Revista ARACÊ
ISSN: 2358-2472

IMPORTANCE OF QUALITY AND UNIFORM DEFINITIONS FOR COVID-19 VACCINATION DATA

https://doi.org/10.56238/arev6n4-026

Submitted on: 03/11/2024 **Publication date:** 03/12/2024

Cristiano Soares da Silva Dell'Antonio[1], Larissa Soares Dell'Antonio[2], Matheus Rocha Short[3] and Ana Luiza Bierrenbach[4]

ABSTRACT

Objective: This study compares typical metrics using different data processing methods in a Covid-19 vaccination database. Methods: Data were extracted from the "Vaccine and Trust" system in March 2022, focusing on the state of Espírito Santo due to inconsistencies in the national database following a hacker attack (2021). Results: The analyses revealed 19,221 duplicate records with a more rigorous methodology. Espírito Santo reached 80% coverage for the second dose, but booster and additional doses were lower. Methodological variations led to metric discrepancies, suggesting the need for clear disclosure of the methodology. The study also noted a high proportion of incorrect vaccine sequences, providing possible explanations. Conclusion: We emphasize the need to ensure the collection and processing of solid data and to standardize the analyses. The authors warn of potential obstacles without these measures, influencing the achievement of established goals.

Keywords: Data Quality. Vaccine. COVID-19. Public health. Electronic Data Processing.

[1] Syrian-Lebanese Hospital, Syrian-Lebanese Institute of Teaching and Research, São Paulo, SP, Brazil.
E-mail: cristianosss@outlook.com
ORCID: 0000-0001-6568-1300
[2] State Department of Health of the State of Espírito Santo, ES, Brazil.
E-mail: lissadellantonio@gmail.com
ORCID: 0000-0001-5084-2829
[3] State Department of Health of the State of Espírito Santo, ES, Brazil.
E-mail: matheusrochacurto@gmail.com
ORCID: 0009-0001-6734-6765
[4] Syrian-Lebanese Hospital, Syrian-Lebanese Institute of Teaching and Research, São Paulo, SP, Brazil.
E-mail: albierrenbach@yahoo.com.br
ORCID: 0000-0002-6837-0636

Revista ARACÊ
ISSN: 2358-2472

INTRODUCTION

In the face of the rapid spread of the SARS-Cov-2 virus and its transformation into a global pandemic declared by the World Health Organization (WHO) in March 2020, the disease caused by the novel coronavirus (COVID-19) has become a challenge on a global scale1,2. By the 29th epidemiological week of 2022, the alarming numbers of the pandemic included more than 6.3 million deaths worldwide and more than 673 thousand in Brazil3,4.

This urgency has led to an unprecedented effort to develop and distribute vaccines in record time. The United Kingdom was the first country to start vaccination against COVID-19 outside of clinical trials, in December 20205. Since then, global vaccination campaigns have played a crucial role in reducing virus transmission and mitigating public health impacts, relieving pressure on health systems in many countries6,7. In Brazil, vaccination began in January 2021 with Coronavac, followed by other vaccines such as AstraZeneca, Pfizer, and Janssen throughout the year8,9.

To keep track of the scale and distribution of vaccines, it was essential to develop reporting and data recording systems. The National Immunization Program Information System (SI-PNI), coordinated by the Ministry of Health, was adapted to record the application of vaccines against COVID-19. In addition, several Brazilian states and municipalities have developed their own registration and registration systems, complementing the national effort9,10. However, the lack of standardization in the collection and processing of these data can result in inconsistencies and challenges in the analysis and comparison of metrics and statistics10,11.

By the 29th epidemiological week of 2022, a total of 11.7 billion vaccine doses had been administered worldwide4. In Brazil, with a population of 212.6 million, 454,264,544 doses were administered in the same period. In the Brazilian state of Espírito Santo, with a population of 4.1 million inhabitants, after several waves of contamination that led to the death of 14 thousand people[12] and a very high occupancy of the more than 1,600 beds available13, official figures indicate that in the same period 8,448,378 doses were administered 14,15.

In the Brazilian context, the Ministry of Health established a centralized database to record COVID-19 vaccination information. However, states have autonomy to develop their own registration systems, as long as they are interoperable with the national system. An example is "Vaccine and Trust", an official platform in Espírito Santo launched in July 2021. This platform not only allows citizens to schedule vaccination, but also provides detailed

Revista ARACÊ
ISSN: 2358-2472

information about the vaccination campaign, including data on dose availability, stock, expired doses, and sociodemographic analyses16.

While there is still much to learn about the virus and its interaction with populations, the availability of data on transmission, case progression, vaccine efficacy, and distribution patterns allows public health authorities to take the necessary steps to manage and contain the spread of the virus17. In particular, the analysis of vaccination data can provide a comprehensive understanding of the performance of vaccines in different contexts and contribute to reducing inequalities in access to vaccines. However, to be useful, data on vaccination against Covid-19 need to be accurate, comparable and up-to-date5,18. The present work aims to compare the values of usual metrics with the application of different data processing methodologies in a Covid-19 vaccination database in the state of Espírito Santo, located in the Southeast of Brazil. The intention here is to emphasize that different ways of processing data actually generate different metrics, making it difficult to compare the results.

OBJECTIVE

To compare the values of usual metrics with the application of different data processing methodologies in a Covid-19 vaccination database in the state of Espírito Santo, located in the Southeast of Brazil. The intention here is to emphasize that different ways of processing data actually generate different metrics, making it difficult to compare the results.

METHODS

This is an observational epidemiological study using secondary data from the "Vaccine and Trust" vaccination database, extracted in March 2022. The choice to use this state database and not the national one was due to the occurrence of inconsistency between the two during the period of system instability caused by a hacker attack on the national database in December 202114.

"Vaccine and Trust" has a standardized data collection instrument, in online format, which is completed by a professional from the immunization team responsible for applying the vaccines. The extracted database contains vaccination records with an application date of February 28, 2022. For the analysis, the following variables were considered: subject's name, date of birth, national individual registration (CPF), national health card (CNS), date

Revista ARACÊ
ISSN: 2358-2472

of application, date of entry into the system, name of the vaccine (AstraZeneca, Coronavac, Janssen and Pfizer), vaccine batch, dose (1=first, 2=second or single dose, R=booster and A=additional), health unit name, registration number of the health unit (CNES), municipality of application of the vaccine, municipality of residence and neighborhood of residence. Personally identifiable information was used only for deterministic linking.

Records of children under 18 years of age and records that presented another vaccine in addition to the four vaccines distributed to adults in the country were excluded from the study. The application of the single dose of the Janssen vaccine was considered as the application of the second dose to facilitate the comparability of the data.

In the surveillance service's data management routine, records with coinciding numbers in the two numerical identification variables (CPF and CNS) are considered to belong to the same individual. However, taking into account the existence of missing values and the possibility of typing errors in both variables, the identification of records belonging to the same individual is not perfect, but can be improved with simple deterministic matching resources. In addition, some types of duplicates and errors in the registered dates can also be corrected through simple procedures.

Starting from the "original" database, which is a direct product of the routine data processing in use in the surveillance service, the following data processing procedures were carried out, which gave rise to the "corrected" database:

1) Correction of dates of birth:

The age of the subjects was calculated by subtracting the date of birth from the date of application of the vaccine. Records with negative ages or over 125 years old had their dates of birth manually verified and eventually the year of birth was corrected.

2) Deterministic linkage of records:

Before starting the linkage procedures, we cut all relevant text variables from consecutive starting, ending, and inner whitespace. Records belonging to the same subject were those that contained the following combinations of variables with the same information or value: (1) same CPF/CNS; (2) same CPF, name and date of birth of the subject; (3) same CNS, subject name and date of birth; (4) name of the same subject, date of birth, neighborhood and municipality of residence; (5) name of the same subject, date of birth, and name of the health

Revista ARACÊ
ISSN: 2358-2472

facility where the vaccine was administered, and (6) name of the same subject, date of birth, and national registration number of the health facility where the vaccine was administered. In all of these rules, variables could not contain missing values.

3) Exclusion of duplicates:

Duplicate records were those that contained identical information about a selection of variables. The selections were applied in the following order, from the most restrictive to the least restrictive: (1) same subject (as identified by the linkage), vaccine, batch, date of application, and dose applied; (2) same subject, vaccine, date of application and dose applied; (3) same subject, vaccine, date of application; (4) same subject, vaccine, date of application and dose applied; (5) same subject, vaccine and dose applied and date of application up to one day apart, and (6) same subject, vaccine and dose applied and date of application up to seven days apart. The goal of having more and less restrictive selections was to be able to count how many duplicates were identified and removed with each of them.

4) Correction of some vaccine application dates:

Another improvement introduced in the database was the correction of the date of application of the vaccine. For each subject, the difference between consecutive application dates was calculated. When this difference was 365 days or more, it was more likely that the year of application of one of the records was incorrect. For these records, the application dates were corrected with an addition of 365 days for those whose year was 2021 and who had a record of insertion date in the system in the year 2022.

To allow the comparison of the original and corrected databases, the following metrics were calculated:

1) Total number of doses applied;
2) Number of people vaccinated;
3) Vaccinated rate (per 1 population) in relation to the total adult population of the State of Espírito Santo;
4) Number of each of the doses applied (1st dose, 2nd dose, Booster Dose and Additional Dose).

This metric was calculated using two dose determination methods:

Revista ARACÊ
ISSN: 2358-2472

a) Maximum dose recorded (Dose recorded): considers all records of each subject and uses the information of the highest dose administered, as presented in the corresponding variable;

b) Maximum Calculated Dose (Calculate Dose): sums the number of records of each subject (1 record = 1 dose, 2 records = 2 doses or single dose of Janssen, 3 records = 3 doses, 4 records or more = 4 or more doses);

5) Proportion of each of the doses applied in relation to the total adult population of the State of Espírito Santo;

6) Interval of days between the 1st dose and the 2nd dose, calculated separately for Coronavac (recommended interval of 14 to 28 days), AstraZeneca (recommended interval of 84 days) and Pfizer (recommended interval of 56 days). Median and interquartile range (IQR)[9];

7) Interval of days between the 2nd dose and the booster dose, calculated separately for Coronavac/AstraZeneca/Pfizer (recommended interval of 120 days) and for Janssen (recommended interval of 60 days). Median and IQR9.

8) Frequency of doses of the complete vaccination schedule (1st, 2nd, booster and additional dose or 1st, 2nd, booster dose or single dose of the Janssen vaccine, booster dose and additional dose or single dose and booster dose), incomplete vaccination schedule (1st dose or 1st and 2nd dose or single dose) and inadequate vaccination schedule (vaccination schedule with sequences that do not start until the 1st dose or single dose);

9) Frequency of the vaccination schedule according to the sequence of doses.

In order to better understand the difference between the recorded dose and the calculated dose, we present below the registration data of two subjects (Table 1). In these examples, we can observe that subject A had all his vaccination records recorded in the correct order. The maximum calculated dose is equal to the maximum recorded dose which is four. For subject B, the maximum calculated dose is equal to four, as there are four records in his name, but the maximum dose recorded is two.

Revista ARACÊ
ISSN: 2358-2472

Table 1. Vaccine data from two records of the study to show the difference between the calculated and recorded maximum dose metrics by date of entry.

Vaccine	Batch Number	Dose administered	Application date	Date entered
		Subject A		
AstraZeneca	ABW4735	1st Dose	15/05/2021	26/05/2021
AstraZeneca	216VCD195Z	2nd Dose	29/07/2021	29/07/2021
Pfizer	31045BD	Booster dose	02/10/2021	02/10/2021
Pfizer	FJ8766	Additional dose	18/02/2022	18/02/2022
Calculated dose: 04; Registered dose: 04; Immunization schedule: 1.2.R.A.				
		Subject B		
Coronavac	202010018	1st Dose	25/01/2021	08/02/2021
AstraZeneca	4120Z005	1st Dose	26/01/2021	10/02/2021
Coronavac	210013	2nd Dose	18/02/2021	19/02/2021
AstraZeneca	213VCD004ZVA	2nd Dose	21/04/2021	23/04/2021
Calculated dose: 04; Dose recorded: 02; Immunization schedule: 1.1.2.2.				

For the population data, population projections from the Brazilian Institute of Geography and Statistics (IBGE) were used, available on the website of the SUS Department of Informatics and used in the Vaccine and Trust system.

Database management and descriptive analysis were performed using the Stata-17 software (Statacorp, College Station, Texas, USA).

This study was authorized by the Espírito Santo Institute of Teaching, Research and Innovation in Health of the State of Espírito Santo and approved by the Research Ethics Committee of the Institute of Teaching and Research of the Syrian-Lebanese Hospital (CAEE No. 45986821.9.0000.5461 of December 2021), with exemption from the Consent Form. All individually identifiable information was kept confidential.

RESULTS

Figure 1 shows the procedures for processing the data carried out in the original database that gave rise to the corrected database. The numbers of records and people in each database are also presented, as well as the number of records that were deleted or modified in each of these processes, totaling 19,221 duplicates.

Revista ARACÊ
ISSN: 2358-2472

Figure 1. Diagram containing the database construction routines and their respective processes.

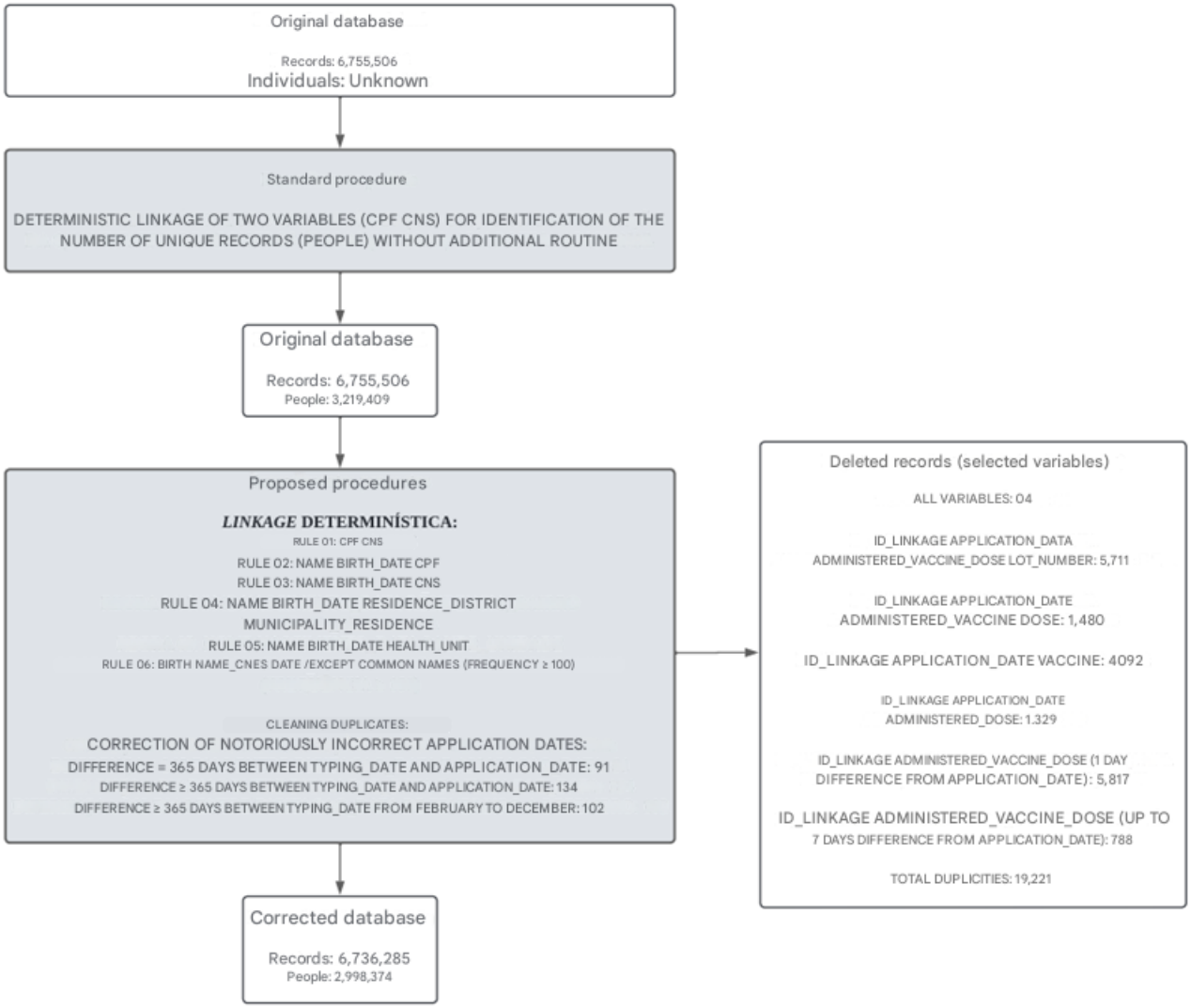

Table 2 shows the significant decline in the proportion of the adult population vaccinated from the first (91.3%) to the second (87%), to the booster dose (40%) and then to the additional dose (2.5%), using the corrected database and the "as recorded" methodology. Table 2 also compares the metrics using the surveillance and corrected databases. In general, the number of doses administered in the original (uncorrected) database is higher than in the corrected database, which is expected since duplicates have been excluded. Similarly, the number of "as recorded" doses is higher on the (uncorrected) surveillance basis. An exception is the number of 2nd doses and booster doses, which is lower on the surveillance basis than on the corrected basis, using the "as calculated" methodology. While the rate is 1.05 vaccines per adult based on surveillance, it drops to 0.98 in the corrected database.

Revista ARACÊ
ISSN: 2358-2472

The interval of days between the 1st and 2nd dose varied greatly in the comparison of Coronavac with the others, as was to be expected given the official recommendations of the manufacturers. The interval was longer between the 2nd dose and the booster, when the 2nd dose vaccine was Janssen's. There was not much variation in the intervals using data from the surveillance (original) or from the corrected databases.

Table 2. Distribution of the most frequent vaccination schedule according to the sequence of doses recorded at the time of application in the original database.

Metric		Original	Fixed
		N (% pop)	N (% pop)
Number of doses administered		6.755.506	6.736.285
Number of individuals	(Rate per population)	3.219.409 (1.05)*	2.998.374 (0.98)*
1st dose	logged	2.783.988 (91.4)*	2.780.148 (91.3)*
	deliberate	3.219.409 (100.0)*	2.998.374 (98.4)*
2nd dose / Individual dose	Registered	2.650.878 (87.1)*	2.650.040 (87)*
	Calculated	2.556.088 (83.9)*	2.602.758 (85.4)*
Booster dose	Registered	1.219.519 (40.0)*	1.218.645 (40.0)*
	Calculated	969.353 (31.8)*	1.131.496 (37.1)*
Additional dose	Registered	78.033 (2.5)*	77.516 (2.5)*
	Calculated	9.791 (0.3)*	3.613 (0.1)*
Interval in days between 1st – 2nd dose	C	34	34
IQR		(28-47)	(28-48)
	AZ	81	82
		(75-88)	(75-88)
	PF	78	78
		(71-85)	(71-85)
Interval of days between 2nd/Single Dose – Booster Dose	C/AZ/PF	149	153
		(132-165)	(133-168)
	J	175	175
IQR		(167-194)	(167-194)

Legend: C: Coronavac vaccine; AZ: AstraZeneca vaccine; PF: Pfizer vaccine; J: Janssen vaccine; IQQ: Interquartile range. * Population ≥ 18 years of age: 3,044,546).

When analyzing the sequence of doses recorded in the original database, it was observed that 29.45% of the records had a complete vaccination schedule (1.2.R.A or 1.2.R or U.R or U.R.A), 56.7% had incomplete vaccination (U or 1 or 1.2) and 13.85% had an inadequate vaccination schedule (other vaccine records). Table 3 presents in detail the inconsistencies in the sequence of immunization and their frequency in the original database. It can be observed that there are several subjects with a lack of adequate sequence of vaccine application, as well as those with the presence of repeated doses, in addition to records in which the single dose presented would be a booster dose (R) or an additional dose (A).

Revista ARACÊ
ISSN: 2358-2472

Table 3. Distribution of the vaccination schedule according to the sequence of doses recorded at the time of application in the original database.

Immunization sequence	Frequency of each sequence	Group Combination Frequency	Frequency (%)
1.2	1.365.164	1.365.164	42.40
1.2.R	892.234	892.234	27.71
1	413.362	413.362	12.84
2.R	115.113	115.113	3.58
R	114.480	114.480	3.56
2	84.189	84.189	2.62
1.2.A	57.062	57.062	1.77
U.R	55.120	55.120	1.71
U	46.961	46.961	1.46
1.R	29.946	29.946	0.93
2.A	11.050	11.050	0.34
The	4.329	4.329	0.13
1.1.2.R	3.773	3.773	0.12
1.1.2	3.559	3.559	0.11
1.1	3.388	3.388	0.11
2.1	1.726	1.726	0.05
1.2.U	1.701	1.701	0.05
1.2.A.R	1.562	1.562	0.05
U.2	1.562	1.562	0.05
1.2.2.R	1.364	1.364	0.04
1.A	1.357	1.357	0.04
1.1.R	1.068	1.068	0.03
1.2.2 / U.1 / 2.1.R / 1.2.R.A / 2.2.R / 1.1.2.2.R / 1.1.2.2 / 2.2 / 1.1.2.A / U.A / 2.A.R / 1.2.R.R / 1.2.2.A / R.A / 2.R.A / A.R / 1.U / U.2.R / 1.R.2 / U.U	100 – 1.000	7.499	0.23
2.U / R.R / R.1.2 / 1.1.1.2 / 1.2.1 / 1.1.1.2.R / U.1.2 / 2.1.A /			
2.2.A / 1.A.R / 1.R.A / 1.U.2 / 1.2.1.R / 1.2.R.U / 2.1.2.R /			
U.U.R/1.1.1/1.1.2.2.A/1.U.R/2.R.R/U.1.R/2.1.2/R.2/			
2.2.1 / 1.A.2 / 1.U.2.R / 1.1.1.R / 1.1.A / R.U / 1.1.2.R.A /	10 - 100	1.578	0.05
1.2.1.2.R / U.1.2.R / U.A.R / 1.2.2.2.R / A.1.2 / U.1.2.A / 1.1.1.2.2.R / 1.1.2.U / 1.A.2.R / R.1 / U.U.2 / 1.1.2.A.R / 1.2.2.2			
/ 1.R.U / U.R.A			
/ 1.1.1.1.2.R / 1.1.2.2.2.2 / 1.1.U / 1.2.1.A / 1.2.2.A.R / 1.2.U.R / 2.1.A.R / 2.1.R.A / 2.2.2.R / 1.1.1.2.A / 2.R.1 / U.2.A.R / U.R.1 /			
.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.			
1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1..1.			
1.2.2.R / 1.1.1.2.2.R / 1.1.1.R.R / 1.1.2.1.2.2.R / 1.1.2.2.2.R / 1.1.2.2.2.R / 1.1.2.2.A.R / 1.1.2.R.U / 1.1.A.2 / 1.1.R.2 / 1.2.1.2.2.R / 1.2.A			
.R.2.2.R / 1.2.2.R.R / 1.2.2.R.R / 1.2.2.R.R / 1.2.2.R.R / 1.2.2.R.R / 1.2.2.R.R / 1.2.2.R.R / 1.2.2.R.R / 1.2.2.R.R / 1.2.2.R.R / 1.2.2.R.R			
R.2.2.R / 1.2.2.R.2.2.R / 1.2.2.R.2.2.R / 1.2.2.R.2.2.R / 1.2.2.R.2.2.R / 1.2.2.R.2.2.R / 1.2.2.R.2.2.R / 1.2.2.R.2.2.R / 1.2.2.R.2.2.R / 1.			
A.A.A / A.R.A / R.2.1.R / R.2.A / R.2.R / U.1.1 / U.1.1.2 / U.1.2.2 /			
Total		3.219.409	100

Legend: 1: 1st dose; 2: 2nd dose; A: Booster Dose; A: Additional dose.

Revista ARACÊ
ISSN: 2358-2472

DISCUSSION

The present study analyzed the inconsistencies in the records of the "Vaccine and Trust" database in the state of Espírito Santo through two methodologies for counting doses, using routine data collected during the pandemic process caused by the new Coronavirus. Considering the results obtained with the most rigorous methodology, it was possible to identify 19,221 duplicate records, which corresponds to 0.2% of the entire database. The number of doses inoculated for each adult was close to only one, although four doses were recommended in the period studied. In an analysis of global public databases entitled "Our World in Data COVID-19" carried out in 2021, the various limitations present in the analysis of data from different sources of information were presented, with the main point being the lack of standardization in the methodology of data collection and management, requiring the implementation of several strategies to ensure the quality of the data presented, through manual and automated validation techniques[5].

According to an analysis carried out by the organization Open Knowledge – Brazil in partnership with other civil society organizations, the Ministry of Health's national vaccination bank had more than 27 thousand duplicate records between the first and second dose, when using simple verification methods (identification of records belonging to the same person)[17]. On the other hand, in a study funded by the federal government and carried out by a group of universities, the number of people vaccinated with at least one dose in the country was 13.5% lower in the database without duplicates than in the original database, a reduction of 7.6 million people[19]. It is natural that studies using different methodologies find different proportions of inconsistencies and duplicate records. In our study, the number of duplicates found also increased when, in addition to the routine methodology, slightly more sophisticated linkage procedures were used, but still feasible to be performed by the responsible teams. We were not surprised that the proportion of duplicates was only 0.2% in the state of Espírito Santo, since there are strong surveillance teams in its various municipalities and most of the inconsistencies are verified and removed at the source, that is, in the municipalities, not reaching the state level[20]. In the rankings on data transparency, the State always tends to occupy the first positions[21]. What matters here is the understanding that the data management methodology must be standardized and presented clearly[10]. It is noteworthy that the inadequacy of the records or the imprecision of the data can directly influence their interpretation, directly influencing the planning and implementation of strategies22.

Revista ARACÊ
ISSN: 2358-2472

Recent studies have shown a high rate of global acceptance of the additional dose of the vaccine against Covid-19[23], which diverges from the findings of the present study. We observed a significant reduction in the immunization rate, especially in relation to booster doses and the additional dose, from more than 90% in the first dose to approximately 2.5% of the population immunized with the additional dose. Factors such as adverse reactions and discomfort experienced in previous doses of the COVID-19 vaccine and concerns about adverse reactions and previous COVID-19 infections are presented in the literature as the main predictors for the decline in the uptake of additional doses [23,24]. It is important to highlight another factor that may have contributed to the drop in doses administered, which highlights the lack of support from the then federal government associated with the large number of Fake News, which considerably harmed the acceptability of the vaccine.

Regarding the interval between doses, it was observed that the median interval exceeds the recommended period. In a study describing the vaccination campaign against SARS-CoV-2 in Brazil, scientific evidence is presented related to the change in the effectiveness of the immunizer when the interval is not respected, thus bringing a greater probability of maintaining high viral circulation and consequently influencing the emergence of new variants. The same text presents possible causes for these delays during the recent pandemic, in which the inefficiency and incompetence of the Brazilian government in the purchase of vaccines certainly played a fundamental role25. As presented by the Pan American Health Association, there is still a need for better knowledge about the duration of vaccine immunity for Covid-19, but the recommendation that the booster dose be administered 6 months after the two primary doses of the immunizer is already well established, with the aim of avoiding the decline of protection over time26.

The study of the vaccination sequence is essential to understand the coverage and effectiveness of vaccination programs. In this context, the analysis of the sequence of doses recorded in a given population can provide important information about the quality of vaccination services offered. When analyzing the sequence of doses recorded, it was found that 86% of patients have logical vaccination sequences, although for some of them the doses were incomplete. But the complement of this observation is that for 14% of the patients the doses recorded contained invalid sequences, which do not present the appropriate pattern of vaccination sequence. When we observe the sequences of inadequate vaccination records of this large number of patients, we must keep in mind that

Revista ARACÊ
ISSN: 2358-2472

they may, on the one hand, be artifacts resulting from the malfunction of the information system. On the other hand, they may be real representations of inadequate vaccine sequences received by each patient.

In several of the patients with invalid sequences, we observed the presence of more than one dose of vaccination, for example, the sequence 1.1.2.R reported in 3,773 people. Although it is possible that this is a system error, a double entry of the first dose, it is also possible that at least a part of these people have been vaccinated twice with the first dose. In other sequences, such as 1.2.A.R which was reported in 1562 people, the most likely explanation seems to be an error in the date of application or in the booster or additional dose. There must be a combination of factors that contribute to these incorrect sequences.

First, when the statewide "Vaccine and Trust" database was implemented in July 2021, vaccine data that had been collected in the national database since the start of the campaign in mid-January of that year had to be imported. In this import process, deduplication was carried out, but it is possible that records of the same person remained that were not recognized as such by the linkage methodology then employed, but that were recognized by our methodology, with greater sensitivity without losing specificity. While we consider our records linkage methodology to be robust, probably with greater sensitivity than that performed by the surveillance system, and without loss of specificity, we recognize that no record linkage is perfect and that there may have been inappropriate peer linkages and failure to link proper peers.

Secondly, another possibility of system error concerns the absence of periodic linking of records between state and national databases. Individuals with vaccinations in transit may also be responsible for the presence of irregular sequences, probably explaining some of the many sequences that have a missing dose, such as 1.2.A where the booster dose (R) is missing, which appears in more than 57 thousand people. This may also explain the higher number of 2nd dose and booster doses observed in the "corrected" database compared to the "original base" identified in our analysis.

Third, we should consider the possibility that people have been vaccinated more than once with the same doses. It is important to mention that the professional responsible for administering the vaccine and recording this data often did not have access to an electronic mechanism for data collection, much less online access to the database, and therefore could not know about the presence of a previous dose. A patient determined to receive more than one dose without the recommended time interval, thinking that two

Revista ARACÊ
ISSN: 2358-2472

doses would confer more protection, or seeking the preferred vaccine brand, could unfortunately bypass the system and receive such doses27,28. The person responsible for entering the vaccination data, in an electronic data collection mechanism when available in their service or on paper, was often the health professional who administered the vaccine and, other times, possibly an employee with little experience.

Fourthly, with the large queues that formed especially in the first months since the arrival of the vaccine in the country, there was a lot of pressure for vaccination and data entry to be done as soon as possible, which increases the likelihood of errors. Another factor to be considered is the training of those responsible for the data. In Brazil, there is still little training of professionals who work in surveillance and vaccination services. Without understanding exactly the reason for data collection, the need for data quality, and without receiving feedback on their work, the professional who enters the data, in addition to being poorly trained, is also poorly motivated to perform the work properly27,28. In any case, it is difficult to determine the individual contribution of each of these possible alternatives to explain the presence of the high proportion of people with incorrect vaccination sequences.

Vaccination is one of the most effective measures to prevent infectious diseases and its importance is recognized by the world scientific community. However, the lack of adherence to vaccination recommendations can generate public health problems, such as outbreaks of diseases that have already been controlled29. Thus, knowledge about vaccination coverage and compliance with recommendations is essential for managers to be able to identify barriers and structure actions that help to change them, given the need for high vaccination coverage rates, which are necessary to break the chain of transmission of the disease30.

We emphasize that our objective is not to point out errors or inaccuracies in the information system. Our goal is to recognize the need for standardization of data collection and management practices. To monitor the vaccination process within a country, compare regions or countries to ensure more equitable access and better control of the pandemic, it is essential to collect solid information, work to reduce errors, and standardize the treatment of data treatment and calculation of indicators, both nationally and internationally. In addition, there are few studies that analyze vaccination data related to Covid-19 with this approach in the literature.

Revista ARACÊ
ISSN: 2358-2472

CONCLUSIONS

As can be seen from the data presented, it is possible to understand that the vaccination campaign against COVID-19 in the State of Espírito Santo was quite efficient in the sense that it reached a vaccination coverage of about 80% with the second dose (or the single dose of Jansen). On the other hand, coverage for booster and additional doses was much lower, which is a very worrying fact. In addition, we show that the use of different methodologies to calculate how many doses everyone received (corrected database vs. original database) actually causes differences in metrics and suggest that the methodology be at least spelled out in state and national reports, or in publications that make an overview of various countries. Finally, we highlight the high proportion (more than 14%) of individuals with incorrect vaccination sequences in their records and try to list explanations for this. With this, we want to draw attention to the need not only to ensure solidity in the collection and processing of data, but also to standardize certain analyses and to be open-minded in the interpretation of the results. Without all this, the data that should serve as a guide for future actions could end up becoming an obstacle to meeting the established goals.

Revista ARACÊ
ISSN: 2358-2472

REFERENCES

1. Strabelli, T. M. V., & Uip, D. E. (2020). COVID-19 e o coração. Arq Bras Cardiol, 114(4), 598–600. https://doi.org/10.36660/abc.20200209

2. Fiocruz. (2021). O que é uma pandemia. Fiocruz. Disponível em: https://www.bio.fiocruz.br/index.php/br/noticias/1763-o-que-e-uma-pandemia. Acesso em: 18 de outubro de 2023.

3. Brasil. (2021). Painel Coronavírus Brasil. Ministério da Saúde. Disponível em: https://covid.saude.gov.br. Acesso em: 28 de agosto de 2023.

4. Johns Hopkins. (2021). Brasil - COVID-19 Overview. Johns Hopkins. Disponível em: https://coronavirus.jhu.edu/region/Brasil. Acesso em: 18 de outubro de 2023.

5. Mathieu, E., Ritchie, H., Ortiz-Ospina, E., et al. (2021). A global database of COVID-19 vaccinations. Nat Hum Behav, 5, 947–953. https://doi.org/10.1038/s41562-021-01122-8

6. Amaral, F., Casaca, W., Oishi, C. M., & Cuminato, J. A. (2021). Simulating immunization campaigns and vaccine protection against COVID-19 pandemic in Brazil. IEEE Access, 9, 126011–126022. https://doi.org/10.1109/ACCESS.2021.3112036

7. Coccia, M. (2022). COVID-19 pandemic over 2020 (with lockdowns) and 2021 (with vaccinations): similar effects for seasonality and environmental factors. Environ Res, 208, 112711. https://doi.org/10.1016/J.ENVRES.2022.112711

8. Stevanim, L. F. (2020). Uma vacina para a humanidade: Da expectativa à realidade, os esforços para se chegar a uma vacina contra Covid-19 acessível à população. RADIS: Comunicação e Saúde, 12–21. Disponível em: https://www.arca.fiocruz.br/handle/icict/43683

9. Brasil. (2022). Plano Nacional de operacionalização da vacinação contra a COVID-19 (13. ed.). Brasília. Disponível em: https://sbim.org.br/images/files/notas-tecnicas/pno-13edicao-230522.pdf. Acesso em: 20 de outubro de 2023.

10. ASPE. (2021). Quality of Federal COVID-19 Data. ASPE. Disponível em: https://www.aspe.hhs.gov/sites/default/files/documents/8467add5edfbaa06f70ade5e4f829655/covid-data-quality-ib.pdf?_ga=2.63299951.1944804403.1666141562-1147266797.1666141562. Acesso em: 18 de outubro de 2023.

11. Brasil. (2022). Campanha Nacional de Vacinação contra Covid-19 - Conjuntos de dados - OPENDATASUS. Ministério da Saúde. Disponível em: https://opendatasus.saude.gov.br/dataset/covid-19-vacinacao. Acesso em: 18 de outubro de 2023.

12. Dell'Antonio, L. S., Leite, F. M. C., Dell'Antonio, C. S. da S., et al. (2022). COVID-19 mortality in public hospitals in a Brazilian state: An analysis of the three waves of the

Revista ARACÊ
ISSN: 2358-2472

pandemic. International Journal of Environmental Research and Public Health, 19, 14077. https://doi.org/10.3390/ijerph192114077

13. Garbin, J. R. T., Leite, F. M. C., Lopes-Júnior, L. C., et al. (2022). Analysis of survival of patients hospitalized with COVID-19 in Espírito Santo, Brazil. International Journal of Environmental Research and Public Health, 19, 8709. https://doi.org/10.3390/ijerph19148709

14. Brasil. (2022). Reunião da Câmara Técnica Assessora de Imunização Covid-19 - 18 de fevereiro de 2022. Brasília. Disponível em: https://www.gov.br/saude/pt-br/composicao/secovid/ctai-covid-19/atas-e-reunioes/2022/atas/ata-da-reuniao-ctai-18-02-2022.pdf. Acesso em: 18 de outubro de 2023.

15. Espírito Santo. (2021). Vacina e Confia ES. Governo do Estado do Espírito Santo. Disponível em: https://vacinaeconfia.es.gov.br/cidadao/. Acesso em: 19 de outubro de 2023.

16. Espírito Santo. (2021). Governo ES - Vacina e Confia ES: Governo lança nova plataforma voltada à imunização contra a COVID-19. Governo do Estado do Espírito Santo. Disponível em: https://www.es.gov.br/Noticia/vacina-e-confia-es-governo-lanca-nova-plataforma-voltada-a-imunizacao-contra-a-covid-19. Acesso em: 1 de fevereiro de 2023.

17. OKBR. (2021). Nota Técnica - Transparência da Vacinação. Uma análise de organizações da sociedade civil. Disponível em: https://www.ok.org.br/wp-content/uploads/2021/03/Nota_Tecnica_Vacinacao_04032021.pdf. Acesso em: 18 de outubro de 2023.

18. United Kingdom. (2022). COVID-19 vaccine weekly surveillance reports. Government of the United Kingdom. Disponível em: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1063023/Vaccine-surveillance-report-week-12.pdf. Acesso em: 24 de março de 2023.

19. Lopes, M. R. (2021, junho 29). Estudo encontra 7,6 milhões de erros em lista de vacinados no país. R7 Brasil. Disponível em: https://noticias.r7.com/brasil/estudo-encontra-76-milhoes-de-erros-em-lista-de-vacinados-no-pais-29062022. Acesso em: 27 de outubro de 2023.

20. Dell, L., Pereira, A., Soares da Silva, C., et al. (2021). Vigilância epidemiológica estadual no enfrentamento da pandemia pela COVID-19 no Brasil: Um relato de experiência. Escola Anna Nery, 25. https://doi.org/10.1590/2177-9465-EAN-2021-0119

21. OKBR. (2021). Transparência Covid-19. OKBR. Disponível em: https://transparenciacovid19.ok.org.br. Acesso em: 20 de outubro de 2023.

22. Lassi, Z. S., Naseem, R., Salam, R. A., et al. (2021). The impact of the COVID-19 pandemic on immunization campaigns and programs: A systematic review. J Environ Res Public Health, 3, 988. https://doi.org/10.3390/ijerph18030988

Revista ARACÊ
ISSN: 2358-2472

23. Abdelmoneim, S. A., Sallam, M., Hafez, D. M., et al. (2022). COVID-19 vaccine booster dose acceptance: Systematic review and meta-analysis. Trop Med Infect Dis, 7, 298. https://doi.org/10.3390/tropicalmed7100298

24. Galanis, P., Vraka, I., Katsiroumpa, A., et al. (2022). First COVID-19 booster dose in the general population: A systematic review and meta-analysis of willingness and its predictors. Vaccines, 10, 1097. https://doi.org/10.3390/vaccines10071097

25. Maciel, E., Fernandez, M., Calife, K., et al. (2022). A campanha de vacinação contra o SARS-CoV-2 no Brasil e a invisibilidade das evidências científicas. Cien Saude Colet, 27, 951–956. https://doi.org/10.1590/1413-81232022273.21822021

26. OPAS. (2022). Perguntas frequentes: Vacinas contra a COVID-19. OPAS/OMS. Disponível em: https://www.paho.org/pt/vacinas-contra-covid-19/perguntas-frequentes-vacinas-contra-covid-19. Acesso em: 8 de janeiro de 2023.

27. Couto, M. T., Barbieri, C. L. A., & De Souza Amorim Matos, C. C. (2021). Considerações sobre o impacto da COVID-19 na relação indivíduo-sociedade: Da hesitação vacinal ao clamor por uma vacina. Saúde e Sociedade, 30(1). https://doi.org/10.1590/S0104-12902021200450

28. Santos Domingues, C. M. A. (2021). Desafios para a realização da campanha de vacinação contra a COVID-19 no Brasil. Cad Saude Publica, 37(1). https://doi.org/10.1590/0102-311X00344620

29. OPAS. (2021). Vaccines and immunization. OPAS/OMS. Disponível em: https://www.who.int/health-topics/vaccines-and-immunization#tab=tab_1. Acesso em: 16 de janeiro de 2023.

30. Kriss, J. L., Reynolds, L. E., Wang, A., et al. (2021). COVID-19 vaccine second-dose completion and interval between first and second doses among vaccinated persons — United States, December 14, 2020−February 14, 2021. Morbidity and Mortality Weekly Report, 70, 389. https://doi.org/10.15585/mmwr.mm7011e2