

Temporal analysis of vaccine adverse effects for causation inference

Carla Sofia Carvalho*

Hitachi Digital Services, Lisbon, Portugal

(Dated: December 6, 2023)

Abstract

We present a formalism to assess the safety of the COVID19 vaccines and infer a causal relation between vaccine administration and adverse effects, in particular death. We use data on adverse effects from VAERS covering the time interval [01-10-2020, 31-12-2021] (downloaded on 1 Feb 2022). We measure a vaccine fatality rate of order 0.01 for all vaccine manufacturers and identify a strong heterogeneity in the vaccine toxicity across vaccine lots, spanning up to four orders of magnitude. We compute the correlation between vaccination dates and death dates, and produce an estimate of the time lag that maximises the correlation, finding that the onset of adverse effects happened statistically (i.e. 32-41 percentiles) within one day after the vaccine administration and that death happened statistically (i.e. 22-50 percentiles) within about two weeks after the vaccine administration. These results provide data-based insights that can guide public health measures and health insurance policies.

* Email: carla.carvalho@hitachids.com

Contents

| | |
|---|-----------|
| I. Introduction | 3 |
| II. The VAERS data | 4 |
| A. VAERS data files | 4 |
| B. Processing of VAERS data | 4 |
| C. Selection of VAERS data | 5 |
| III. Demographic analysis of the VAERS data | 5 |
| A. Organisation of VAERS data per vaccine manufacturer | 5 |
| B. Organisation of VAERS data per vaccine manufacturer, per symptom | 7 |
| C. Comparison of VAERS data per vaccine name | 9 |
| IV. Homogeneity analysis of the VAERS data | 11 |
| A. Organisation of VAERS data per vaccine manufacturer, per vaccine lot | 11 |
| V. Time analysis of the VAERS data | 13 |
| A. Correlations in the VAERS data | 13 |
| B. Percentile corresponding to the maximum correlation | 16 |
| VI. Conclusions | 17 |
| References | 18 |

I. INTRODUCTION

The COVID19 vaccines were authorised for sale in 2020 under an Emergency Use Authorisation (EUA). This authorisation was based on clinical trials of inappropriate design and of incomplete execution. In particular the clinical trials used disease specific endpoints instead of all-case mortality or morbidity, thus overestimating the benefit and concealing the harm [1]. Consequently, the clinical trials could not (nor had been designed to) conclude on prevention of infection or transmission, or on prevention of hospitalisation, severe disease or death [2]. Moreover, clinical trials were still on-going, hence limited data were available and absolute risk reduction metrics were not reported, which led to reporting bias [3]. Hence, at the time of the EUA, vaccine safety – and still less vaccine efficiency – had not been proven.

The roll-out of the COVID19 vaccination was followed by an unprecedented increase in the number of reports on adverse effects submitted to the Vaccine Adverse Event Reporting System (VAERS) database in the United States of America (USA).

Several studies have analysed the VAERS data focusing on different aspects, namely occurrence of myocarditis and pericarditis [4], and occurrence of deaths with age and comorbidity [5, 6]. Other studies have analysed similar vaccine reporting systems, namely the Suspected Adverse Effects (SAE) in the European Union/European Economic Area (EU/EEA), focusing on the inhomogeneous toxicity across vaccine batches [7]. Vaccine toxicity should pose a scientific concern for vaccine formulation and a quality concern for vaccine manufacturing, given the implications to public health [8].

The quantitative relations inferred from these studies establish a strong correlation between COVID19 vaccines and adverse effects. However, to our knowledge no study was focused on establishing a causal relation. In this study we present a formalism to infer a causal relation between COVID19 vaccine and adverse effects, and hence assess the safety of COVID19 vaccines. To minimise the impact of confounding factors [9], among the adverse effects we focused on deaths as the indisputable measure of vaccine safety.

Since the VAERS database relies on a spontaneous (or passive) reporting of adverse events that occur following vaccination, it is inherently limited and hence it should be analysed with caution [10]. Having these limitations in mind, we specify the working assumptions on the data, the theoretical properties of the selected methods, and the limitations of the current formalism.

The study is organised as follows. In section II, we present the VAERS data set and describe the data processing and data selection for the subsequent analysis. In section III, we organise the data per vaccine manufacturer and per symptom, producing a demographic analysis of the patient population. In section IV, we organise the data per vaccine manufacturer and per vaccine lot, producing a homogeneity analysis of each COVID19 vaccine. In section V, we organise the data per vaccine manufacturer in time and measure the lag-correlation between pairs of dates. Finally, in section VI, we conclude with insights from the current analysis and plans for the future analysis.

II. THE VAERS DATA

A. VAERS data files

Data on the adverse effects of the COVID19 vaccine were downloaded on 1 Feb 2022 from VAERS¹ in the form of comma-separated value files. We selected this data download to minimise contamination from subsequent vaccine doses. For each year [YEAR], we obtain the following three files:

- the [YEAR]VAERSDATA.csv file contains information related to the vaccine’s recipient (e.g. age, sex, state, patient outcomes, dates) and consists of: an array of size 50036 rows and 35 columns for year = 2020, and an array of size 748690 rows and 35 columns for year = 2021;
- the [YEAR]VAERSVAX.csv file contains information related to the vaccine (e.g. vaccine name, manufacturer, lot number, number of previous doses administered) and consists of: an array of size 60054 rows and 8 columns for year = 2020, and an array of size 792969 rows and 8 columns for year = 2021;
- the [YEAR]VAERSSYMPTOMS.csv file contains information on the adverse events, grouped into five columns, with entries in text (e.g. SYMPTOM1, ..., SYMPTOM5) and in coded terms using the MedDRA dictionary (e.g. SYMPTOMVERSION1, ..., SYMPTOMVERSION5) and consists of: an array of size 61335 rows and 11 columns for year = 2020, and an array of size 1001256 rows and 11 columns for year = 2021.

The files have a column named VAERS_ID containing the VAERS identification number, which we use to match the entries across the three files.

B. Processing of VAERS data

For each year we combined the data from the three files on the column VAERS_ID, obtaining: an array of size 50036 rows and 52 columns for year = 2020, and an array of size 748687 rows and 52 columns for year = 2021. We then concatenated the combined data from all years, resulting an array of size 798723 rows and 52 columns. We removed rows with null entries in the column named VAX_MANU containing the vaccine manufacturer (which corresponds to a fraction of 0.02) or with null entries in the column named VAX_LOT containing the vaccine lot (which corresponds to a fraction of 0.32). Hence, the resulting array has size 535526 rows and 52 columns.

We checked the consistency between deaths (as read off of column DIED) and death dates (as read off of column DATEDIED). Of all 7047 deaths (i.e. DIED = True), 6604 had valid death date (which corresponds to a fraction of 0.94) and 6762 had valid onset date (which corresponds to a fraction of 0.96). While all valid death dates correspond to deaths, there are deaths with an invalid death date. Of the 443 deaths with invalid death dates, 213 had valid onset dates (as read off of

¹ www.vaers.hhs.gov

column `ONSET_DATE`); conversely, of the 285 deaths with invalid onset dates, 55 had valid death dates. An idea was to assign the corresponding onset date (if valid) to the invalid death date, and conversely to assign the corresponding death date (if valid) to the invalid onset date.

Before making the assignment, we proceeded to check if the death cases (i.e. `DIED = True`) contained death in any of the five symptom columns. Of the 213 deaths with valid onset dates and invalid death dates, 156 contained death among the symptoms, so we assigned the onset date to the death date. Of the 55 deaths with valid death date and invalid onset dates, 34 contained death among the symptoms, so we assigned the death date to the onset date. Of the 230 death cases with invalid death date and invalid onset date, 152 contained death among the symptoms, so we removed them. We also checked if non-death cases (i.e. `DIED \neq True`) indeed did not contain death among the symptoms. Of the 26 non-death cases with death listed as a symptom, all had invalid death date but valid onset date, so we assigned the onset date to the death date and moreover changed `DIED` to `DIED = True`.

Finally, we removed the rows with null onset date or null vaccination date (as read off of column `VAX_DATE`). Hence, the resulting array has size 499824 rows and 52 columns.

We also checked the consistency between pairs of date columns. Noting that the onset date must come after the vaccination date, and that the received date (as read off of column `RECVDATE`) must come after the onset date, then the differences `ONSET_DATE - VAX_DATE` and `RECVDATE - ONSET_DATE`s must always be positive. In the case of death, nothing that the death date must come after the vaccination date, then the difference `DATEDIED - VAX_DATE` must always be positive. We removed all rows that violated these conditions, and set the lower time bound as the minimum value of `VAX_DATE` and the upper time bound as the maximum value of `RECVDATE`. The resulting array has size 491402 rows and 52 columns, and covers the interval [12-02-1900, 31-12-2021].

C. Selection of VAERS data

Since we are interested in analysing the adverse effects of the COVID19 vaccines only, we selected the entries with COVID19 in the column `VAX_NAME` and set a lower bound to the vaccination date of 01-10-2020. The resulting array has size 454573, rows and 52 columns, and covers the interval [01-10-2020, 21-12-2021].

Moreover, since we are interested in identifying vaccine lots, we removed the entries with invalid values in the columns `VAX_LOT`. The resulting array has size 452091 rows and 52 columns.

III. DEMOGRAPHIC ANALYSIS OF THE VAERS DATA

A. Organisation of VAERS data per vaccine manufacturer

We selected the VAERS data on the COVID19 vaccine manufacturers (as read off of column `VAX_MANU`). In particular, we filtered the data by the three vaccine manufacturers, namely:

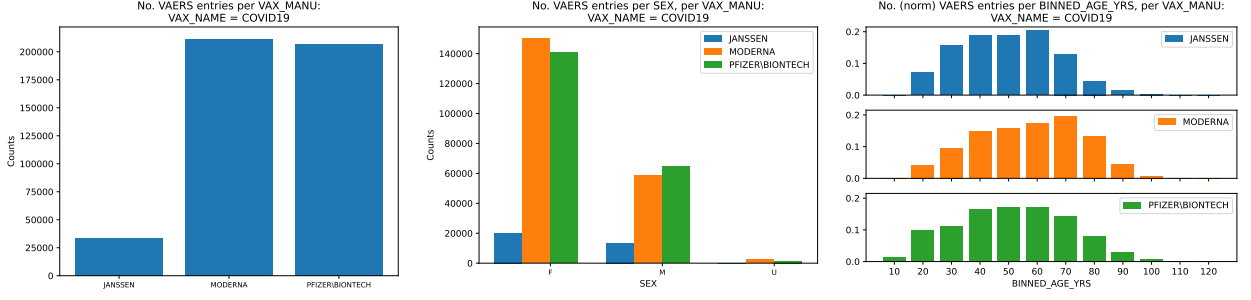


Figure 1: **Distribution of the COVID19 vaccine manufacturer, by sex and age bin, in the entire VAERS sample.** Left panel: Size of the sample of each vaccine manufacturer; Centre panel: Size of the sample per sex. Right panel: Size of the sample per age bin, normalised to the size of the sample.

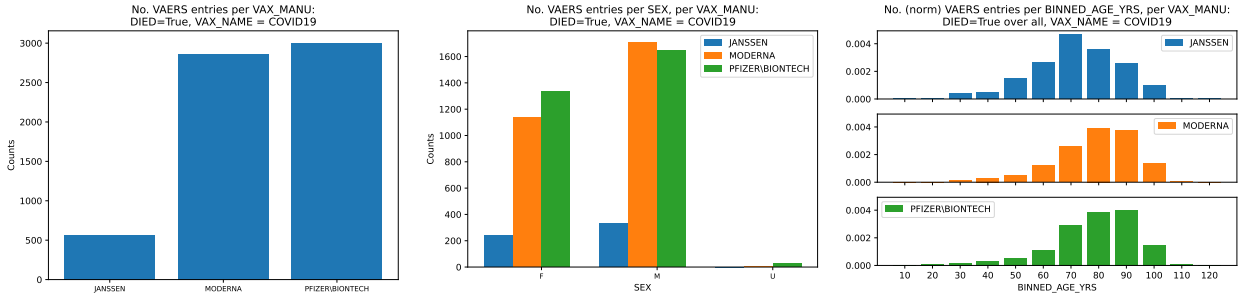


Figure 2: **Distribution of the COVID19 vaccine manufacturer, by sex and age bin, in the VAERS subsample restricted to deaths.** Left panel: Size of the sample of each vaccine manufacturer; Centre panel: Size of the sample per sex. Right panel: Size of the sample per age bin, normalised to the size of the sample.

JANSSEN, MODERNA and PFIZER\BIONTECH.² We produced the distribution of the VAERS data by `VAX_MANU` (Fig. 1, left panel). The fraction of entries per `VAX_MANU` is $[0.08, 0.47, 0.45]$.

For each `VAX_MANU`, we produced distributions of the demographical characteristics of the corresponding sample, in particular by sex, read off of column `SEX` (Fig. 1, centre panel) and by age bin of length 10, obtained from column `AGE` (Fig. 1, right panel), so that the age bin labelled x corresponds to the interval $[x - 10, x[$. Regarding sex per `VAX_MANU`, we observe that the fraction of each sex is $[0.60, 0.71, 0.68]$ for females, $[0.40, 0.28, 0.31]$ for males, with the remaining fraction referring to unknown value. Regarding age per `VAX_MANU`, we observe that the median age is $[45, 54, 47]$ and that the 90 percentile corresponds to the age bins labelled $[60, 70, 70]$, with the top age bins being $[60, 70, 60]$.

We also produced the distribution of the VAERS data by vaccine manufacturer, grouped by adverse effects, in particular grouped by hospitalisation (read off of column `HOSPITAL`, i.e. whether the patient required hospitalisation) and by deaths (read off of column `DIED`, i.e. whether the patient died). We observe that fractions $[0.089, 0.054, 0.076]$ of entries required hospitalisation, whereas fractions $[0.017, 0.013, 0.015]$ of entries led to death.

To contextualise the occurrence of deaths, from the subsample defined by `DIED = True`, we pro-

² Unless indicated otherwise, the subsequent results per vaccine manufacturer are presented in this order.

duced the same distributions as above for the VAERS entries corresponding to deaths (Fig. 2). We show the distribution of deaths per VAX_MANU in Fig. 2, left panel. The fraction of entries per VAX_MANU is [0.09, 0.44, 0.47].

Regarding sex per VAX_MANU, we observe that the fraction of each sex is [0.42, 0.40, 0.44] for females, [0.58, 0.60, 0.55] for males, with the remaining fraction referring to unknown value (Fig. 2, centre panel). Regarding age per VAX_MANU, we observe that the median death age is [67, 76, 76] and that the 90 percentile corresponds to the age bins labelled [80, 90, 90], with the top age bins being [70, 80, 90] (Fig. 2, right panel). Regarding hospitalisation per VAX_MANU, we observe that fractions [0.46, 0.33, 0.43] of entries corresponding to deaths required hospitalisation which correspond to fractions [0.008, 0.004, 0.006] of all entries.

We then compared the COVID19 vaccine fatality rate per VAX_MANU with the COVID19 infection fatality rate (IFR) obtained in the latest meta-regression analysis [11]. According to Ref. [11], the median IFR of COVID19 per age bin ranges between 3×10^{-6} and 5×10^{-3} , whereas the global IFR of COVID19 is 7×10^{-4} for 0–69 years old (Table I, last column). Note that the focus of Ref. [11] was on an accurate estimate of the IFR of COVID19 among non-elderly people, motivated by the fact that non-elderly people (i.e. younger than 70 years old) represent 94% of the global population.

Conversely, in the sample of vaccinated patients covered by the VAERS records, our results show vaccine fatality rates per VAX_MANU between 10 and 10^2 times (i.e. between one and two orders of magnitude) larger than the IFR of COVID19 in almost all age bins, the sole exception being the [60, 70[bin where the fatality rates are of the same order of magnitude (Table I). This amounts to a global vaccine fatality over all age bins of order 0.02. We also computed the global vaccine fatality over all age bins up to 70, finding [0.010, 0.006, 0.006] per VAX_MANU, hence of order 0.01.³ For an easier visualisation, we plotted the fatality rates per age bin in Fig. 3. The global vaccine fatality rate for non-elderly people is non-negligible and moreover up to two orders of magnitude larger than the IFR of COVID19 in the sample of infected patients in Ref. [11].

B. Organisation of VAERS data per vaccine manufacturer, per symptom

We produced the distribution of the VAERS data by COVID19 vaccine manufacturer, grouped by symptom (read off of columns SYMPTOM#). We observe that SYMPTOM1 is the only symptom column that does not have missing values, with the columns from SYMPTOM2 to SYMPTOM5 having progressively more missing values.

³ Actually we are measuring the probability of death by the vaccine conditional on the adverse event being reported $P(\text{death, vaccine}|\text{report})$ which relates to the probability of death by the vaccine $P(\text{death, vaccine})$ as follows:

$$P(\text{death, vaccine}|\text{report}) = P(\text{death, vaccine}) \times \frac{P(\text{report}|\text{death, vaccine})}{P(\text{report})}. \quad (1)$$

Since the probability of reporting conditional on there being a death is expectedly larger than the probability of reporting, then $P(\text{report}|\text{death, vaccine})/P(\text{report}) \geq 1$. It follows that $P(\text{death, vaccine}|\text{report}) \geq P(\text{death, vaccine})$, hence we are estimating an upper bound of the lethality.

| Age bin | JANSSEN | MODERNA | PFIZER\BIONTECH | COVID19 |
|-----------|---|---|---|--------------------|
| [0, 20[| $3 \times 10^{-5} \pm 3 \times 10^{-5}$ | $1 \times 10^{-5} \pm 7 \times 10^{-6}$ | $1 \times 10^{-4} \pm 3 \times 10^{-5}$ | 3×10^{-6} |
| [20, 30[| $4 \times 10^{-4} \pm 1 \times 10^{-4}$ | $1 \times 10^{-4} \pm 3 \times 10^{-5}$ | $2 \times 10^{-4} \pm 4 \times 10^{-5}$ | 3×10^{-5} |
| [30, 40[| $4 \times 10^{-4} \pm 1 \times 10^{-4}$ | $3 \times 10^{-4} \pm 4 \times 10^{-5}$ | $3 \times 10^{-4} \pm 6 \times 10^{-5}$ | 1×10^{-4} |
| [40, 50[| $1 \times 10^{-3} \pm 2 \times 10^{-4}$ | $5 \times 10^{-4} \pm 6 \times 10^{-5}$ | $5 \times 10^{-4} \pm 1 \times 10^{-1}$ | 4×10^{-4} |
| [50, 60[| $3 \times 10^{-3} \pm 3 \times 10^{-4}$ | $1 \times 10^{-3} \pm 1 \times 10^{-4}$ | $1 \times 10^{-3} \pm 2 \times 10^{-4}$ | 1×10^{-3} |
| [60, 70[| $5 \times 10^{-3} \pm 4 \times 10^{-4}$ | $3 \times 10^{-3} \pm 2 \times 10^{-4}$ | $3 \times 10^{-3} \pm 5 \times 10^{-4}$ | 5×10^{-3} |
| [70, 80[| $4 \times 10^{-3} \pm 3 \times 10^{-4}$ | $4 \times 10^{-3} \pm 2 \times 10^{-4}$ | $4 \times 10^{-3} \pm 6 \times 10^{-4}$ | |
| [80, 90[| $3 \times 10^{-3} \pm 3 \times 10^{-4}$ | $4 \times 10^{-3} \pm 2 \times 10^{-4}$ | $4 \times 10^{-3} \pm 6 \times 10^{-4}$ | |
| [90, 100[| $1 \times 10^{-3} \pm 2 \times 10^{-4}$ | $1 \times 10^{-3} \pm 1 \times 10^{-4}$ | $2 \times 10^{-3} \pm 2 \times 10^{-4}$ | |
| [0, 70[| $1 \times 10^{-2} \pm 6 \times 10^{-4}$ | $6 \times 10^{-3} \pm 4 \times 10^{-4}$ | $6 \times 10^{-3} \pm 3 \times 10^{-4}$ | 7×10^{-4} |

Table I: **Fatality rate per age bin.** Column 1: Age bin over which the infection rates were computed. Columns 2–4: Infection fatality rate of COVID19 vaccine per vaccine manufacturer per age bin. Column 5: Median infection fatality rate of COVID19 infection per age bin. The errors of the fatality rate were computed by the propagation of the errors of a division of counts, where the error of each count n is given by \sqrt{n} .

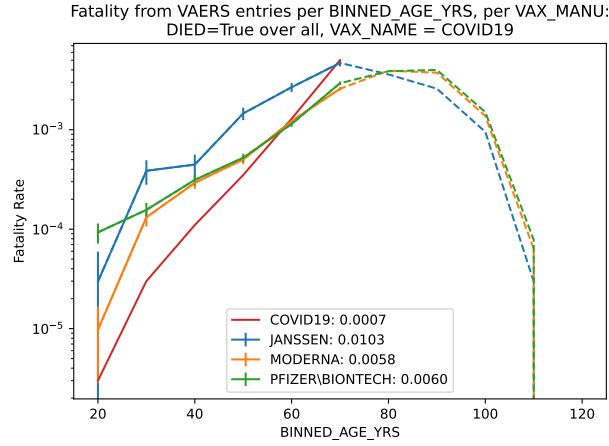


Figure 3: **Fatality rate per age bin of COVID19 infection and of COVID19 vaccination (per vaccine manufacturer).** Filled lines: Fatality rates (of infection [11] and of vaccination) per age bin up to 70 years old. Dashed lines: Fatality rates (of vaccination only) per age bin larger than or equal to 70 years old. In the legend, we indicate the fatality rate over all age bins up to 70 years old.

We also produced the distribution of the top-10 most frequent symptoms per symptom column separately for each COVID19 vaccine manufacturer. We observe that the top-10 most frequent entries in column SYMPTOM_1 over all VAX_MANU are {Chills, Arthralgia, Dizziness, COVID-19, Asthenia, Fatigue, Injection site erythema, Headache, Expired product administered, Erythema}.

To contextualise the severity of the reported symptoms, using the subsample defined by DIED=True, we produced the distribution of the top-10 most frequent symptoms corresponding to deaths. We observe that the top-10 most frequent entries in column SYMPTOM_1 over all VAX_MANU are {Death, COVID-19, Acute respiratory failure, Asthenia, Acute kidney injury, Cardiac arrest, Autopsy, Abdominal pain, COVID-19 pneumonia, Cerebrovascular accident}.

For an easier reading, in Table II, we list the top-10 most frequent symptoms in column

| | JANSSEN | | MODERNA | | PFIZER\BIONTECH | |
|--------------------|---------------|------------------|--------------|--------------------|-----------------|-------------------|
| SYMPTOM1 | All | DIED = True | All | DIED = True | All | DIED = True |
| No. | 2047 | 88 | 14874 | 567 | 11266 | 396 |
| (fraction wrt All) | | (0.043) | | (0.038) | | (0.035) |
| Top-10 | Chills | Death | Chills | Death | Dizziness | Death |
| | Dizziness | COVID-19 | Arthralgia | COVID-19 | Chills | COVID-19 |
| | Arthralgia | Acute resp fail | Inj erythema | Asthenia | Arthralgia | Acute resp fail |
| | Asthenia | Asthenia | Dizziness | Acute resp fail | COVID-19 | Acute kidney inj |
| | COVID-19 | Acute kidney inj | Exp product | Cardiac arrest | Asthenia | Asthenia |
| | Headache | Autopsy | Asthenia | Acute kidney inj | Product err | Cardiac arrest |
| | Fatigue | Abd pain | Erythema | Autopsy | Fatigue | Autopsy |
| | Blood test | Cardiac arrest | Fatigue | Cerebrovasc acc | Headache | Abd pain |
| | Product admin | Acute RDS | COVID-19 | COVID-19 pneum | Chest discom | Acute myocard inf |
| | Abd pain | COVID-19 pneum | Headache | Cardio-resp arrest | Anxiety | COVID-19 pneum |

Table II: **Symptoms with the top number of entries per column SYMPTOM1, per COVID19 vaccine manufacturer.** Row 1: COVID19 vaccine manufacturers. Row 2: Specification of the sample used to count vaccine lots: entire sample (denoted by “All”) and subsample restricted to deaths (denoted by **DIED = True**); Row 3: Number of vaccine lots per sample (fraction of lots with respect to the number of lots in entire sample); Row 4: List of symptoms with the top number of entries per column SYMPTOM1, per sample.

SYMPTOM_1 per VAX_MANU, both for the entire sample and for the subsample **DIED=True**.

C. Comparison of VAERS data per vaccine name

To contextualise the counts of symptoms, we looked at the VAERS data over all vaccine names (read off of column **VAX_NAME**) for the same time interval and produced the distribution of the number of VAERS entries per **VAX_NAME** (Fig. 4, top panel). Note that Fig. 1, left panel, covers a subset of Fig. 4 corresponding to **VAX_NAME = COVID19**. We observe that the number of VAERS entries for each COVID19 vaccine is 10^2 times (i.e. two orders of magnitude) larger than for the vaccines with the next largest number of entries (i.e. ZOSTER, INFLUENZA and HPV).⁴

To contextualise the counts of symptoms within the occurrence of deaths, using the subsample defined by **DIED=True**, we produced the distribution per **VAX_NAME** of the number of VAERS entries corresponding to deaths (Fig. 4, middle panel), and the corresponding distribution normalised to the number of VAERS entries in each **VAX_NAME** (Fig. 4, bottom panel). We observe that the number of VAERS deaths for each COVID19 vaccine is 10^2 times (i.e. two orders of magnitude) larger than for the vaccines with the next largest number of deaths (i.e. DTAP+IPV+HIB, INFLUENZA and DTAP+HEPB+IPV). We also observe that, within the VAERS data, the fatality rate of the COVID19 vaccines is only surpassed by the fatality rate of the DTAP+IPV+HIB vaccine and of the DTAP+HEPB+IPV vaccine.

⁴ Note that, for a fair comparison, we should normalise the number of VAERS entries per **VAX_NAME** to the number of vaccines of each type administered during the same time interval.

IV. HOMOGENEITY ANALYSIS OF THE VAERS DATA

A. Organisation of VAERS data per vaccine manufacturer, per vaccine lot

We produced the distribution of the number of adverse effects per COVID19 vaccine lot. We used the counts of the listed symptoms as a measure of the vaccine toxicity across vaccine lots, per vaccine manufacturer. The number of vaccine lots per COVID19 vaccine manufacturer is indicated in Table III. Moreover, we assume that approximately the same number of doses of each lot were

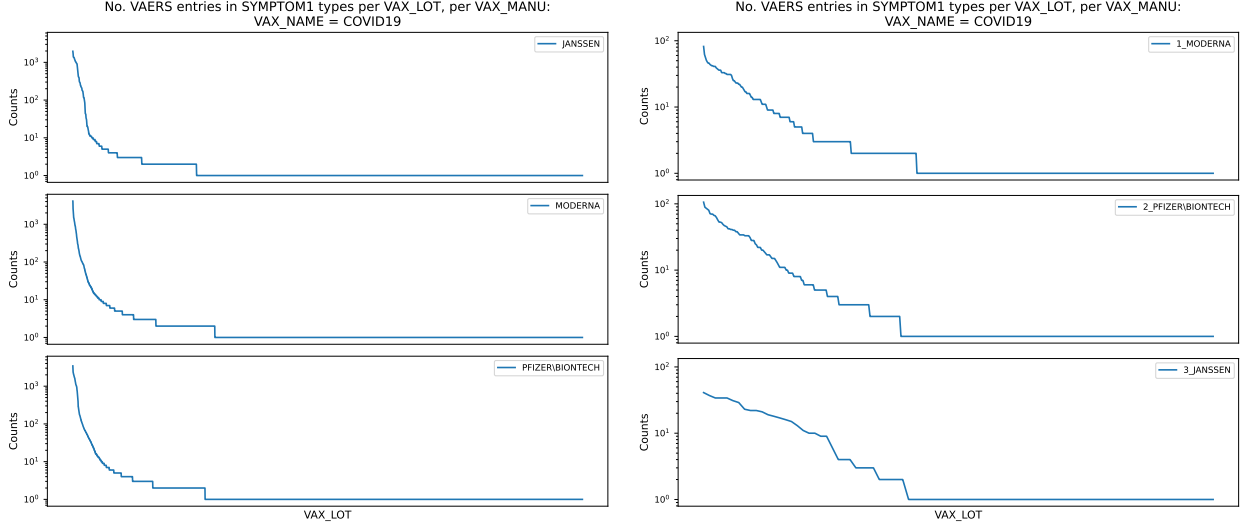


Figure 5: **Distribution of the COVID19 vaccine manufacturer, by vaccine lot, in the VAERS data.** Left panel: Counts of column SYMPTOM1 per vaccine lot for entire sample; Right panel: Counts of column SYMPTOM1 per vaccine lot for the DIED = True subsample.

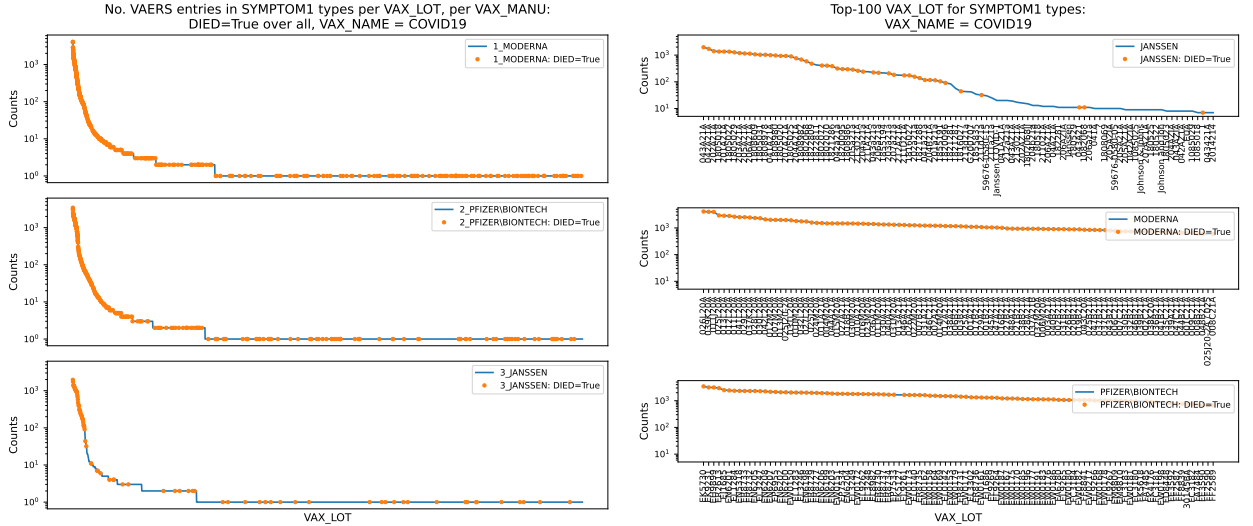


Figure 6: **Distribution of the COVID19 vaccine manufacturer, by vaccine lot, in the VAERS data, for the entire sample (blue line) and for the DIED = True subsample (orange dots).** Left panel: Counts of column SYMPTOM1 per vaccine lot; Right panel: Top-100 vaccine lots in the counts of column SYMPTOM1.

| | JANSSEN | | MODERNA | | PFIZER\BIONTECH | |
|--------------------|---------|-------------|---------|-------------|-----------------|-------------|
| VAX_LOT | All | DIED = True | All | DIED = True | All | DIED = True |
| No. | 2047 | 88 | 14874 | 567 | 11266 | 396 |
| (fraction wrt All) | | (0.043) | | (0.038) | | (0.035) |
| Top-10 | 043A21A | 1805031 | 026L20A | 039K20A | EK5730 | EN6201 |
| | 042A21A | 043A21A | 039K20A | 012L20A | EK9231 | EL9261 |
| | 202A21A | 1805018 | 011J20A | 013L20A | EH9899 | EN5318 |
| | 1805018 | 042A21A | 025L20A | 037K20A | ER2613 | EL3248 |
| | 201A21A | 1805029 | 013L20A | 025L20A | EJ1685 | EN6200 |
| | 1805022 | 205A21A | 012L20A | 010M20A | EN6201 | EL9269 |
| | 1805029 | 1805022 | 011L20A | 012M20A | EL1284 | EN6202 |
| | 203A21A | 203A21A | 041L20A | 029L20A | EN5318 | EN6198 |
| | 205A21A | 202A21A | 029L20A | 007M20A | ER8733 | EM9810 |
| | 206A21A | 1805025 | 037K20A | 030L20A | ER8732 | EL3249 |

Table III: **Vaccine lots with the top number of adverse effects per column SYMPTOM1, per COVID19 vaccine manufacturer.** Row 1: COVID19 vaccine manufacturers. Row 2: Specification of the sample used to count vaccine lots: entire sample (denoted by “All”) and subsample restricted to deaths (denoted by `DIED = True`); Row 3: Number of vaccine lots per sample (fraction of lots with respect to the number of lots in entire sample); Row 4: List of vaccine lots with the top number of adverse effects per column SYMPTOM1, per sample.

administered.

We plot the vaccine lots in decreasing order of the number of adverse effects in SYMPTOM1 (Fig. 5, left panel). We observe that the number of adverse effects vary by up to four orders of magnitude over the vaccine lots of the same manufacturer, which deviates strongly from a uniform distribution.⁵ This indicates that the vaccine lots are strongly inhomogeneous. The number of adverse effects in logarithmic scale not only deviates from the uniform distribution, but appears to vary exponentially. In Table III, for each VAX_MANU, we list the top-10 vaccine lots in the number of entries in column SYMPTOM1.

Using the subsample defined by `DIED = True`, we produced the corresponding distribution (Fig. 5, right panel). We observe that the number of adverse effects vary by up to two orders of magnitude over the the vaccine lots of the same manufacturer. The number of adverse effects in logarithmic scale not only deviates from the uniform distribution, but appears to vary linearly, followed by a plateau. In Table III, for each VAX_MANU and `DIED = True`, we list the top-10 vaccine lots in the number of entries in column SYMPTOM1.

For an easier visualisation, in Fig. 6 (left and right panels), we plot the vaccine lots in decreasing order of the number of adverse effects in SYMPTOM1 for both the entire sample (line) and the `DIED = True` subsample (dots superposed to the line). This shows that the vaccine lots associated with deaths are among the lots with the largest number of adverse effects, hence among the most toxic vaccine lots.

⁵ Homogeneous vaccine lots have similar (or of the same order) numbers of adverse effects, hence the distribution of the numbers over the vaccine lots will be uniform. A deviation from a uniform distribution is thus a measure of the deviation from homogeneity.

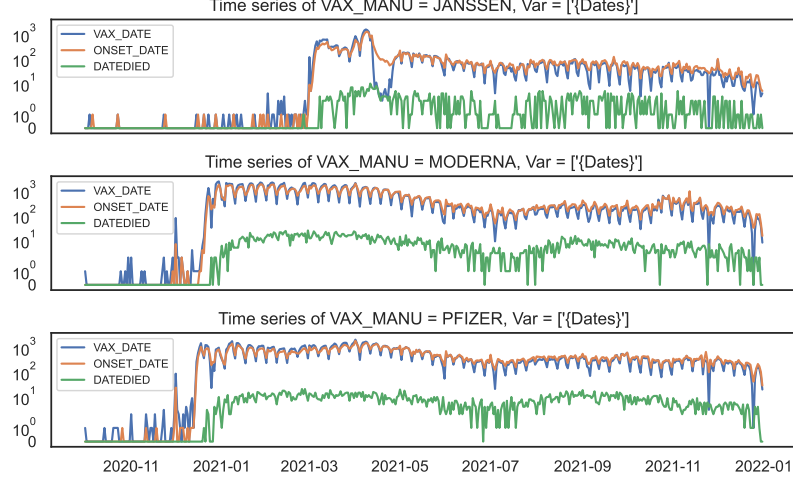


Figure 7: **Time series of counts by date, by COVID19 vaccine manufacturer, in the VAERS data.** Top panel: JANSSEN; Middle panel: MODERNA; Bottom panel: PFIZER\BIONTECH.

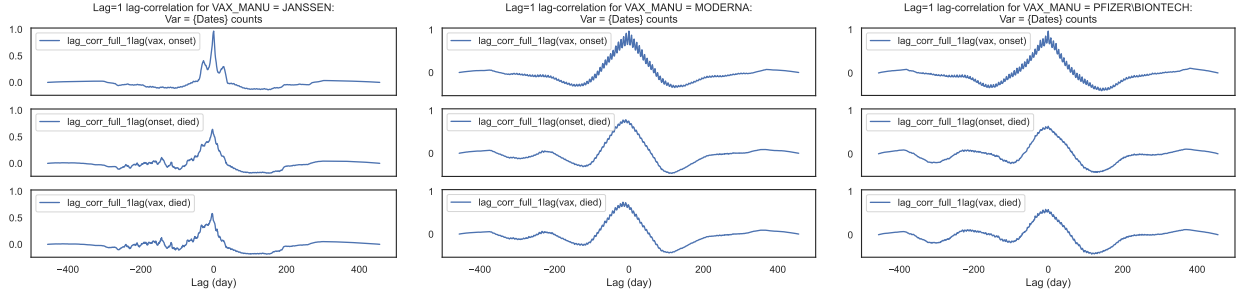


Figure 8: **Lag-correlation between the time series at VAX_DATE, ONSET_DATE and DATEDIED, per COVID19 vaccine manufacturer, in the entire VAERS sample.** Left panel: JANSSEN; Centre panel: MODERNA; Right panel: PFIZER\BIONTECH.

V. TIME ANALYSIS OF THE VAERS DATA

A. Correlations in the VAERS data

In order to establish whether there is a causal relation between vaccine administration and adverse effect onset, we proceeded to compute the correlation between pairs of dates for different time lags.

We selected the following date columns: **VAX_DATE**, **ONSET_DATE** and **DATEDIED**, which we identify generically as \mathbf{x}_i . We built time series $\mathbf{x}_{i,t}$ of the number of entries at each date column \mathbf{x}_i for the same, equally spaced time array (Fig. 7).

We computed the lag-correlation between any two time series $\mathbf{x}_{i,t}$ and $\mathbf{x}_{i',t}$ for a varying time lag l as follows

$$\text{LagCorr}(\mathbf{x}_i, \mathbf{x}_{i'}, l) = \frac{\sum_t (\mathbf{x}_{i,t} - \text{mean}(\mathbf{x}_{i,t})) (\mathbf{x}_{i',t-l} - \text{mean}(\mathbf{x}_{i',t-l}))}{\sqrt{\sum_t (\mathbf{x}_{i,t} - \text{mean}(\mathbf{x}_{i,t}))^2 (\mathbf{x}_{i',t-l} - \text{mean}(\mathbf{x}_{i',t-l}))^2}}. \quad (2)$$

We plot the lag-correlations in Fig. 8.

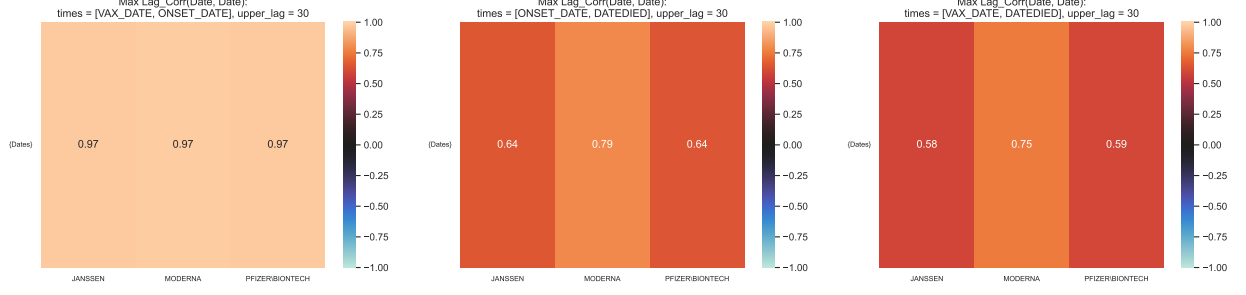


Figure 9: **Maximum lag-correlation between the time series at VAX_DATE, ONSET_DATE and DATEDIED, per COVID19 vaccine manufacturer, in the entire VAERS sample.** Left panel: $\text{LagCorr}(\text{VAX_DATE}, \text{ONSET_DATE}, l_{\max})$; Centre panel: $\text{LagCorr}(\text{ONSET_DATE}, \text{DATEDIED}, l_{\max})$; Right panel: $\text{LagCorr}(\text{VAX_DATE}, \text{DATEDIED}, l_{\max})$.

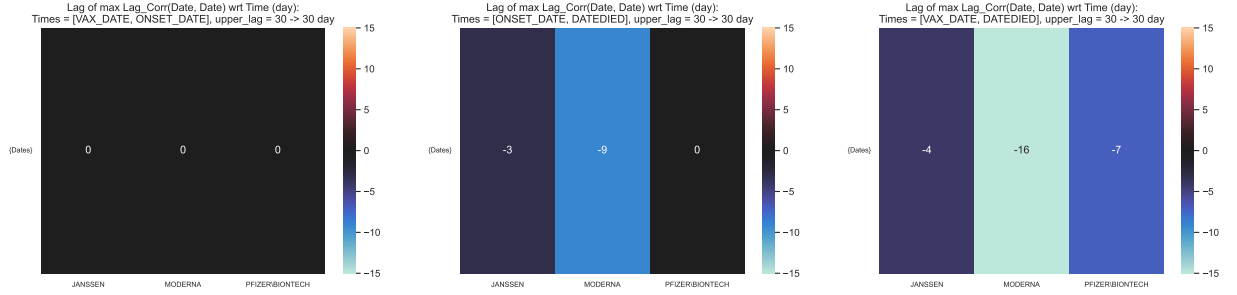


Figure 10: **Time lag corresponding to the maximum lag-correlation between the time series at VAX_DATE, ONSET_DATE and DATEDIED, per COVID19 vaccine manufacturer, in the entire VAERS sample.** Left panel: $\text{LagCorr}(\text{VAX_DATE}, \text{ONSET_DATE}, l_{\max})$; Centre panel: $\text{LagCorr}(\text{ONSET_DATE}, \text{DATEDIED}, l_{\max})$; Right panel: $\text{LagCorr}(\text{VAX_DATE}, \text{DATEDIED}, l_{\max})$.

We then found the time lag $l_{\max, ii'}$ such that

$$\text{LagCorr}(\mathbf{x}_i, \mathbf{x}_{i'}, l_{\max, ii'}) = \max(\text{LagCorr}(\mathbf{x}_i, \mathbf{x}_{i'}, l_{ii'})) \quad (3)$$

for $l_{ii'} \leq \text{upper_lag}$. By definition, if $\text{LagCorr}(\mathbf{x}_i, \mathbf{x}_{i'}, l_{\max, ii'}) < 0$, then \mathbf{x}_i comes before $\mathbf{x}_{i'}$; conversely, if $\text{LagCorr}(\mathbf{x}_i, \mathbf{x}_{i'}, l_{\max, ii'}) > 0$, then \mathbf{x}_i comes after $\mathbf{x}_{i'}$. We plot the maxima of the lag-correlations in Fig. 9 and the corresponding time lags in Fig. 10. We note that the maxima of the lag-correlations over all vaccine manufacturers are positive values (which means that each two dates considered are in-phase) and all equal to/larger than 0.6 (which means that the correlations are significant). We describe the results below:

- For $\text{LagCorr}(\text{VAX_DATE}, \text{ONSET_DATE}, l_{\max})$, the time lags are consistent with zero over all vaccine manufacturers; this implies that high/low counts of the onset of symptoms occur statistically within 1 day after high/low counts of vaccine administration.
- For $\text{LagCorr}(\text{ONSET_DATE}, \text{DATEDIED}, l_{\max})$, the time lags are negative over all vaccine manufacturers, ranging between 0 and -9; this implies that high/low counts of deaths occur statistically within 9 days after high/low counts of the onset of symptoms.

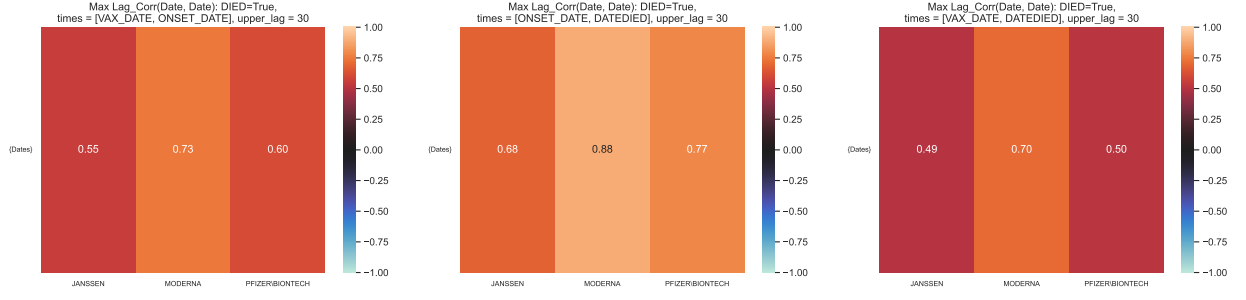


Figure 11: **Maximum lag-correlation between the time series at VAX_DATE, ONSET_DATE and DATEDIED, per COVID19 vaccine manufacturer, in the VAERS subsample restricted to deaths.** Left panel: $\text{LagCorr}(\text{VAX_DATE}, \text{ONSET_DATE}, l_{\max})$; Centre panel: $\text{LagCorr}(\text{ONSET_DATE}, \text{DATEDIED}, l_{\max})$; Right panel: $\text{LagCorr}(\text{VAX_DATE}, \text{DATEDIED}, l_{\max})$.

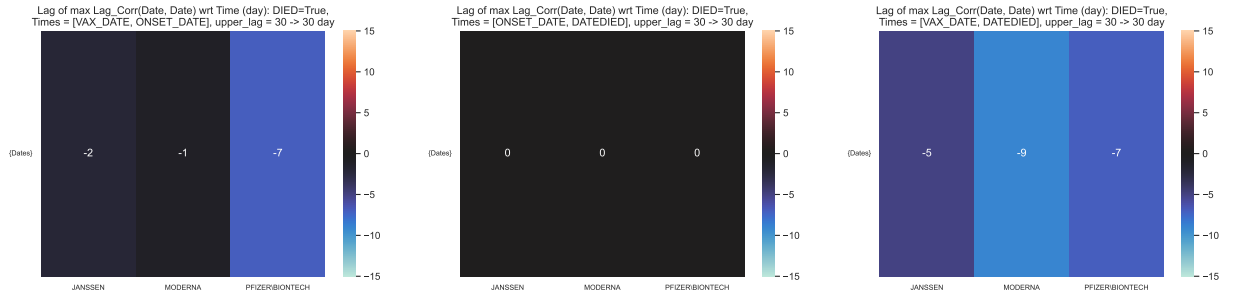


Figure 12: **Time lag corresponding to the maximum lag-correlation between the time series at VAX_DATE, ONSET_DATE and DATEDIED, per COVID19 vaccine manufacturer, in the VAERS subsample restricted to deaths.** Left panel: $\text{LagCorr}(\text{VAX_DATE}, \text{ONSET_DATE}, l_{\max})$; Centre panel: $\text{LagCorr}(\text{ONSET_DATE}, \text{DATEDIED}, l_{\max})$; Right panel: $\text{LagCorr}(\text{VAX_DATE}, \text{DATEDIED}, l_{\max})$.

- For $\text{LagCorr}(\text{VAX_DATE}, \text{DATEDIED}, l_{\max})$, the time lags are negative over all vaccine manufacturers, ranging between -4 and -16; this implies that high/low counts of deaths occur statistically within 16 days after high/low counts of vaccine administration.

Using the subsample defined by $\text{DIED} = \text{True}$, we repeated the procedure above, producing equally spaced time series of the selected date columns, computing the lag-correlation between pairs of time series, and finding the maximum lag-correlation and the corresponding time lag (Figs. 11 and 12). We obtained qualitatively similar results. The maxima of the lag-correlations over all vaccine manufacturers are positive values (which means that each two dates considered are in-phase) and all equal to/larger than 0.5 (which means that the correlations are significant). We describe the results below:

- For $\text{LagCorr}(\text{VAX_DATE}, \text{ONSET_DATE}, l_{\max})$, the time lags are negative over all vaccine manufacturers, ranging between -1 and -7; this implies that high/low counts of the onset of symptoms occur statistically within 7 days after high/low counts of vaccine administration.
- For $\text{LagCorr}(\text{ONSET_DATE}, \text{DATEDIED}, l_{\max})$, the time lags are consistent with zero over all vaccine manufacturers; this implies that high/low counts of deaths occur statistically within 1 day after high/low counts of the onset of symptoms.

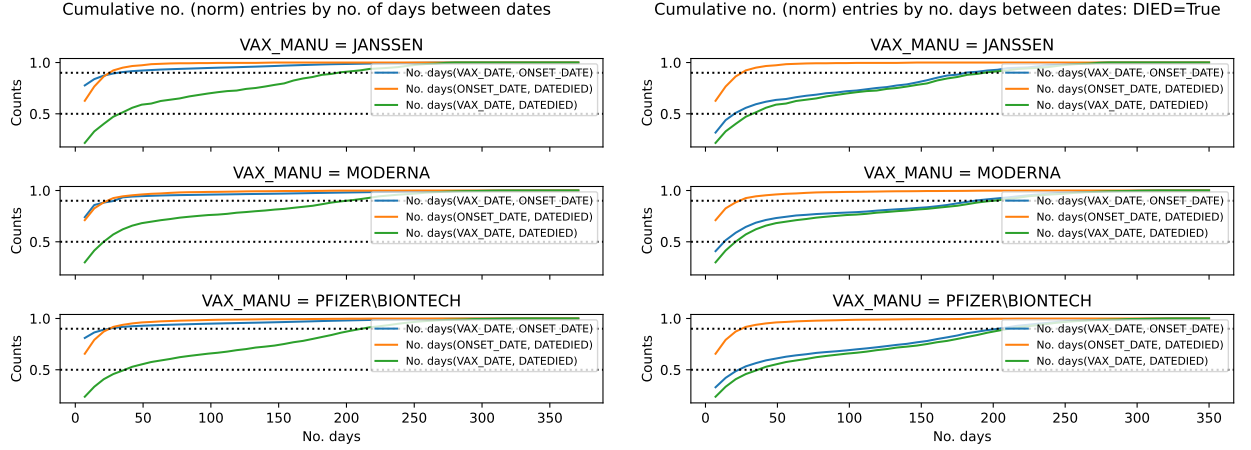


Figure 13: **Cumulative distribution of the COVID19 vaccine manufacturer, by the difference between pairs of dates, in the VAERS data.** Normalised cumulative number of entries for the difference between pairs of dates, per COVID19 vaccine manufacturer, with the horizontal lines at 0.5 and 0.9 marking the 50 and 90 percentile respectively. Left panel: Entire sample. Right panel: Subsample restricted to deaths.

- For $\text{LagCorr}(\text{VAX_DATE}, \text{DATEDIED}, l_{\max})$, the time lags are negative over all vaccine manufacturers, ranging between -5 and -9; this implies that high/low counts of deaths occur statistically within 9 days after high/low counts of vaccine administration.

B. Percentile corresponding to the maximum correlation

In order to quantify the confidence of the time lag corresponding to the maximum lag-correlation, we proceeded to compute the distribution of the difference between the selected pairs of dates.

For the selected date columns (i.e. `VAX_DATE`, `ONSET_DATE` and `DATEDIED`) and for each `VAX_MANU`, we computed the following date differences in number of days:

$$\text{No.days}(\text{VAX_DATE}, \text{ONSET_DATE}) = \text{ONSET_DATE} - \text{VAX_DATE}, \quad (4)$$

$$\text{No.days}(\text{ONSET_DATE}, \text{DATEDIED}) = \text{DATEDIED} - \text{ONSET_DATE}, \quad (5)$$

$$\text{No.days}(\text{VAX_DATE}, \text{DATEDIED}) = \text{DATEDIED} - \text{VAX_DATE}, \quad (6)$$

and produced the corresponding cumulative distributions (Fig. 13, left panel). We selected the bin size equal to seven days. We then found the percentile corresponding to the time lag of maximum lag-correlation. We describe the results below:

- For $\text{LagCorr}(\text{VAX_DATE}, \text{ONSET_DATE}, l_{\max})$, the time lags are consistent with zero over all vaccine manufacturers; this implies that the maximum lag-correlation happens within the percentile [32, 41, 33].
- For $\text{LagCorr}(\text{ONSET_DATE}, \text{DATEDIED}, l_{\max})$, the time lags are negative over all vaccine manufacturers, ranging between 0 and -9; this implies that the maximum lag-correlation happens within the percentile [63, 83, 66].

- For $\text{LagCorr}(\text{VAX_DATE}, \text{DATEDIED}, l_{\max})$, the time lags are negative over all vaccine manufacturers, ranging between -4 and -16; this implies that the maximum lag–correlation happens within the percentile [22, 50, 24].

This means that, over all vaccine manufacturers: a) for 32–41% of the dates, less than a week has elapsed from the vaccination date until the time of the onset of symptoms; b) for 63–83% of the dates, less than two weeks have elapsed from the time of the onset of symptoms until the death date; c) for 22–50% of the dates, about two weeks have elapsed from the vaccination date until the death date.

Using the subsample defined by $\text{DIED} = \text{True}$, we produced the corresponding cumulative distributions (Fig. 13, right panel) and found the percentile corresponding to the time lag of maximum lag–correlation. We describe the results below:

- For $\text{LagCorr}(\text{VAX_DATE}, \text{ONSET_DATE}, l_{\max})$, the time lags are negative over all vaccine manufacturers, ranging between -1 and -7; this implies that the maximum lag–correlation happens within the percentile [32, 41, 33].
- For $\text{LagCorr}(\text{ONSET_DATE}, \text{DATEDIED}, l_{\max})$, the time lags are consistent with zero over all vaccine manufacturers; this implies that the maximum lag–correlation happens within the percentile [63, 71, 66].
- For $\text{LagCorr}(\text{VAX_DATE}, \text{DATEDIED}, l_{\max})$, the time lags are negative over all vaccine manufacturers, ranging between -5 and -9; this implies that the maximum lag–correlation happens within the percentile [49, 42, 24].

Recall that, for this data subset, all vaccinations lead to deaths. This means that, over all vaccine manufacturers: a) for 32–41% of the dates, about a week has elapsed from the vaccination date until the time of the onset of symptoms; b) for 24–49% of the dates, less than two weeks have elapsed from the vaccination date until the time of death. Recall also that, for this data subset, all symptom onsets lead to deaths. This means that for 63–71% of the dates, it takes less than one week between the onset of symptoms and death.

VI. CONCLUSIONS

In this study, we analysed a VAERS data set covering [01–10–2020, 31–12–2021]. We focused on the COVID19 vaccine, thus filtering the data by the three vaccine manufacturers, namely: JANSSEN, MODERNA and PFIZER\BIONTECH. We summarise our results below.

- We measured the COVID19 vaccine fatality rate per vaccine manufacturer. We found that the vaccine fatality rate (conditional on being reported in VAERS) is of order 0.01 across the vaccine manufacturers for the age interval of [0, 70[years, hence about ten times larger than the COVID19 infection fatality rate (conditional on being infected) for the same age interval (which is estimated to be 7×10^{-4}). Since vaccines are prophylactic, they are administered

mostly to healthy people. Such fatality rate implies that the vaccine is causing more harm than the disease, thus rendering the vaccines unsafe. This should suffice to prompt the immediate halt of the clinical trial and/or withdrawal of the vaccines from the market.

- We also measured the number of vaccine adverse events per vaccine manufacturer. We found that the number of adverse events spans four orders of magnitude across the vaccine lots, regardless the vaccine manufacturer. A ready-for-market vaccine should have a uniform distribution of the number of adverse effects over the vaccine lots. Such inhomogeneity attests to a poor safety record and a poor manufacturing process of the vaccines, thus rendering the vaccines improper for general administration. This should suffice to prompt the immediate withdrawal of the vaccines from the market.
- We also measured the lag-correlation between pairs of dates, namely vaccination date and death date. We found that, over all vaccine manufacturers, the maximum lag-correlation is positive and that the corresponding time lag is negative. These observations imply that a peak in deaths occurs statistically within a time lag after a peak in vaccinations, thus suggesting that vaccination is causing deaths.

We conclude that the analysis of the VAERS data can provide epidemiological insights following the introduction of a new drug, and hence guide the formulation of procedures (e.g. hospital admittance or insurance coverage) based on the estimated time scales (e.g. time lags between drug administration and off-set of symptoms). In a forthcoming study, we plan to analyse the geographic distribution of adverse effects as a result of vaccine administration, which will provide further epidemiological insights.

-
- [1] J.B. Classen, US COVID-19 vaccines proven to cause more harm than good based on pivotal clinical trial data analyzed using the proper scientific endpoint ‘all cause severe morbidity’, *Trends Int Med* 2021, 1 (1): 1–6 (<https://www.scivisionpub.com/abstract-display.php?id=1811>) [accessed 01 Aug 2023]
 - [2] P. Olliaro, E. Torreele, M. Vaillant, COVID-19 vaccine efficacy and effectiveness – the elephant (not) in the room, *The Lancet Microbe* 2021, 2, 7, E279–E280 ([https://doi.org/10.1016/S2666-5247\(21\)00069-0](https://doi.org/10.1016/S2666-5247(21)00069-0)) [accessed 01 Aug 2023]
 - [3] R. B. Brown, Outcome reporting bias in COVID-19 mRNA vaccine clinical trials, *Medicina* 2021, 57 (3), 199 (<https://doi.org/10.3390/medicina57030199>) [accessed 01 Aug 2023]
 - [4] M.E. Oster, D.K. Shay, J.R. Su, et al., Myocarditis cases reported after mRNA-based COVID-19 vaccination in the US from December 2020 to August 2021, *JAMA* 2022, 327(4): 331–340 (<https://jamanetwork.com/journals/jama/fullarticle/2788346>) [accessed 01 Aug 2023]
 - [5] S. McLachlan, M. Osman, K. Dube, P. Chikatero, Y. Choi, N. Fenton, Analysis of COVID-19 vaccine death reports from the Vaccine Adverse Events Reporting System (VAERS) database: Interim results and analysis, Preprint 2021 (<https://www.scienceopen.com/document?vid=2e4c057d-5b64-4c45-ae7c-9507e9dbb74a>) [accessed 01 Aug 2023]
 - [6] S. McLachlan, M. Neil, Y. Choi, C. Craig, J. Engler, K. Dube, M. Osman, N. Fenton, Extended: Analysis of the COVID-19 Vaccine Death Reports from the Vaccine Adverse Events Reporting System

- (VAERS) database, Preprint 2023 (https://www.researchgate.net/publication/367030584_Extended_Analysis_of_COVID-19_Vaccine_Death_Reports_from_the_Vaccine_Adverse_Events_Reporting_System_VAERS_Database) [accessed 01 Aug 2023]
- [7] M. Schemling, V. MAnniche, P.R. Hansen, Batch-dependent safety pf the BNT162b2 mRNA COVID-19 vaccine, *Eur J Clin Invest* 2023, 53, 8, e13998 (<https://onlinelibrary.wiley.com/doi/full/10.1111/eci.13998>) [accessed 01 Aug 2023]
 - [8] Y.B. Yu, M.B. Taraban, K.T. Briggs, All vials are not the same: Potential role of vaccine quality in vaccine adverse reactions *Vaccine* 2013, 39, 45: 6565–6569
 - [9] N.E. Fenton, M. Neil, S. McLachlan, Paradoxes in the reporting of Covid19 vaccine effectiveness: Why current studies (for or against vaccination) cannot be trusted and what we can do about it, Preprint 2021 (https://www.researchgate.net/publication/354601308_Paradoxes_in_the_reporting_of_Covid19_vaccine_effectiveness_Why_current_studies_for_or_against_vaccination_cannot_be_trusted_and_what_we_can_do_about_it) [accessed 01 Aug 2023]
 - [10] T.T. Shimabukuro, M. Nguyen, D. Martin, F. DeStefano, Safety monitoring in the Vaccine Adverse Event Reporting System (VAERS) *Vaccine* 2015, 33 36: 4398–4405 [accessed 01 Aug 2023] (<https://doi.org/10.1016/j.vaccine.2015.07.035>) [accessed 01 Aug 2023]
 - [11] A.M. Pezzullo, C. Axfors, D.G. Contopoulos-Ioannidis, A. Apostolatos, J.P.A. Ioannidis, Age-stratified infection fatality rate of COVID-19 in the non-elderly informed from pre-vaccination national seroprevalence studies, Preprint 2022 (<https://www.medrxiv.org/content/10.1101/2022.10.11.22280963v1>) [accessed 01 Aug 2023]