARS-CoV-2 infection. We identified and selected subjects with a positive result, a subsequent negative result, and then another positive result. Reinfection was defined as a second positive result > 90 days from the initial positive result [4]. To identify possible reinfection of enrolled subjects, data analyzed was extended up to 22th February 2022. Data regarding the vaccination status of the positive subject was also acquired. Subjects were grouped according to age categories: 0–14, 15–29, 30–49, 50–69 and ≥ 70 years old.

Total RNA was extracted from the clinical samples using a commercial RNA-extraction kit and was reverse transcribed. The cDNA was then amplified by real-time qualitative PCR, using a commercial kit (Alinity m SARS-CoV-2 Assay—Abbott Molecular). The procedures and interpretation of the results were carried out in compliance with the manufacturers' instructions.

For this retrospective and observational study, the sample size was not calculated a priori, but all data collected in the database up to 22th February 2022 were considered. Due to the lack of informed consent and the impossibility of its subsequent acquisition, data were pseudonymized. One researcher extracted data from the database using specific search queries without including any personal identification codes. A second researcher analyzed the data extracted. Any researcher using the data was able to trace the identity of the subject analyzed.

We calculated the frequency of reinfection of SARS-CoV-2 over the study period. We also stratified data into subgroups based on subjects' vaccination status. The reinfection rate was determined by dividing the number of reinfected subjects by the total number of initially positive subjects.

Review of systematic reviews

We conducted an overview of systematic reviews (SRs) according to the Cochrane methodological guidance [13], reporting the findings according to applicable items from the PRISMA statement [14].

This overview included any type of SR that reported the number of reinfections, defined as a positive RT-PCR test

carried out > 90 days after the initial test, in subjects who had been considered cured or a positive RT-PCR test carried out > 45 days after the initial positive test, accompanied by compatible symptoms or epidemiological exposure [4]. Further inclusion criteria specified that SRs: (i) provided specific eligibility criteria allowing to define the clinical question; (ii) described a search strategy; (iii) reported sufficient details for included studies; (iv) reported the definition of reinfection; and (v) reported the pooled estimate or the exact number of reinfections. Eligibility was not restricted by language, patient age or study setting. We excluded any SRs that did not report sufficient information about reinfections, were narrative reviews, evaluated sample specimens or included animal models.

To identify all SRs of interest, we searched the clinical queries of MedLine, including appropriate filters for SRs conducted on COVID-19. The search strategy included the following keywords “reinfection”, “COVID-19”, “SARS-CoV-2”. Further, the reference lists of potentially eligible SRs were also screened. We limited the search to studies published between 2020 and 2022. The literature search was conducted by one investigator in February 2022. One author screened the titles and abstracts of SRs retrieved from the database searches and selected the studies for inclusion according to eligibility criteria. A second author checked the selection. Disagreements were resolved by consensus.

From each included SR, one author extracted the necessary data, and a second author validated the data collected. The following information was recorded: (1) characteristics of SR (authors, year, country); (2) definition of reinfection; (3) methodological design of included studies (i.e., case report, cohort); (4) characteristics of participants (i.e., sample size, age, gender, symptoms); (5) investigated outcomes

(number of reinfections, time interval between first and second infection). Disagreement between reviewers was resolved by consensus.

One author assessed the methodological quality of the included SRs, using the AMSTAR 2 tool [15]. A second author checked the evaluation. AMSTAR 2 does not generate an overall score, and we report an overall rating of confidence in the results of the SRs as follows: high (zero non-critical weaknesses), moderate (≥ 1 non-critical weakness), low (1 critical flaw with or without non-critical weaknesses), very low (> 1 critical flaw with or without non-critical weaknesses). We synthesised data into tables, including SR characteristics and summarised narrative findings, according to quality and outcomes of interest for the current overview.

Results

Analysis of laboratory data

During the 18-month study period, 178,948 subjects (25.5% of all assisted subjects in the Province of Modena) performed a molecular test for the search for SARS-CoV-2 RNA. Of these, 20% ($n = 35,743/178,948$) had at least one dose of vaccine for SARS-CoV-2 up to 30th June, especially subjects over the age of 70. During the study’s time period, vaccines were not available for children < 14 years old (Table 1).

The 20% (35,692) of tested subjects had a positive PCR result for SARS-CoV-2, with highest among the 50–69 age group ($n = 9488/40813$, 23%). Among all positive subjects, 5% ($n = 1,794/35,692$) were vaccinated. Likewise, among

Table 1 Subjects tested, infected and reinfected, according to age group

Range of age, years	All	0–14	15–29	30–49	50–69	> = 70
N. of tested subjects	178,948	32,825	32,406	47,145	40,813	25,759
Vaccinated						
Up to 30/06/2021, <i>n</i> (%)	35,743 (20%)	0 (0%)	3868 (12%)	9811 (21%)	10,905 (26.7%)	11,153 (43.3%)
Up to 22/02/2022, <i>n</i> (%)	141,936 (79%)	16,291 (49.6%)	27,753 (85.6%)	40,305 (85.5%)	35,769 (87.6%)	21,818 (84.7%)
Unvaccinated						
Up to 30/06/2021, <i>n</i> (%)	143,205 (80%)	32,825 (100%)	28,538 (88%)	37,334 (79.2%)	29,908 (73.3%)	14,606 (56.7%)
Up to 22/02/2022, <i>n</i> (%)	37,012 (20.7%)	16,534 (50.4%)	4653 (14.4%)	6840 (14.5%)	5044 (12.4%)	3941 (15.3%)
Infection, <i>n</i> (%)	35,692 (20%)	4926 (15%)	6347(19.6%)	10,166 (21.5%)	9488 (23.2%)	4765 (18.5%)
Male, <i>n</i> (%)	17,718 (49.7%)	2563 (52%)	3330 (52.5%)	4881 (48%)	4752 (50%)	2192 (46%)
Vaccinated, <i>n</i> (%)	1794 (5%)	0 (0%)	112 (2.9%)	431 (4.4%)	480 (4.4%)	771 (7%)
Unvaccinated, <i>n</i> (%)	33,898 (23.7%)	4926 (15%)	6235 (21.9%)	9735 (26%)	9008 (30.1%)	3994 (27.3%)
Reinfection, <i>n</i> (%)	1258 (3.5%)	208 (4.2%)	252 (4%)	414 (4%)	246 (2.6%)	138 (2.9%)
Male, <i>n</i> (%)	585 (46.5%)	110 (52.9%)	117 (46.4%)	173 (41.8%)	117 (47.6%)	68 (49.3%)
Vaccinated, <i>n</i> (%)	833 (0.6%)	31 (0.2%)	203 (0.7%)	330 (0.8%)	176 (0.5%)	93 (0.4%)
Unvaccinated, <i>n</i> (%)	425 (1.15%)	177 (1.07%)	49 (1.05%)	84 (1.23%)	70 (1.4%)	45 (1.14%)

vaccinated subjects, 5% ($n = 1794/35,743$) had a positive result on molecular test (Table 1).

Of the subjects with an initial infection, 3.5% ($n = 1,258/35,692$) had reinfection, with the highest incidence among the children groups (Table 1, Fig. 1). Among re-positive subjects, 66.2% (833/1,258) were vaccinated.

At the closure of the extended data analysis for reinfections (up to 22th February 2022), 79% of tested subjects (141,936/178,948) were vaccinated. Reinfection rates for vaccinated and non-vaccinated subjects were 0.59% (833/141,936) and 1.15% (425/37012) respectively (Fig. 2).

The mean time between the initial and second positive molecular results according to age groups was 248 days

(> 70 years), 313 days (< 14 years), 309, 310 and 302 for subjects with age range from 15 to 69 years, respectively.

Synthesis of systematic reviews

The literature search identified 22 references. Of these, 3 were excluded because they did not meet the inclusion criteria. There were 19 references considered eligible for inclusion and details were obtained from full texts. Further, 10 texts were excluded, leaving a total of 9 SRs [16–24] selected for this overview. (Fig. 3). All included publications reported the database and data search, the description of inclusion criteria and the confirmation of the reinfection by RT-PCR. Three SRs defined the reinfection following CDC

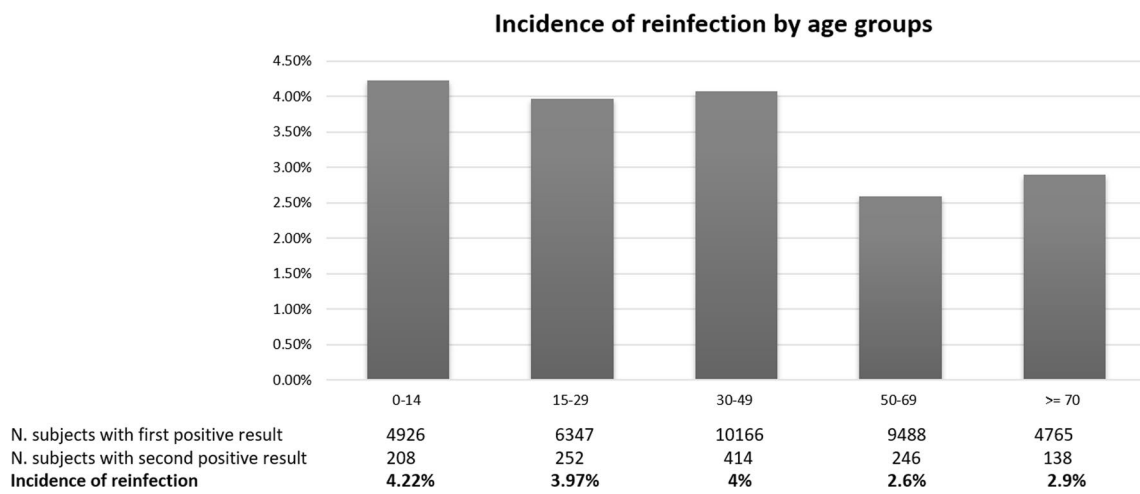


Fig. 1 Incidence of reinfection by age

Fig. 2 Incidence of reinfection by vaccination status

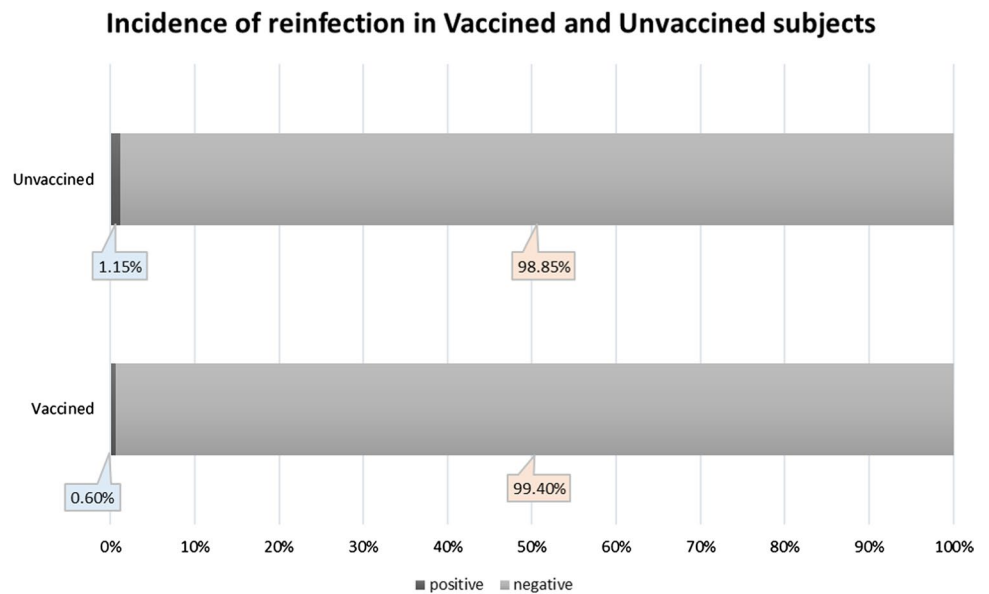
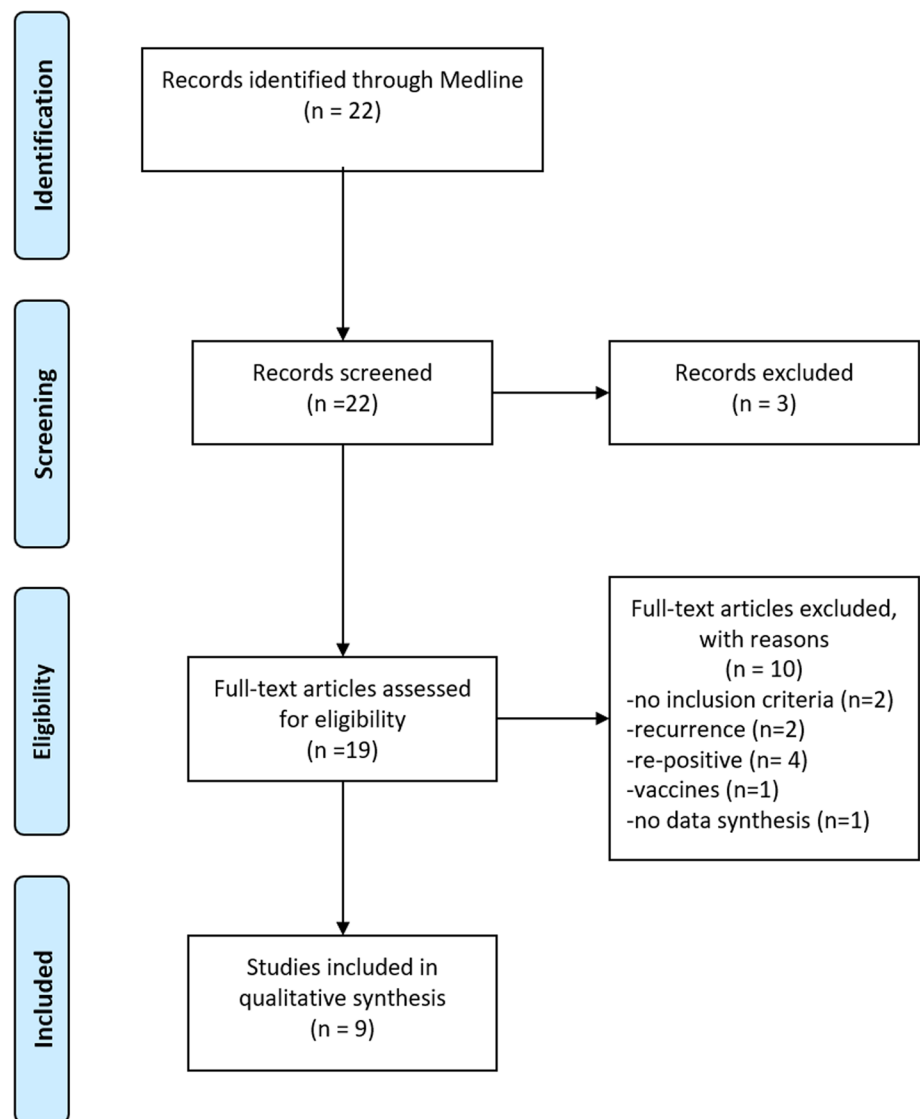


Fig. 3 PRISMA flow diagram of the selection process for systematic reviews



criteria, three reported a general definition, one used the RKI definition, and two did not report any definition (Table 2).

The methodological quality of SR was judged moderate for 5 SRs, low for 3 SRs and very low for one SRs.

The selected SRs reported few cases of reinfections in studies published up to August 2021. The reporting of this information is more heterogeneous among SRs included. Cases ranged from 35 [17] to 260 [20]. Abrokwa et al. [16] reported that the rate of re-positive result ranged from 0 to 50%, likewise, Mao et al. [21] showed a pooled reinfection rate of 0.65%. Sotoodeh Ghorbani et al. [23] estimated reinfection to be 3 per 1000 patients (95% CI 0.8–5), and Chivese et al. [18] reported that patients previously infected by SARS-CoV-2 had a 81% odds reduction for reinfection (Table 2). Two SRs reported that protection against reinfection was 87% [21] and 90.4% [19], respectively. Authors agreed that the protective effect of prior SARS-CoV-2

infection on re-infection is high and similar to the protective effect of vaccination.

Four SRs [17, 20, 22, 24] reported information about the severity of symptoms in reinfection compared with initial infection; two reported similar severity [17, 24], whereas Lo Muzio et al. [20] reported less severe symptoms with reinfection. Massachi et al. [22] reported the same percentage of patients experiencing either greater or milder symptoms with reinfection.

Discussion

The available literature for SARS-CoV-2 reinfection risk suggests that although rare, reinfection is possible, but available estimates vary considerably and are mostly based on case reports/series and some cohort studies. We

Table 2 Characteristics of systematic review included in the overview

Author	Abrokwa 2021	Chamblain 2021	Chivese 2022	Kojima 2021	Lo Muzio 2021	Mao 2022	Massachi 2021	Sotoodeh Ghorbani 2021	Wang 2021
Database and data search	PubMed, Embase, preprint servers (01/DIC/2019—2020)	Pubmed (up to 30/NOV/2020)	Cochrane Library, WOS Clarivate, Cina Academica Journals Full Text Database, PubMed, Scopus, preprint servers (01/JAN/2020—01/APR/2021)	PubMed, preprint servers, MedRxiv, (up to 18/AUG/2021)	PubMed, Scopus, Web of Science Direct, EMBASE, preprint servers (MedRxiv, BioRxiv, and SSRN) (up to 31/JUL/2021)	PubMed, EMBASE, Cochrane Library, Scopus, Web of Science, and ClinicalTrials.gov (up to 01/MAY/2021)	The PubMed Medline, Embase, Web of Science, and medRxiv (up to 01/MAY/2021)	PubMed, Web of Science, Scopus, and Google Scholar (up to 06/MAR/2021)	PubMed, MedRxiv, and Social Science Research Network from January 1, 2020, to October 12, 2020,
Inclusion criteria	nr	Case reports or case series of reinfection published as either letters to the editor or original studies	Observational studies reported prevalence of IgG, CD4 +, CD8 + and memory cell B	(1) Baseline PCR testing, (2) uninfected comparison group, (3) longitudinal follow-up, (4) a human cohort, i.e. no case reports or case series, and (5) outcome determined by PCR	Positive RT-PCR testing in recovered patients either (i) > 90 days after initial test or (ii) positive RT-PCR test > 45 days from initial test, accompanied by symptoms or epidemiological exposure	(1) Studies reporting <i>n</i> of COVID-19 reinfections (US CDC criteria) and (2) original research including cohort, ecological, and cross-sectional studies	Sequencing \geq moderate quality evidence (according to CDC investigative criteria) or had supporting evidence of reinfection	(1) studies published in English; (2) peer-reviewed and published articles included desired indicators and definitions	Only studies with viral genome sequencing available for both initial and reinfection events
Definition of reinfection	RKI definition	nr	Positive PCR test \leq 3 months after recovery (documented by negative PCR test)	nr	US CDC criteria (protocol designed to identify real cases of SARS-CoV-2 reinfection)	\geq 90 days after initial positive test or \geq 45 days (with background contact information with confirmed cases or the reappearance of COVID-19 like symptoms)	CDC investigative criteria	A person was infected with an agent, recovered, and then become infected later	Individuals infected with different genetic strains of SARS-CoV-2, confirmed by PCR
Reinfection confirmed by PCR	Yes	nr	yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2 (continued)

Author	Abrokwa 2021	Chamblain 2021	Chivese 2022	Kojima 2021	Lo Muzio 2021	Mao 2022	Massachi 2021	Sotoodeh Ghorbani 2021	Wang 2021
Genome sequencing	Yes	nr	nr	nr	Yes	nr	Yes	nr	Yes
Study design	Case report, case series, observational studies, review	Case report, case series	Case series, cross sectional, cohort	Cohort studies	Case report, case series	Cohort, cross sectional studies	Cohort, case series, report	Observational studies	Case report
Methodological quality (guidelines) of included studies	Joanna Briggs Institute (JBI) rate from 6 to 10	nr	MASTER	nr	Murad's quality checklist	Newcastle–Ottawa Scale	nr	STROBE checklist	nr
N publication included	133	17	54 (5 including reinfection)	10	117	19	35	25	17
Patients with reinfection	237; rate of positives range 0–50%. Pooled recurrence rate range 12–17.7%	35 cases	pooled OR 0.19 (95%CI 0.11–0.32)	relative decreased risk of reinfection range 80.5–100%, compared to those without prior infection	260	1,096/325,225 (0.33%). Incidence of reinfection among hospitalized COVID-19 patients, range 0–20%	60	prevalence of reinfection 3/1,000 (range 0.8–5)	17
Interval between first and second infection	12–33 weeks	median 2 months (IQR 1.5–5 months)	7–8 months	median observation period range 1–10.3 months	>45 days	nr	median 116 days (range 18–308)	45–212 days	mean 76 days (range 19–142)

Table 2 (continued)

Author	Abrokwa 2021	Chamblain 2021	Chivese 2022	Kojima 2021	Lo Muzio 2021	Mao 2022	Massachi 2021	Sotoodeh Ghorbani 2021	Wang 2021
Disease Severity	nr	69% = mild symptoms at initial infection vs 37% = mild symptoms at reinfection	Nr	Nr	reinfection occurred in 80% (207/260) with asymptomatic/mild disease and 20% (53/260) with moderate/severe disease at initial infection. The observation that many reinfection cases were less severe than initial cases suggests possible partial protection from disease	Nr	41.5% (22/53) = greater symptom burden than initial infection; 17% (9/53) = same symptom burden for both the initial and reinfection; 41.5% (22/53) = milder symptom burden at reinfection	nr	68.8% (11/16) = similar severity; 18.8% (3/16) = worse symptoms; 12.5% (2/16) = milder symptoms with reinfection
Methodological quality of SR (AMSTAR)	Moderate	Very low	Moderate	Low	Moderate	Moderate	Low	Moderate	Low

retrospectively analyzed a large repository of complete data for a single province, with the assistance of big data analysis, identifying more than 35 thousand subjects with an initial positive result for SARS-CoV-2, registered over an 18 months period, and with extended analysis of at least 6 months follow-up. We report 1,258 cases of SARS-CoV-2 reinfection (3.5%) compared to a total of 1,705 cases of reinfection reported among the SRs included in our SR overview. Despite inherent limitations of big data analysis, our observation suggests that its application in the context of a complete database of reported infections for a single population can more accurately estimate true rates of SARS-CoV-2 reinfection.

Multiple questions regarding reinfection associated with SARS-CoV-2 are still ongoing. What is the pathophysiological mechanism for reinfection, who are the subjects with a higher risk of reinfection and what is the clinical burden for reinfected patients? Reinfection with the SARS-CoV-2 virus can be mainly attributed to two phenomena: decay of the immune response and viral mutations that favor the appearance of new variants. The decay of immunity or the failure of naturally acquired immunity may result in reinfection with the same virus strain [25], whilst viral mutations may make subjects vulnerable to reinfection [26, 27]. New virus variants could evade immune responses acquired in subjects with infections from previous variants or reduce the capacity for neutralization by polyclonal antibodies [26, 28].

To understand the causes of reinfections it is necessary to know the nature, duration and kinetic of the anti-SARS-CoV-2 antibody response, as well as the variables associated with the virus itself. Based on the several reinfection cases reported, now we know that the possible causes of reinfection include the decay of antibodies titer, the exposure to a higher dose of the virus, immunological comorbidities associated with some patients and the diffusion of more virulent virus due to new genome mutations that allow them to evade the host's immune response [29]. The viral variations occur as result of nucleotide changes in the viral genome during replication and confer advantages with respect to viral replication, transmission, and immune evasion. Most of the mutations, occurring at the Spike (S) protein and especially in the Receptor-Binding Domain (RBD), could affect the entry of the virus into the host cells and the efficacy of vaccines and neutralizing antibodies. Since the beginning of the pandemic, several viral variations have followed one another. CDC distinguishes between variant of concern (VOC) and variant of interest (VOI) based on their ability to cause severe disease, infectivity, or reduced Abs response. In particular, VOC is associated with higher transmissibility, severe disease, and escaping natural and vaccine-induced immunity. Alpha (B.1.1.7) variant has 23 mutations and it is associated with an increase transmissibility, hospitalizations, mortality rates and burden to health care systems. The Beta

(B.1.351) variant has mutations in S, N, E e ORF proteins. In particular, the N501Y in RBD confers an increased binding affinity for the ACE2 receptor, and E484K mutation is associated with reduced vaccine efficacy and increase immune escape. Similarly, the Gamma (P.1) variant contained more mutation in S protein, among which E484K that is associated with immune evasion and higher risk of reinfection. The Delta (B.1.617.2) variant is more virulent than other variants, is associated with severe disease, hospitalization and resistance to preventive measures. Lastly, the Omicron (B.1.1.529) variant has 32 mutations in the spike (S) protein and is associated with higher virulence and increase risk of re-infection. The VOC share common mutations in S protein. The first is N501Y founding in the ACE2 binding site of the RBD and is common to Alpha, Beta, Gamma and Omicron variant. The second and third, E484K/Q/A and K417T/N, are present in the Beta, Gamma and Omicron variants. The fourth, L452R, is unique to the Delta variant [30].

The efficacy of current vaccines is due to the ability to stimulate the neutralizing antibodies production against the S protein of the SARS-CoV-2 wild-type strain. The advent of new viral variants, evading the immunity acquired from previous infection, could reduce the efficacy of vaccine and cause re-infection, with heterogeneous clinical severity and great difficulties for healthcare system.

This issue suggests the need to increase the current knowledge about the degree of protection provided against SARS-CoV-2, leading the development of vaccines and the creation and implementation of appropriate interventional strategies.

Current evidence confirms that patients infected by SARS-CoV-2 produce antibodies against Spike and N proteins within 30 days from infection [31] but the mechanism of mediate immunity are not fully understood. Infection by SARS-CoV-2 activates T and B cells, leading to the production of neutralizing protein inhibiting viral infectivity through various mechanisms of action, including blocking the binding of the Spike protein with the ACE2 receptor [32]. IgM appears quickly but has a very short half-life. Specific IgG develops a few days after IgM and can be determined in serum about 7–14 days from symptom onset [33]. A recent systematic review [34] reported differences in the presence of antibodies during the first infection (56%) and reinfection (63%), suggesting that waning antibodies could place individuals at a risk for reinfection. The presence of antibodies could provide a protective role, but it does not specifically prevent reinfection [35]. Furthermore, it has also been suggested that a previous COVID-19 infection may not confer total immunity, paving the way for a potential second infection by a different variant, with the second infection being potentially more severe than the first [24].

Currently, there are discordant rates of reinfection reported in SRs (ranging from 0 to 50%), which could

partially be explained by the heterogeneous adopted definitions of reinfection. Today, there is still no universal agreement on the determination of the correct time period between positive results for SARS-CoV-2 for the definition of reinfection, although the definition provided by CDC is the most accredited [4]. Further, most SRs mainly include case series or case reports, with limited examples of reinfection. Our big data analysis was conducted in a unique environment of complete data stored in a single warehouse including all SARS-CoV-2 testing with PCR in a single province, analyzed according to the most commonly accepted definition of reinfection in literature. With the collection of a large number of reinfected cases, possible causes, important information for the discrimination of reinfection from recurrence, and the definition of subjects with higher risk of reinfection can be evaluated.

It has been pointed out that the severity of reinfection depends on the individual immune response, as well as both the viral load and the SARS-CoV-2 variants causing the reinfection [36]. A reinfection can then be of the same or greater intensity, and it is probable that it is mostly due to a new species of coronavirus [37]. Garduno-Orte et al. [38] described 4 cases of reinfected healthcare workers, showing that in two cases the reinfection resulted in a more severe case. Likewise, Massachi et al. [22] reported that the 41% of reinfected patients experienced greater symptom burden than initial infection. Conversely, Wang et al. [24] reported that 69% had similar severity, 19% had worse symptoms, and 12% had milder symptoms with a second episode. Our study does not include information regarding disease severity, making any examination of the clinical and social implication of SARS-CoV-2 reinfection impossible to make.

Our study does, however, include subjects' vaccination status, providing important considerations for the risk of reinfection provided by natural immunity and vaccines. In this study, the rate of reinfection among vaccinated subjects was lower than that observed among non-vaccinated subjects. If antibody decay is associated with susceptibility to reinfection, we may observe further reinfections over the next months. Likewise, if the immune response vaccine-induced is likely to decay as the natural immune response, the need for booster immunization will require re-evaluated to maintain ongoing protection.

Our work is an example of application of big data analysis in laboratory setting, enabling real estimates of incidence of reinfection, to identify factors affecting reinfection, such as strains of the virus or patient immune characteristics, and ponder the involvement of the vaccine in this pandemic. The Big Data concept refers to a complex analysis of a huge set of data, which requires the use of dedicated analytical and statistical approaches. The method uses advanced computational methods to extract information from datasets and build new association models. There are multiple sources

of data (administrative databases, electronic health records, epidemiological studies), so it is important to develop an appropriate integration and analysis system to translate the information from analysed data to appropriate clinical decisions.

The growing data availability and greater analytical capacity, can improve results not only in the economic and financial fields, but also in public health, supporting diagnostic pathways, developing prognostic predictive models of disease, personalizing therapeutic regimens and, can also find an application in prevention initiatives [39, 40]. The application of Big Data analysis in healthcare has numerous advantages as it enables: (i) the integration of different datasets and builds different algorithms and more complex learning models to find new genetic, biological and clinical associations [41], (ii) direct analyses of data from an entire population, overcoming the limitations associated with statistical approaches applied to the analysis of data from a representative sample to make inference on a population (even if randomized controlled trials remain the "gold standard" to study treatment efficacy), (iii) the observation of effects of long-term treatments. However, this approach has limitations due to the high variability of data and data collection methodologies. These limits can be resolved by the use of adequate computation systems, thereby helping to reduce bias and make data more functional.

In the interpretation of our results, some limitations due to the lack of information about symptoms and immunological status of subjects analyzed, and the viral strain causing infection and reinfection, should be considered. The definition of a positive result 90 days after an initial infection as reinfection used in this analysis cannot exclude a possible reactivation of a latent infection. Further, the true rate of infection is assumed to be underestimated, as many asymptomatic subjects are not tested for viral RNA research and, among those tested, genomic sequencing is not always performed, rendering the identification of the precise variants causing infection and reinfection very difficult to make. Considering the official data from Istituto Superiore di Sanità, between January and June 2021, in Italy, the most common variant was the alpha variant (88.1%) followed by gamma variant (7.3%), instead between January and March 2022 was the omicron variant (98.3%) (<https://www.epicentro.iss.it/coronavirus/sars-cov-2-monitoraggio-varianti-rapporti-periodici>), and probably the re-positive results observed in our study may be re-infections. The third factor to consider in the evaluation of reinfection, is a potential false negative molecular result at discharge and a subsequent positive result being due to persistent infection [42].

Our big data analysis of a complete population confirms an overall risk of reinfection by SARS-CoV-2 of 3.5%, with unvaccinated and younger subjects more susceptible to reinfection. More data will become available over time,

and big data analysis will enable its timely integration into considerations of targeted strategies to control and prevent reinfection, increasing value in the patient's care pathway and supporting healthcare systems. In the meanwhile, social distancing, the use of masks and hand hygiene remain the main preventative measures against primary infection and reinfection of SARS-CoV-2. A standardized approach to identify and report reinfection cases is necessary.

Acknowledgements The authors would like to thank Johanna Chester for her critical revision and editing assistance.

Author contributions All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission. VP conceived and designed the study and the systematic review. TP extract data from laboratory database, VP analysed data. VP designed and implemented the search strategy for the overview of systematic reviews. VP and TP extracted and analysed the data of systematic review identified. VP wrote the paper. All authors were involved in the critical revision of the intellectual content of the manuscript.

Funding The authors received no financial support for the research, authorship, and/or publication of this article.

Declarations

Competing interests The authors declare no competing interests.

Conflict of interest The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Dao T, Hoang V, Gautret P. Recurrence of SARS-CoV-2 viral RNA in recovered COVID-19 patients: a narrative review. *Eur J Clin Microbiol Infect Dis*. 2021. <https://doi.org/10.1007/s10096-020-04088-z>.
2. Lu J, Peng J, Xiong Q, Liu Z, Lin H, Tan X, et al. Clinical, immunological and virological characterization of COVID-19 patients that test re-positive for SARS-CoV-2 by RT-PCR. *EBioMedicine*. 2020. <https://doi.org/10.1016/j.ebiom.2020.102960>.
3. Agarwal V, Venkatakrishnan A, Puranik A, Kirkup C, Lopez-Marquez A, Challener DW, et al. Long-term SARS-CoV-2 RNA shedding and its temporal association to IgG seropositivity. *Cell Death Discov*. 2020;6(1):138. <https://doi.org/10.1038/s41420-020-00375-y>.
4. Centers for Disease Control and Prevention. Investigative Criteria for Suspected Cases of SARS-CoV-2 Reinfection (ICR) [Internet]. <https://www.cdc.gov/coronavirus/2019-ncov/php/invest-criteria.html> (accessed on 2 February 2022)
5. Yahav D, Yelin D, Eckerle I, Eberhardt CS, Wang J, Cao B, et al. Definitions for coronavirus disease 2019 reinfection, relapse and PCR re-positivity. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis*. 2021;27(3):315–8. <https://doi.org/10.1016/j.cmi.2020.11.028>.
6. Gao G, Zhu Z, Fan L, Ye S, Huang Z, Shi Q, et al. Absent immune response to SARS-CoV-2 in a 3-month recurrence of coronavirus disease 2019 (COVID-19) case. *Infection*. 2021;49(1):57–61. <https://doi.org/10.1007/s15010-020-01485-6>.
7. Gousseff M, Penot P, Gally L, Batisse D, Benech N, Bouiller K, et al. Clinical recurrences of COVID-19 symptoms after recovery: Viral relapse, reinfection or inflammatory rebound? *J Infect*. 2020;81(5):816–46. <https://doi.org/10.1016/j.jinf.2020.06.073>.
8. Ye G, Pan Z, Pan Y, Deng Q, Chen L, Li J, et al. Clinical characteristics of severe acute respiratory syndrome coronavirus 2 reactivation. *J Infect*. 2020;80(5):e14–7. <https://doi.org/10.1016/j.jinf.2020.03.001>.
9. Ravioli S, Ochsner H, Lindner G. Reactivation of COVID-19 pneumonia: a report of two cases. *J Infect*. 2020;81(2):e72–3. <https://doi.org/10.1016/j.jinf.2020.05.008>.
10. Li N, Wang X, Lv T. Prolonged SARS-CoV-2 RNA shedding: not a rare phenomenon. *J Med Virol*. 2020;92(11):2286–7. <https://doi.org/10.1002/jmv.25952>.
11. Wong ZSY, Zhou J, Zhang Q. Artificial intelligence for infectious disease big data analytics. *Infect Dis Health*. 2019;24(1):44–8. <https://doi.org/10.1016/j.idh.2018.10.002>.
12. Viceconti M, Hunter P, Hose R. Big data, big knowledge: big data for personalized healthcare. *IEEE J Biomed Health Inform*. 2015;19(4):1209–15. <https://doi.org/10.1109/JBHI.2015.2406883>.
13. Pollock M, Fernandes R, Becker L, et al. Chapter V: overviews of reviews. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (Eds) *Cochrane handbook for systematic reviews of interventions version 62* (updated February 2021) [Internet].; 2021. www.training.cochrane.org/handbook
14. Page MJ, McKenzie JE, Bossuyt P, Met Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. <https://doi.org/10.1136/bmj.n71>.
15. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358: j4008. <https://doi.org/10.1136/bmj.j4008>.
16. Abrokwa SK, Müller SA, Méndez-Brito A, Hanefeld J, El Bcheraoui C. Recurrent SARS-CoV-2 infections and their potential risk to public health—a systematic review. *PLoS ONE*. 2021;16(12): e0261221. <https://doi.org/10.1371/journal.pone.0261221>.
17. Chamblain M, Dawkins E, Lane J, Ghany R, Tamariz L, Palacio A, et al. A case of COVID-19 reinfection and systematic review of patterns of reinfection. *Infect Dis Clin Pract Baltim Md*. 2021;29(6):e409–11. <https://doi.org/10.1097/IPC.0000000000001055>.
18. Chivese T, Matiznadzo JT, Musa OAH, Hindy G, Furuya-Kanamori L, Islam N, et al. The prevalence of adaptive immunity to COVID-19 and reinfection after recovery—a comprehensive systematic review and meta-analysis. *Pathog Glob Health*. 2022. <https://doi.org/10.1080/20477724.2022.2029301>.
19. Kojima N, Shrestha NK, Klausner JD. A systematic review of the protective effect of prior SARS-CoV-2 infection on repeat infection. *Eval Health Prof*. 2021;44(4):327–32. <https://doi.org/10.1177/01632787211047932>.
20. Lo Muzio L, Ambosino M, Lo Muzio E, Quadri MFA. SARS-CoV-2 reinfection is a new challenge for the effectiveness of global vaccination campaign: a systematic review of cases reported in literature. *Int J Environ Res Public Health*. 2021. <https://doi.org/10.3390/ijerph182011001>.
21. Mao Y, Wang W, Ma J, Wu S, Sun F. Reinfection rates among patients previously infected by SARS-CoV-2: systematic review and meta-analysis. *Chin Med J (Engl)*. 2021;135(2):145–52. <https://doi.org/10.1097/CM9.0000000000001892>.
22. Massachi J, Donohue KC, Kelly JD. Severe acute respiratory syndrome coronavirus 2 reinfection cases corroborated by

- sequencing. *Am J Trop Med Hyg.* 2021;105(4):884–9. <https://doi.org/10.4269/ajtmh.21-0365>.
23. Sotoodeh Ghorbani S, Taherpour N, Bayat S, Ghajari H, Mohseni P, Hashemi Nazari SS. Epidemiologic characteristics of cases with reinfection, recurrence, and hospital readmission due to COVID-19: a systematic review and meta-analysis. *J Med Virol.* 2022;94(1):44–53. <https://doi.org/10.1002/jmv.27281>.
 24. Wang J, Kaperak C, Sato T, Sakuraba A. COVID-19 reinfection: a rapid systematic review of case reports and case series. *J Investig Med Off Publ Am Fed Clin Res.* 2021;69:1253–5. <https://doi.org/10.1136/jim-2021-001853>.
 25. Zhao J, Yuan Q, Wang H, Liu W, Liao X, Su Y, et al. Antibody responses to SARS-CoV-2 in patients with novel coronavirus disease 2019. *Clin Infect Dis Off Publ Infect Dis Soc Am.* 2020;71(16):2027–34. <https://doi.org/10.1093/cid/ciaa344>.
 26. To KK-W, Hung IF-N, Ip JD, Chu AW, Chan WM, Tam AR, et al. Coronavirus disease 2019 (COVID-19) re-infection by a phylogenetically distinct severe acute respiratory syndrome coronavirus 2 strain confirmed by whole genome sequencing. *Clin Infect Dis Off Publ Infect Dis Soc Am.* 2021;73(9):e2946–51. <https://doi.org/10.1093/cid/ciaa1275>.
 27. Van Elslande J, Vermeersch P, Vandervoort K, Wawina-Bokalanga T, Vanmechelen B, Wollants E, et al. Symptomatic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) reinfection by a phylogenetically distinct strain. *Clin Infect Dis Off Publ Infect Dis Soc Am.* 2021;73(2):354–6. <https://doi.org/10.1093/cid/ciaa1330>.
 28. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe.* 2021;29(3):463–476.e6. <https://doi.org/10.1016/j.chom.2021.02.003>.
 29. Fakhroo A, AlKhatib HA, Al Thani AA, Yassine HM. Reinfections in COVID-19 patients: impact of virus genetic variability and host immunity. *Vaccines (Basel).* 2021;9(10):1168. <https://doi.org/10.3390/vaccines9101168>.
 30. Mistry P, Barmania F, Mellet J, Peta K, Strydom A, Viljoen IM, et al. SARS-CoV-2 variants, vaccines, and host immunity. *Front Immunol.* 2022;12:809244. <https://doi.org/10.3389/fimmu.2021.809244>.
 31. Perera RA, Mok CK, Tsang OT, Lv H, Ko RL, Wu NC, et al. Serological assays for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull.* 2020. <https://doi.org/10.2807/1560-7917.ES.2020.25.16.2000421>.
 32. Carrillo J, Izquierdo-Useros N, Ávila-Nieto C, Pradenas E, Clotet B, Blanco J. Humoral immune responses and neutralizing antibodies against SARS-CoV-2; implications in pathogenesis and protective immunity. *Biochem Biophys Res Commun.* 2021;538:187–91. <https://doi.org/10.1016/j.bbrc.2020.10.108>.
 33. Sethuraman N, Jeremiah SS, Ryo A. Interpreting diagnostic tests for SARS-CoV-2. *JAMA.* 2020;323:2249–51. <https://doi.org/10.1001/jama.2020.8259>.
 34. Dhillon RA, Qamar MA, Gilani JA, Irfan O, Waqar U, Sajid MI, Mahmood SF. The mystery of COVID-19 reinfections: a global systematic review and meta-analysis. *Ann Med Surg.* 2021;72:103130. <https://doi.org/10.1016/j.amsu.2021.103130>.
 35. Piri SM, Edalatfar M, Shool S, Jalalian MN, Tavakolpour S. A systematic review on the recurrence of SARS-CoV-2 virus: frequency, risk factors, and possible explanations. *Infect Dis Lond Engl.* 2021;53:315–24. <https://doi.org/10.1080/23744235.2020.1871066>.
 36. Ledford H. Coronavirus reinfections: three questions scientists are asking. *Nature.* 2020;585:168–9. <https://doi.org/10.1038/d41586-020-02506-y>.
 37. Kellam P, Barclay W. The dynamics of humoral immune responses following SARS-CoV-2 infection and the potential for reinfection. *J Gen Virol.* 2020;101:791–7. <https://doi.org/10.1099/jgv.0.001439>.
 38. Garduño-Orbe B, Sánchez-Rebolledo JM, Cortés-Rafael M, García-Jiménez Y, Pérez-Ortiz M, Mendiola-Pastrana IR, et al. SARS-CoV-2 reinfection among healthcare workers in Mexico: case report and literature review. *Med Kaunas Lith.* 2021. <https://doi.org/10.3390/medicina57050442>.
 39. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst.* 2014;2:3. <https://doi.org/10.1186/2047-2501-2-3>.
 40. Alonso SG, de la Torre DI, Rodrigues JJPC, Hamrioui S, López-Coronado M. A systematic review of techniques and sources of big data in the healthcare sector. *J Med Syst.* 2017;41(11):183. <https://doi.org/10.1007/s10916-017-0832-2>.
 41. Kantarjian H, Yu PP. Artificial intelligence, big data, and cancer. *JAMA Oncol.* 2015;1(5):573–4. <https://doi.org/10.1001/jamaoncol.2015.1203>.
 42. SeyedAlinaghi S, Oliaei S, Kianzad S, Afsahi AM, Mohsseni-Pour M, Barzegary A, et al. Reinfection risk of novel coronavirus (COVID-19): a systematic review of current evidence. *World J Virol.* 2020;9(5):79–90. <https://doi.org/10.5501/wjv.v9.i5.79>.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.