Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Detecting post-acute COVID-19 vaccination syndrome in pharmacovigilance data: a pre-registered analysis of v-safe free text

Matthew Halma

mhalma@opensourcemed.info

Open Source Medicine Foundation OU https://orcid.org/0000-0003-2487-0636

Research Article

Keywords: post-acute COVID-19 vaccination syndrome, PACVS, pharmacovigilance, v-safe, VAERS, natural language processing, MedDRA, COVID-19 vaccines, pre-registration

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11721193/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare potential competing interests as follows: MH is Director of the Open Source Medicine Foundation and has previously received funding from the Independent Medical Alliance. He has published on PACVS and vaccine-injury recognition. He has no role with, and has received no payment from, OpenVAERS, ICAN or any litigation body, and the language-model services used were paid for at list price.

Detecting post-acute COVID-19 vaccination syndrome in pharmacovigilance data: a pre-registered analysis of v-safe free text

Matthew Halma[1]

[1] Open Source Medicine, Tallinn, Estonia. ORCID: 0000-0003-2487-0636. Corresponding author: Matthew Halma, Director, Open Source Medicine, Pärnu mnt. 139c, 11317 Tallinn, Estonia; mhalma@opensourcemed.info.

Running title: PACVS signature in v-safe free text. Article type: Original Research Article.

Abstract

Introduction. Post-acute COVID-19 vaccination syndrome (PACVS) is a chronic multi-system condition that pharmacovigilance systems code one symptom at a time. CDC's v-safe collected free-text answers at 3, 6 and 12 months after a dose; one prior analysis used its structured answers, none its free text.

Objective. To test, under a pre-registered protocol, whether long-term v-safe free text shows a coherent, persistent and under-captured multi-domain symptom pattern.

Methods. Four questions were registered (osf.io/gnfpd) before any long-term text was read: clustering of nine candidate symptom domains, early onset and persistence, under-capture and dispersed coding in VAERS, and burden. Text was classified by a frozen keyword lexicon plus a language model and validated against CDC's MedDRA coding, with a confirmatory gate of precision and recall of at least 0.85.

Results. Only fatigue passed the gate, so the primary analysis could not be run; every result is the registered nine-domain sensitivity analysis. Of 7,819,251 entries, 1,595,121 were long-term, and of 8.62 million person-doses with a long-term survey, 87,229 (1.01%) met the case definition. Domain pairs co-occurred 1.25 times as often as chance ($p = 0.001$) in every dose and manufacturer stratum and in 1.06 million VAERS reports. Nulls matched on how much each person wrote, or fixing each person-dose's own domain count, removed that aggregate excess but left a core in which cognitive symptoms co-occurred with sleep 2.3 to 2.7 times and with fatigue 2.1 to 2.3 times more than chance, in women and in men. The registered onset test was uninformative (28% early against a schedule-dependent 71% expected). Symptoms recurred at a later survey in 47 to 60% of cases, depending on how cases qualified, against 1.5 to 10.5% of comparison groups. Cases reached VAERS slightly less often than people with named acute diagnoses (rate ratio 0.94, 95% CI 0.90 to 0.98), and their reports were coded to a narrower, not wider, spread of terms.

Conclusion. Of person-doses with a long-term v-safe survey, 1.01% describe a multi-domain pattern whose cognitive, fatigue, sleep and dysautonomia core clusters beyond chance after adjustment for verbosity. When the symptoms began could not be established, and the data cannot attribute the pattern to vaccination rather than to intercurrent illness or selection into

long-term response. The classifier did not meet its pre-registered accuracy standard, although a blind single-coder audit of 306 disagreements found it right in 86% of its unmatched positives against 60% for CDC's coding; the findings are hypothesis-generating.

Key Points

- The text classifier failed its pre-registered accuracy gate for eight of nine domains, so every result is a sensitivity analysis; within it, 1.01% of 8.6 million person-doses with a long-term v-safe survey described a multi-domain symptom pattern meeting a pre-registered PACVS-like definition.
- Cognitive symptoms, fatigue, sleep disturbance and dysautonomia co-occurred 1.4 to 2.7 times more often than chance under every null tried, including one that fixes each person's own number of symptoms; the aggregate excess did not survive those nulls, the core did.
- The registered onset test could not establish when symptoms began, and recorded VAERS filing for these cases was only slightly lower than for named acute events; the findings are hypothesis-generating and call for independent, multi-coder validation of the classifier before any confirmatory claim.

Introduction

Post-acute COVID-19 vaccination syndrome (PACVS) is a multi-system condition following COVID-19 vaccination, with symptoms resembling long COVID: fatigue, cognitive difficulty ("brain fog") and post-exertional malaise among them. Causality has been hard to assign for a multi-system disorder [1, 2], which leaves patients without the clinical entity on which compensation, treatment and research depend [3]. Prevalence estimates range from 0.003% to 0.9% of recipients [2]; the higher figure comes from an occupational survey, the lower from verified reports, which understate the rate. Case series describe a coherent presentation [4–15], but their patients came forward attributing symptoms to vaccination.

Routine surveillance is poorly placed to see such a syndrome. Spontaneous reports are coded one symptom at a time to MedDRA Preferred Terms, so a multi-system pattern is split across many terms, and their denominators are unknown, so reporting can track awareness of the system or the salience of the campaign rather than the rate of events [16]. Active surveillance counts events inside a fixed risk window after the dose, so later problems fall outside it. There is no disease entity to code to: a system can flag fatigue or brain fog, but not a constellation.

v-safe, CDC's smartphone check-in system, followed about 10 million COVID-19 vaccine recipients in the United States for up to a year, soliciting their condition at fixed time points, and has been used to examine preliminary VAERS signals [17]. A followed cohort supplies the denominator that spontaneous reporting lacks, and v-safe asked open questions at 3, 6 and 12 months after a dose; released under FOIA, these answers are the largest body of long-term, participant-written text on post-vaccination health. Levi and Lu [22] analysed v-safe's structured answers: among 9,904,924 enrollees aged 16 or older, 185,285 (1.87%) reported worse health lasting at least three months that they attributed to vaccination, and 1.92% reported prolonged worse health without attributing it. Their phenotype rests on attribution. The present study uses

the free text to define a symptom-based phenotype that does not, tests whether symptom domains cluster against a null model, and examines onset and recorded VAERS filing; the two analyses were independent.

We asked four pre-registered questions:

1. **Coherence (Q1).** Do nine candidate PACVS symptom domains co-occur within people more than their individual frequencies predict?
2. **Natural history (Q2).** Do cases' symptoms begin within six weeks of a dose, and do they persist more than in people who had a severe first week and recovered?
3. **Detection (Q3).** Are cases less likely than people with named acute diagnoses to reach VAERS, and are their VAERS reports coded to a wider spread of terms?
4. **Burden (Q4).** How often do cases report health impact, medical care and call-centre contact? (Descriptive.)

Methods

Data sources

Table 1. Data sources, content and use.

Source	Content	Use
v-safe free text (FOIA 24-01322, original and updated releases, under-3 file) [19]	7,819,251 entries after de-duplication	Classification
v-safe registrants, vaccinations, health check-ins [19]	145.0 million check-ins; dose dates, manufacturer, sex (no age field was present in the linked release)	Dose anchor, survey type, health answers
CDC MedDRA coding of the free text (November 2024 release) [19, 21]	Preferred Terms per entry; 99.9% of entries coded	Reference standard
v-safe call centre (FOIA 23-00925, 23-00926) [19]	Cases, care, VAERS-filed dispositions	Q3, Q4
OpenVSafe person pages [18]	Best-matching VAERS report per VAERS-filed person, score out of 11	Q3
VAERS public data (dated 2026-06-03) [20]	Reports, MedDRA terms, symptom text	Q3, Q1 replication

Each entry was linked to its registrant, the check-in it answered, the most recent preceding dose, CDC's MedDRA codes and call-centre outcomes within 90 days (Table 1); entries redacted under FOIA exemption (b)[6] were excluded. OpenVSafe match data were read from its public per-registrant pages [18] on 2026-11-10 and 2026-11-11 for the 44,954 people whom the call-centre data marked as VAERS-filed, at one request per second and within the access the site's robots.txt allows; no bulk file exists, and no permission beyond the site's published access rules was sought. Only the VAERS ID, match score, confidence and runner-up score were kept.

Pre-registration

The protocol was registered on OSF (osf.io/gnfpd) on 2026-11-11, before any 3-, 6- or 12-month text was classified or read. Code enforced this: long-term text could not be classified until the registration link was set. The taxonomy (v1.0) was frozen on 2026-11-10. The closest prior work, Levi and Lu's analysis of v-safe's structured answers [22], was identified on 12 November 2026, after registration and after the registered analyses had been run; it did not change the registered plan, and no amendment was filed for it. Two amendments were filed before the results they govern were computed:

- **Amendment 1** replaced the planned two-coder gold set, for which no coders were available, with CDC's MedDRA coding as the reference standard.
- **Amendment 2** replaced the registered language model (GLM-5.3) with DeepSeek V4.1 Flash, because the first model's quota cost made the full run infeasible.

Classification

Layer 1 is a keyword lexicon for 21 themes, including the nine PACVS domains: fatigue and post-exertional malaise, cognitive, dysautonomia, sensory-neuropathic, sleep, pain, gastrointestinal, menstrual and cardiorespiratory. It also tags three exclusion themes (named acute diagnosis, unrelated cause, upper-respiratory infection) and flags entries about the time before vaccination. It was run on all entries.

Layer 3 is a language model (deepseek/deepseek-v4.1-flash, reasoning off, prompt v0.1) given the same closed theme list. It labelled every unique long-term text to which layer 1 gave no theme: 649,516 entries, 386,862 unique texts. Requests went through a zero-data-retention route; texts containing a detectable e-mail address or phone number were not sent. The final classifier is the union of the two layers.

Validation and the confirmatory gate

An entry was positive for a domain in the reference if CDC coded it to at least one Preferred Term listed for that domain in the frozen map. All long-term entries with a CDC code were evaluated, excluding redacted and pre-vaccine entries. We computed precision and recall with 95% Wilson intervals. A domain was confirmatory if both were at least 0.85 for the final classifier. As registered, if any domain failed, Q1 to Q3 were to be run on passing domains (primary) and on all nine (sensitivity).

Case definition and comparison groups

The unit is a person-dose. A **case** meets three criteria:

1. At a 3-, 6- or 12-month survey S, the person reported health worse than before vaccination, or a health impact (unable to do normal activities, missed work or school, or needed care).
2. Free text at S names at least two PACVS domains, or free text across at least two long-term surveys names at least two between them.
3. Free text at S names no exclusion theme.

Three mutually exclusive comparison groups were defined, assigned in the priority case, (b), (c), (a):

- **(a) No symptoms:** no long-term survey worse or with impact, and no PACVS text.
- **(b) Acute event:** free text at any survey of that dose names an acute diagnosis.
- **(c) Severe week 1, recovered:** an impact or at least two domains in week 1, then no long-term survey worse or with impact.

Statistical analysis

Q1. For person-doses with long-term text we built a person-dose by domain matrix. The statistic is the sum, over the 36 domain pairs, of the people having both. The null permutes each domain column independently 1,000 times (seed 20261110), keeping each domain's frequency; the one-sided p-value is (1 + number of null values at least as large as the observed) / 1,001, and per-pair p-values were Holm-adjusted. The test was repeated within each dose and manufacturer with at least 1,000 person-doses, and on VAERS COVID-19 report text classified by layer 1.

Q2. Onset is the first survey type (week 1, six weeks, 3, 6 or 12 months) whose text names any of the case's long-term domains. Each case's null probability of early onset is the share of their completed check-ins for that dose that fell in week 1 or the six-week period; the number of early-onset cases was compared with this Poisson-binomial null by an exact one-sided test. Persistence compares cases with group (c) on the share naming a PACVS domain at a later long-term survey, among those who completed one, as a risk difference with a Newcombe interval and a one-sided z-test.

Q3. Analysis was per person: a case if any dose was a case, otherwise group (b) if any dose was. A person reached VAERS if a call-centre contact within 90 days of an entry recorded a VAERS filing, or if OpenVSafe [18], a project of OpenVAERS that matches v-safe records to VAERS reports on timing, dose and shared symptoms, matched them to a report (score at least 8 of 11; reports matched to more than one person dropped; main set 25,913 people, high-confidence set 23,925). We call this recorded VAERS filing; a report filed without call-centre contact and not matched by OpenVSafe is not counted. Filing was compared as a rate ratio with a log-scale Wald interval. Dispersion compared the Shannon entropy of matched reports' MedDRA terms between groups by a 1,000-fold permutation of group labels.

Q4 and attrition. Q4 gives rates per 1,000 for health impact, care and call-centre cases by group. Case prevalence was bounded by assuming every person-dose without a long-term survey was well (lower bound) or a case (upper bound).

Disclosure. Only aggregates were exported. Any cell with fewer than 10 people was suppressed, and nothing was reported below the national level.

Decisions where the registration was silent

- Survey types 'day 0' to 'day 7' count as week 1 and 'week 2' to 'week 6' as six weeks; impact is any health-impact answer other than 'N/A' or blank; entries flagged as pre-vaccine by either layer were dropped.
- The registered layer 2 (embedding classifier) was not built; the final classifier is layers 1 and 3. Layer 3 labelled long-term text only, so week-1 and six-week text was classified by layer 1 alone.
- Persistence uses each case's first worse-or-impact long-term survey, and group (c)'s first long-term survey, as the reference point; Q3 is run per person, because VAERS outcomes attach to people.
- Q1's permutation was re-implemented after registration for speed (each domain's positive rows drawn directly), which gives the same null distribution.

Model-sensitivity check

As registered, a second model, claude-opus-5-5 through the Anthropic API, labelled a fixed sample of long-term text: 2,000 long-term residue entries drawn with seed 1 (the pilot's sample size, since the registration fixed none), 1,687 unique texts. A theme is model-sensitive if its entry count differs by more than 20% between the two models on the texts both labelled.

Analyses not in the pre-registration

These were specified after the registered results were seen, each for the reason given, and are reported as exploratory. They do not replace any registered result.

1. Text-matched null (Q1): the registered null ignores how much each person-dose wrote, so each domain's positives are placed at random within deciles of long-term characters, and within entries (1, 2, 3, 4 or more) crossed with deciles.
2. Curveball null (Q1): the text-matched null still lets a person-dose's own domain count vary. Curveball trades keep every person-dose's domain count and every domain's frequency (30 burn-in rounds, 3 between samples, 1,000 samples, seed 20261110), unstratified and within entries and text-length tertiles; the total pair count is then fixed, so only the 36 pairs are tested, with upper- and lower-tail p-values.
3. Core-subgraph statistic and Q1 within each sex, to locate the excess and to check it in each sex: co-occurrences summed over the six pairs among cognitive symptoms, fatigue, sleep and dysautonomia and over the other 30 pairs, under each null, with no test attached; and the registered, text-matched and curveball nulls within women and within men, and the registered null within sex × dose strata.
4. Exclusion criterion variants: criterion 3 and group (b) from keyword tags only, and criterion 3 dropped, because the exclusion tags had low precision; and Q1 with redacted entries included.
5. Persistence by qualification route (cases meeting criterion 2 at a single survey) and against group (d), long-term responders worse or impacted who named fewer than two domains, excluding cases and acute events.

6. Onset re-examined: layer 3 also run on cases' early residue (32,212 entries) with a second null uniform over completed survey types; and week-1 check-box symptoms and severity for fatigue, pain and gastrointestinal symptoms among cases with the domain, group (a) and group (c).
7. Validation audit: 306 long-term entries where layer 3 and CDC disagree on a domain (17 per domain per side, seed 20261112), coded by the author blind to side.
8. Dispersion with Miller-Madow correction, rarefaction to equal code counts and a two-sided test.
9. Levi and Lu's cohorts rebuilt from the structured answers by their rules (no age field exists, so the adult file stands in for 16 or older) as a linkage check, and cross-tabulated with the text-defined case with Cohen's kappa.

Results

Corpus

The twelve updated FOIA releases and the under-3 file hold 7,819,275 rows, 7,819,251 after de-duplication; the release description gives 7,554,929 entries, and the difference is unexplained. Of the 7,819,251 entries, 1,595,121 came from long-term surveys, written by 1,207,426 people: 565,626 at 3 months, 586,468 at 6 months and 443,027 at 12 months. A further 5,791,000 came from week-1 and six-week surveys, and 433,130 could not be matched to a survey. The 144,972,101 check-ins are the 144,856,043 rows of the adult file and 116,294 of the under-3 file in our download, de-duplicated by response ID; Levi and Lu report 151,581,805 adult rows. Layer 3 labelled all 386,862 unique long-term residue texts with no failed requests after retry; logged provider charges were US$0.022 for that run and US$0.0005 for later runs, and the main run's token count was not logged.

Classifier validation: one domain passed

Only fatigue passed the gate (precision 0.961, recall 0.859; Figure 1). Layer 3 raised recall in every domain but lowered precision. Dysautonomia, which passed on keywords alone (0.978 / 0.872), failed on precision once layer 3 was added (0.774). Eight domains are therefore exploratory, and the registered primary analysis, which needs at least two passing domains, could not be run. Every result below is the registered nine-domain sensitivity analysis.

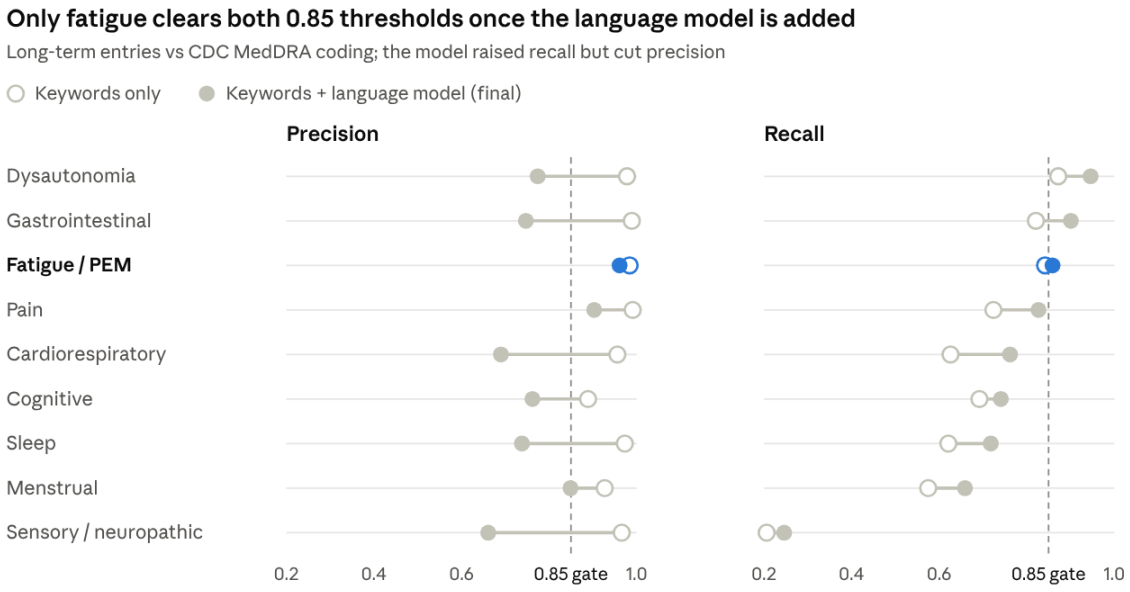


Figure 1. Precision and recall of each PACVS domain against CDC's MedDRA coding of long-term entries, for keywords alone (open circles) and for the final classifier, keywords plus language model (filled). The dashed line marks the pre-registered 0.85 gate; a domain passes only if both filled dots lie to its right. Only fatigue / post-exertional malaise passed.

The exclusion themes also had low precision against the reference: unrelated cause 0.144, acute diagnosis 0.311. A post-hoc look at the CDC terms on these entries suggests most of the shortfall comes from gaps in the frozen map rather than from wrong tags. The commonest terms were COVID-19, a positive SARS-CoV-2 test, nasopharyngitis, cerebrovascular accident and thrombosis, none of which are in the map. Two real errors were seen: entries mentioning a negative COVID-19 test were tagged as an unrelated cause, and chronic conditions were tagged as acute diagnoses.

Case and comparison groups

Table 2. Case and comparison groups: person-doses, people and long-term response.

Group	Person-doses	People	Share of person-doses with a long-term survey
Case	87,229	86,323	1.01%
(a) No symptoms	5,877,624	4,760,357	68.2%
(b) Acute event	104,736	102,616	1.08% (93,028 with a long-term survey)
(c) Severe week 1, recovered	1,602,523	1,518,229	18.6%
Other	2,267,821	2,036,216	11.2% (963,254 with a long-term survey)

Of 17.46 million person-doses with any check-in, 49.4% completed a long-term survey; response was higher after a week-1 health impact (60.1%) than without one (42.4%). The five groups in Table 2 cover 9,939,933 person-doses; the other 7.52 million completed no long-term survey and had neither an acute-diagnosis entry nor a severe week 1. Case prevalence among responders was 1.01%, with bounds of 0.50% (all non-responders well) to 51.1% (all non-responders cases).

Characteristics of the groups

Cases were more often women and more often second-dose person-doses than the no-symptoms group (Table 3); manufacturer shares were similar across groups. Age was not present in the linked release.

Table 3. Sex, manufacturer and dose by group, % of person-doses.

	Case	(a) No symptoms	(b) Acute event	(c) Severe week 1, recovered
Female	77.5	58.1	68.9	72.8
Male	21.3	41.1	30.3	26.2
Other or unknown sex	1.2	0.8	0.8	1.0
Pfizer-BioNTech	48.4	53.2	48.3	39.5
Moderna	45.9	43.4	47.3	55.8
Janssen	5.7	3.3	4.3	4.7
Dose 1	14.3	13.7	18.3	9.6
Dose 2	69.7	60.6	60.6	79.5
Dose 3	14.2	20.2	16.3	9.6
Dose 4 or later	1.8	5.6	4.8	1.4

Q1: the domains cluster

Among 1,267,509 person-doses with long-term text, domain pairs co-occurred 235,184 times against 188,776 expected under the null (ratio 1.25, p = 0.001). The test rejected in all seven strata with at least 1,000 person-doses: doses 1 to 4 (ratios 1.21 to 1.39) and Pfizer-BioNTech, Moderna and Janssen (1.22 to 1.26). This meets the registered criterion for reproducibility, but at these sample sizes any real dependence reaches p = 0.001, so the ratios carry the information, and the registered null does not account for how much each person wrote (see Analyses not in the pre-registration).

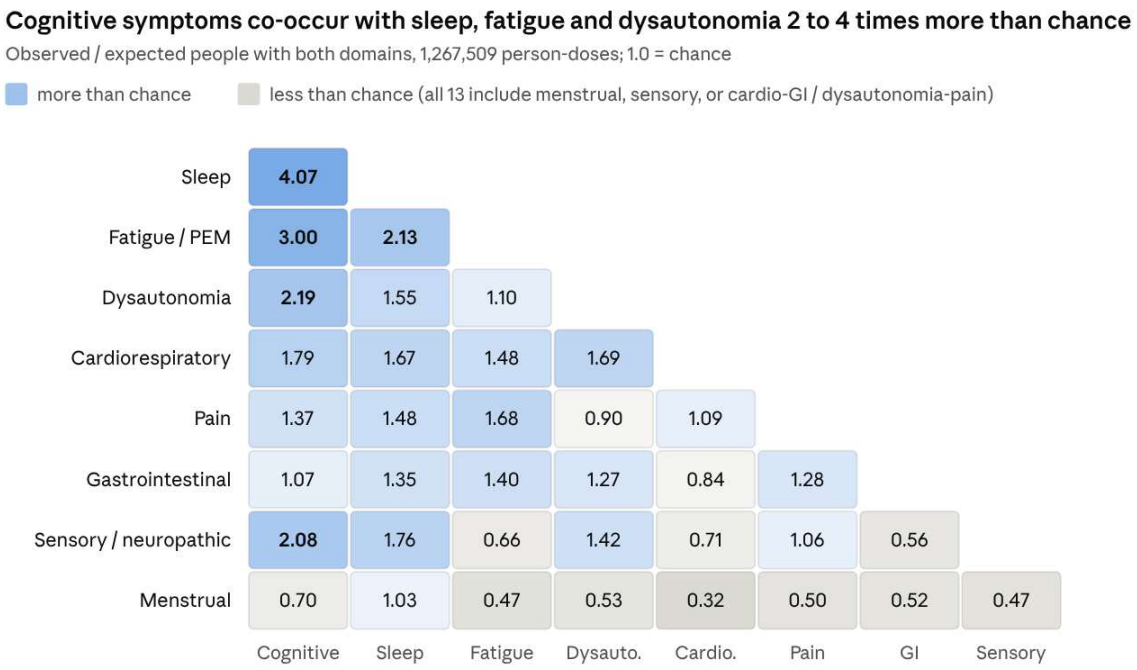


Figure 2. Q1 under the registered null: observed / expected number of person-doses naming both domains, for the 36 domain pairs among 1,267,509 person-doses with long-term text. Values above 1 indicate co-occurrence beyond chance.

Twenty-three of the 36 pairs exceeded the null after Holm adjustment (adjusted p = 0.036; Figure 2). The strongest ties were cognitive symptoms with sleep (4.1 times expected), fatigue (3.0) and dysautonomia (2.2). The other 13 pairs were at or below chance: all eight involving menstrual symptoms, sensory symptoms with fatigue, gastrointestinal or cardiorespiratory symptoms, cardiorespiratory with gastrointestinal symptoms, and dysautonomia with pain.

Replication in VAERS. We applied the keyword classifier to the symptom text of 1,057,077 VAERS reports for COVID-19 vaccines (2020 to 2026), one row per report. Pairs co-occurred 486,503 times against 262,471 expected (ratio 1.85, p = 0.001). Thirty-one of 36 pairs exceeded the null after Holm adjustment. As in v-safe, the strongest tie was cognitive symptoms with sleep (4.2 times expected), followed by cognitive symptoms with dysautonomia (2.9) and fatigue (2.8). Menstrual symptoms again did not cluster with fatigue, pain or sensory symptoms. VAERS reports mostly describe acute reactions, in which several symptoms often appear together, so this replication shows the same structure but not its timing.

Q2: onset test uninformative, strong persistence

Onset. Of 87,217 cases with an onset survey, 24,316 (27.9%) first described their case domains in week 1 or by six weeks, against 62,192 (71.3%) expected under the null, so H0 is not rejected (p = 1.0). Most (72%) first described them at 3 months or later, and 3 months was the commonest first survey (34%; Figure 3).

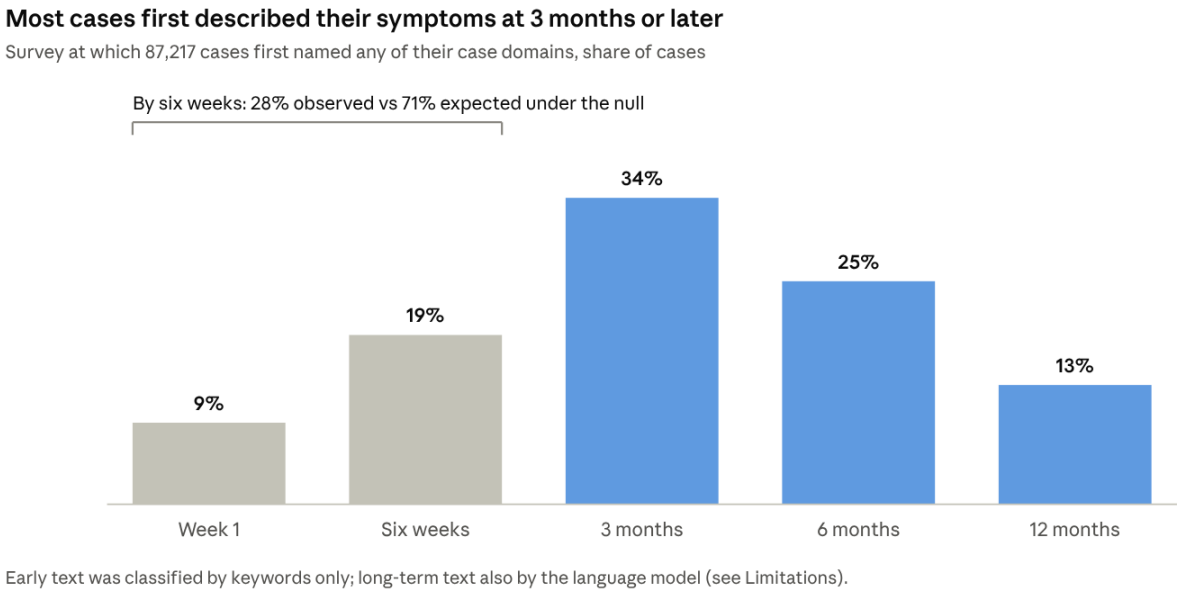


Figure 3. Survey at which each of 87,217 cases first named any of their case domains. By six weeks 28% had done so, against 71% expected under the registered null. Early text was classified by keywords only, long-term text also by the language model.

Persistence. Among those who completed a later long-term survey, 60.3% of cases (30,907 of 51,221) named a PACVS domain again, against 1.5% of group (c) (17,416 of 1,169,538). The risk difference was 58.9 percentage points (95% CI 58.4 to 59.3, p < 0.001).

Q3: recorded VAERS filing slightly lower for cases, coding not more dispersed

Table 4. Recorded VAERS filing and dispersion of MedDRA Preferred Terms on matched reports, cases against group (b), per person.

Match set	Cases with a recorded VAERS filing	Group (b) with a recorded VAERS filing	Rate ratio (95% CI)	PT entropy, cases vs (b)	Dispersion p
Main	3,937 / 86,323 (4.56%)	4,944 / 101,690 (4.86%)	0.94 (0.90 to 0.98)	8.31 vs 8.59	1.0
High confidence	3,925 / 86,323 (4.55%)	4,921 / 101,690 (4.84%)	0.94 (0.90 to 0.98)	8.31 vs 8.61	1.0

Cases reached VAERS about 6% less often than people with named acute diagnoses (4.6% against 4.9%; Table 4); the upper confidence limit is below 1, so H0 is rejected. Group (b) here has 101,690 people rather than 102,616 because 926 people who were a case at one dose and an acute event at another count as cases. On matched reports of 2,122 cases and 2,642 group (b) people, cases' reports were coded to a narrower, not wider, spread of Preferred Terms (entropy 8.31 against 8.59 bits), so H0 is not rejected; small-sample corrections did not change this (below).

Q4: burden (descriptive)

Table 5. Health impact, care and call-centre cases at long-term surveys by group, per 1,000.

Group	Health impact at a long-term survey, per 1,000	Care at a long-term survey, per 1,000	Call-centre case, per 1,000
Case	952.7	462.6	587.2
(b) Acute event	680.7	518.1	771.4
(a) No symptoms	0	0	19.5
(c) Severe week 1, recovered	0	0	54.5

Nearly all cases (952.7 per 1,000) reported a health impact at a long-term survey and 462.6 per 1,000 sought care, against 680.7 and 518.1 in group (b); 587.2 per 1,000 cases and 771.4 per 1,000 in group (b) had a call-centre case (Table 5). Zero rates in groups (a) and (c) follow from their definitions.

Model sensitivity (registered check)

On the fixed sample of 2,000 long-term residue entries (1,687 unique texts, all labelled by both models without failures or refusals), four of nine PACVS domains were model-sensitive by the registered 20% rule: fatigue, menstrual, cardiorespiratory and cognitive symptoms (Table 6). Sleep had too few entries to report; dysautonomia, gastrointestinal, pain and sensory-neuropathic symptoms were stable. Among the other themes, unrelated cause, persistence and care sought were flagged, each applied more often by Claude Opus. Counts per domain are small on residue text, where the keyword layer found nothing, and the shift statistic is noisy below about 50 entries, but the direction for fatigue is clear: DeepSeek labelled it less than half as often as Claude. A pilot on 2,000 random early entries had found every PACVS domain within 13% between the two models, so the disagreement is specific to the harder long-term residue.

Table 6. Entries labelled with each theme by the two models on the fixed long-term residue sample.

Theme	DeepSeek V4.1 Flash	Claude Opus 5.5	Shift	Model-sensitive
Fatigue / PEM	20	46	0.57	yes
Menstrual	18	10	0.44	yes
Cardiorespiratory	68	41	0.40	yes
Cognitive	13	10	0.23	yes
Sensory / neuropathic	57	50	0.12	no
Pain	205	224	0.09	no
Dysautonomia	52	54	0.04	no
Gastrointestinal	105	107	0.02	no
Sleep	<10	<10	suppressed	

Theme	DeepSeek V4.1 Flash	Claude Opus 5.5	Shift	Model-sensitive
Unrelated cause (exclusion)	429	621	0.31	yes
Acute diagnosis (exclusion)	131	112	0.15	no
Upper respiratory (exclusion)	232	229	0.01	no
Persistence	52	102	0.49	yes
Care sought	404	553	0.27	yes
Pre-existing worsened	49	61	0.20	no

Analyses not in the pre-registration

Text-matched null for Q1. Mean domains per person-dose rose from 0.26 in the lowest decile of long-term characters to 1.07 in the highest. Matching the null on amount of text cut the ratio of observed to expected pairs from 1.25 to 1.06 (character deciles) and 1.04 (entries crossed with deciles), still p = 0.001 at this sample size (Table 7).

Table 7. Q1 under the registered null and two text-matched nulls.

Null	Strata	Observed pairs	Expected pairs	Ratio	Pairs above null (Holm)
Registered (domain frequency only)	1	235,184	188,776	1.25	23 of 36
Matched on character deciles	10	235,184	221,060	1.06	19 of 36
Matched on entries × character deciles	39	235,184	226,337	1.04	16 of 36

The specific structure survived (Figure 4): under the strictest matching, cognitive symptoms co-occurred with sleep 2.5 times, with fatigue 2.3 and with dysautonomia 1.6 times more than expected, and fatigue with sleep 1.6 and with pain 1.4; pairs that had been modestly above chance moved close to it, and the 13 pairs below chance stayed below.

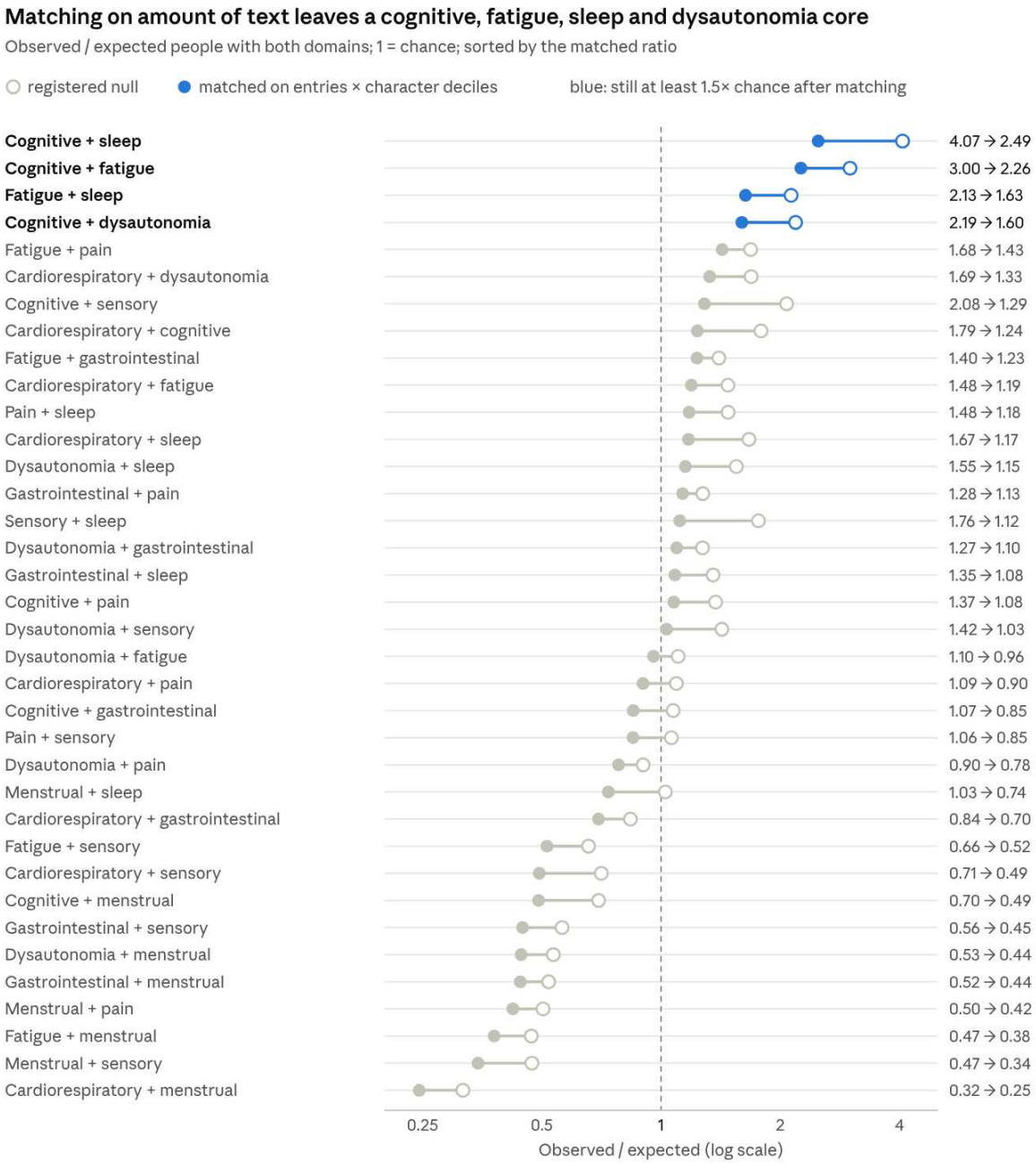


Figure 4. Observed / expected co-occurrence for the 36 domain pairs under the registered null (open circles) and under a null matched on entries × character deciles (filled), log scale. Pairs in blue remain at least 1.5 times chance after matching. Not pre-registered.

Curveball null. With both margins fixed the total pair count is invariant (235,184 in every sample), so only the 36 pairs can be tested (Table 8). The same core stood out under all three stratifications: cognitive symptoms with sleep 2.3 to 2.7 times expected, with fatigue 2.1 to 2.3 and with dysautonomia 1.4 to 1.5, and fatigue with sleep 1.5 to 1.6. Thirteen to fifteen pairs lay

above the null after Holm adjustment and seventeen to nineteen below it, every menstrual pair among the latter.

Table 8. Q1 domain pairs under the curveball null (both margins fixed), observed / expected.

Pair	Unstratified	Within entries 1 / 2 / 3+	Within text-length tertiles
Cognitive + sleep	2.70	2.44	2.30
Cognitive + fatigue	2.27	2.29	2.06
Fatigue + sleep	1.60	1.62	1.47
Cognitive + dysautonomia	1.55	1.42	1.53
Fatigue + pain	1.45	1.48	1.39
Cognitive + sensory	1.43	1.23	1.25
Cardiorespiratory + dysautonomia	1.22	1.16 or less	1.29
Pairs above null after Holm	15	13	15
Pairs below null after Holm	18	19	17
Fatigue + menstrual	0.36	0.38	0.35
Cardiorespiratory + menstrual	0.22	0.23	0.22

Curveball samples: 1,000 after 30 burn-in rounds, 3 rounds between samples, seed 20261110. The full 36-pair table is in the supplement.

Core subgraph. Summed over the six core pairs, the ratio of observed to expected co-occurrences fell from 1.58 under the registered null to 1.30 to 1.32 under the text-matched nulls and 1.21 to 1.22 under the curveball null; for the other 30 pairs it fell from 1.22 to 1.02 to 1.04 and then 0.98 (Table 9). The excess that survives the strict nulls is confined to the core. In VAERS the core ratio was 1.95 against 1.84 for the other pairs.

Table 9. Co-occurrences summed over the six core pairs and over the other 30 pairs, observed / expected (ratio), under each null.

Null	Core pairs	Other pairs
Registered (domain frequency only)	22,587 / 14,251 (1.58)	212,597 / 174,533 (1.22)
Matched on character deciles	22,587 / 17,083 (1.32)	212,597 / 203,952 (1.04)
Matched on entries × character deciles	22,587 / 17,345 (1.30)	212,597 / 209,012 (1.02)
Curveball, unstratified	22,587 / 18,624 (1.21)	212,597 / 216,560 (0.98)
Curveball within entries 1 / 2 / 3+	22,587 / 18,573 (1.22)	212,597 / 216,611 (0.98)
Curveball within text-length tertiles	22,587 / 18,654 (1.21)	212,597 / 216,530 (0.98)
VAERS replication, registered null	46,291 / 23,769 (1.95)	440,212 / 238,707 (1.84)

Computed from the pair tables (supplement S2, S3, S5, S7); no test is attached to the sums.

Q1 within each sex. The registered test rejected in 884,746 women and 370,519 men with long-term text (ratios 1.22 and 1.27) and in every sex × dose stratum (1.19 to 1.40, all p = 0.001); text-matching brought the ratios to 1.03 and 1.06. Under the curveball null the core was the same in both sexes: cognitive symptoms co-occurred with sleep 2.8 and 2.5 times expected in women and men, with fatigue 2.3 and 2.2 and with dysautonomia 1.6 and 1.4, all above the null after Holm adjustment; the six core pairs summed to 1.24 and 1.14, against 0.98 and 0.99 for the other pairs (supplement S29 to S31).

Exclusion criterion variants. Criterion 3 removed about a third of would-be cases (135,252 without it against 87,229). Keyword-only exclusions changed the case count little (89,198) but shrank group (b) from 104,736 to 64,629, because most acute-diagnosis tags on the residue came from the language model and often named chronic conditions. Q1 was unaffected and Q2 held (early onset 27.7% and 23.3%; persistence 60.4% and 52.3% against 1.5%). Recorded VAERS filing fell to rate ratios of 0.82 (95% CI 0.79 to 0.86) and 0.78 (0.75 to 0.81), because the keyword-tagged group (b) filed more often (5.6%); dispersion was never higher for cases (supplement S10 to S13, S18). Including redacted entries among the Q1 rows (1,269,337 person-doses) left the ratio at 1.247.

Onset re-examined. Applying the language model to cases' early residue (32,212 entries) raised the early share from 27.9% to 32.7%, still below the check-in null (71.3%) and a null spread evenly over completed survey types (42.2%; p = 1.0 for both). Check-boxes measure something else (Table 10): 80.9% of the 68,386 cases with a week-1 check-in and a fatigue, pain or gastrointestinal case domain had ticked a symptom in that domain, against 24.2% whose week-1 text named it, but so had 89 to 92% of the severe-then-recovered group, whose persistence was 1.5%, and 35 to 40% of the no-symptoms group. Cases with fatigue had ticked week-1 fatigue in 78.9% (64.4% moderate or severe) against 35.5% (14.6%) of group (a), with similar gaps for pain and gastrointestinal symptoms (supplement S15). The check-boxes record the common acute reaction, not the start of a persistent syndrome: they separate cases from the well, not from the recovered.

Table 10. Onset of cases' domains under four readings.

Measure	Early (week 1 or six weeks)	Comparison
Registered: first text mention, layer 1 on early text	24,316 of 87,217 (27.9%)	check-in null 71.3%; survey-type null 42.2%
Symmetric: language model also on cases' early residue	28,490 of 87,217 (32.7%)	same nulls; p = 1.0
Week-1 text names a fatigue, pain or gastrointestinal case domain	19,833 of 81,826 (24.2%)	
Week-1 checkbox in the same domain	55,342 of 68,386 with a week-1 check-in (80.9%)	no-symptoms group 35 to 40%; severe-then-recovered 89 to 92%

By-group check-box rates use all cases and a 1-in-20 sample of other person-doses.

Persistence by qualification route and against group (d). Of case person-doses, 69,816 (80.0%) met criterion 2 at a single survey and 17,413 only across surveys. Persistence was 47.1% for same-survey cases against 60.3% for all, both against 1.5% in group (c); group (d) had 10.5%. The risk difference is 58.9 points for the registered comparison and 36.6 (95% CI 36.1 to 37.1) for same-survey cases against group (d) (Table 11).

Table 11. Persistence of a PACVS domain at a later long-term survey, by case route and comparison group.

Comparison	Cases	Comparison group	Risk difference, points (95% CI)
Registered: all cases vs group (c)	30,907 / 51,221 (60.3%)	17,416 / 1,169,538 (1.5%)	58.9 (58.4 to 59.3)
Same-survey cases vs group (c)	17,177 / 36,478 (47.1%)	17,416 / 1,169,538 (1.5%)	45.6 (45.1 to 46.1)
All cases vs group (d)	30,907 / 51,221 (60.3%)	32,391 / 308,869 (10.5%)	49.9 (49.4 to 50.3)
Same-survey cases vs group (d)	17,177 / 36,478 (47.1%)	32,391 / 308,869 (10.5%)	36.6 (36.1 to 37.1)

Validation audit. Layer 3 and CDC coding disagreed on a PACVS domain for 223,652 entry-domain pairs where only CDC was positive and 111,523 where only layer 3 was; a seeded sample of 306 (17 per domain per side) was coded by the author, blind to side but not to the hypotheses. Where only the classifier was positive, the coder agreed with it in 131 of 153 entries (85.6%); where only CDC was positive, with CDC in 92 of 153 (60.1%) (Table 12). The classifier's unmatched positives were right in every sampled dysautonomia and sensory-neuropathic entry; CDC's unmatched codes were right least often for fatigue (6 of 17) and menstrual symptoms (2 of 17). Most of what the validation counted as false positives were symptoms CDC had not coded, so measured precision understates the classifier for several domains; the registered gate is applied as registered.

Table 12. Blind audit of 306 classifier-versus-CDC disagreements: entries in which the coder agreed with the side that had called the domain present (n of 17 per cell).

Domain	Only the classifier positive: classifier right	Only CDC positive: CDC right
Dysautonomia	17	10
Sensory / neuropathic	17	13
Cardiorespiratory	16	14
Gastrointestinal	16	14
Cognitive	15	9
Pain	15	11
Sleep	15	13
Fatigue / PEM	14	6
Menstrual	6	2

Domain	Only the classifier positive: classifier right	Only CDC positive: CDC right
All domains	131 of 153 (85.6%)	92 of 153 (60.1%)

Dispersion with small-sample corrections. Group (b) contributed more codes (19,063) than cases (18,072). Neither the Miller-Madow correction nor rarefaction to 18,072 codes changed the direction, the two-sided permutation test rejected equality (Table 13), and the high-confidence match set gave the same picture (raw 8.31 against 8.61; p = 0.001).

Table 13. Shannon entropy (bits) of MedDRA Preferred Terms on matched VAERS reports, main match set.

Measure	Cases (2,122 people, 18,072 codes)	Group (b) (2,642 people, 19,063 codes)
Raw	8.31	8.59
Miller-Madow corrected	8.38	8.67
Rarefied to 18,072 codes (mean of 200 draws)	8.31	8.59
Difference, raw	-0.28	
p, one-sided (cases more dispersed)	1.0	
p, two-sided	0.001	

Reproduction of Levi and Lu's cohorts. Rebuilt from the structured answers alone, their dose-1 and dose-2 Prolonged-AE counts agree within 0.5% (32,732 against 32,838; 101,064 against 100,564), which supports the linkage. The shortfalls of 3.6% in enrollees, 9.7% in Prolonged-AE and 11.6% in Prolonged-NoAttr sit almost entirely in the dose-3-or-more sub-cohorts (35% fewer in each), consistent with our check-in file holding 6.7 million fewer rows than theirs; the released files carry no age field, so their 16-or-older restriction could not be applied (supplement S26). Among 6,754,831 people in both denominators, 29,593 were both a text-defined case and in Prolonged-AE, 56,730 text-defined only and 137,742 Prolonged-AE only: 34.3% of text-defined cases attributed worse health to vaccination on a check-box, and 17.7% of Prolonged-AE met the text definition (Cohen's kappa 0.22; S27).

Symptom frequencies among cases, compared with published case series

Among the 87,229 case person-doses, pain was the most often named domain (70.2%), then fatigue (51.5%), gastrointestinal symptoms (33.1%) and dysautonomia (25.1%); cognitive (8.5%), menstrual (5.3%) and sleep (4.0%) symptoms were named least. Cases named a median of 2 domains (mean 2.29). Table 14 sets these beside published PACVS series [4–6, 10, 12, 13], using each paper's own symptom label.

Table 14. Share of patients with each symptom domain: this study and published PACVS case series, %.

Domain	This study, v-safe cases, free text (n = 87,229 person-doses)	Mundorf 2024, online checklist (n = 191) [5]	Krumholz 2023, online checklist (n = 241) [6]	Dinser 2026, CCC questionnaire (n = 26) [4]	Gerhard 2023, neurology clinic (n = 50) [12]	Tokumasu 2023, clinic records (n = 121) [10]
Fatigue / PEM	51.5	exhaustion 84.8; PEM 70.7	excessive fatigue 69; exercise intolerance 71	new exhaustion 76.9; deterioration with exertion 96.2	fatigue 46	general fatigue 24.8
Cognitive	8.5	brain fog 76.4	brain fog 63	concentration impairment 88.5	cognitive impairment 36	not reported
Dysautonomia	25.1	dizziness 80.1; perceptible heartbeat 72.8	palpitations 60; dizziness 50	positional dizziness 65.4; POTS 50.0	vertigo 16	dizziness 4.1
Sensory / neuropathic	14.9	paresthesia 79.6; numbness 62.8	numbness 63; neuropathy 63	perceptual and sensory disorders 46.2	paraesthesia 56	sensory impairment 28.1
Sleep	4.0	unrestful sleep 80.6; insomnia 36.1	insomnia 61	sleep disturbances 57.7	not reported	not reported
Pain	70.2	muscle pain 80.6; headache 63.4	myalgia 55; headache 53	headache 65.4; muscle pain 61.5	headache 30; myalgia 22	headache 17.4
Gastrointestinal	33.1	nausea 52.4; diarrhoea 40.3	abdominal pain 33; nausea or vomiting 32	intestinal dysfunction 65.4	not reported	nausea or vomiting 5.8
Menstrual	5.3	abnormal cycle 42.9	irregular cycles 20	not reported	not reported	not reported
Cardiorespiratory	16.3	exertional dyspnoea 59.7; angina 46.4	shortness of breath 37; chest pain 26 to 28	dyspnoea 38.5	not reported	chest discomfort 3.3

Percentages are of each series' whole sample; menstrual items in this study and in [5, 6] use denominators that include men. Finsterer and Scorza [13] describe three patients and are not tabulated. Symptoms were elicited by checklists of 38 to 96 items [4–6], clinician-recorded complaints [10, 12] or, here, unprompted free text at 3 to 12 months.

The rank order overlaps but the levels do not: checklist series report most domains in 60 to 90% of patients, clinician-recorded series in 4 to 56%, and v-safe free text sits between for fatigue,

pain and gastrointestinal symptoms and well below both for cognitive and sleep symptoms. The 8.5% naming cognitive symptoms, against 36 to 88% elsewhere, reflects the classifier's recall for that domain (0.74), what people name unprompted, and a question about health after a dose rather than about a syndrome.

Discussion

The pre-registered confirmatory analysis could not be run, because only one of nine domains met the accuracy gate. In the registered sensitivity analysis three of five hypotheses were supported (Figure 5): symptoms cluster (Q1), they persist (Q2), and cases reach VAERS less often than people with named acute diagnoses (Q3). Early onset (Q2) and dispersed VAERS coding (Q3) were not supported, and for both the observed value lay on the other side of the null. The supported effects differ greatly in size: clustering and persistence are large, while under-capture is a 6% relative difference between two rates near 5%.

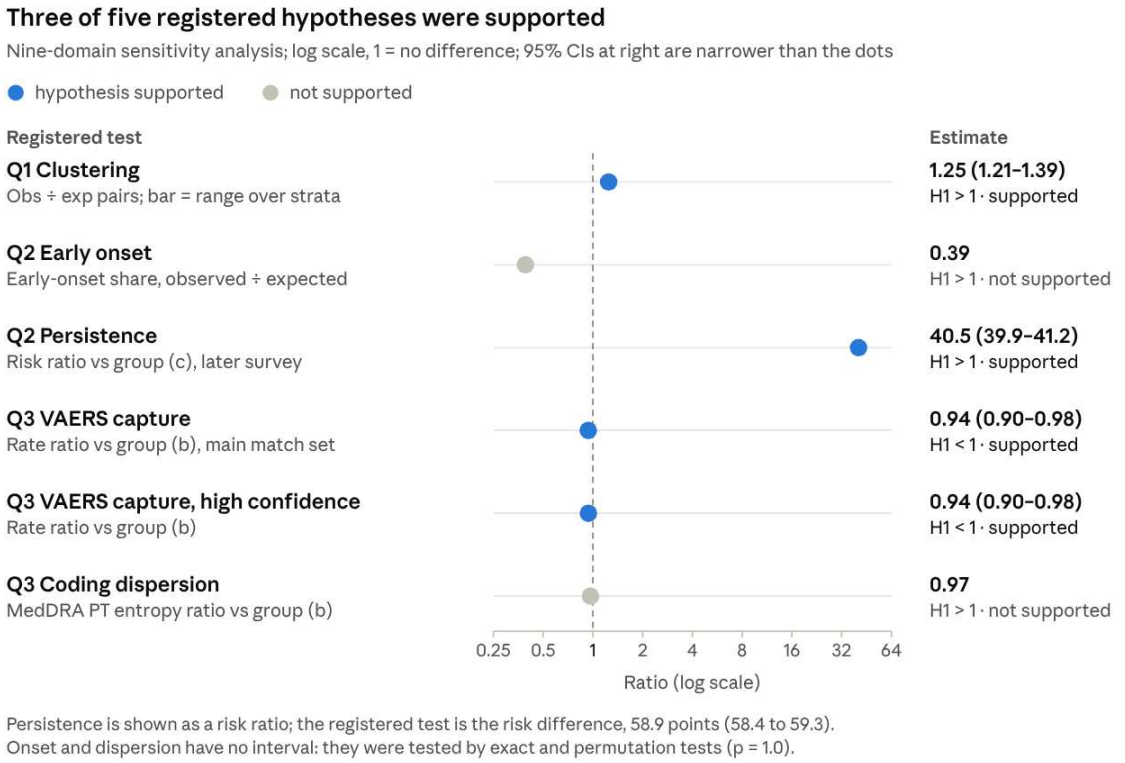


Figure 5. Registered tests of the nine-domain sensitivity analysis, log scale (1 = no difference). Persistence is shown as a risk ratio; the registered test is the risk difference, 58.9 points (95% CI 58.4 to 59.3). Onset and dispersion were tested by exact and permutation tests and have no interval. 'VAERS capture' is the recorded VAERS filing defined in Methods.

Clustering. The excess co-occurrence centres on cognitive symptoms, fatigue, sleep and dysautonomia, the core described for PACVS and for long COVID, and it held in every dose and manufacturer stratum and in both sexes. Pain and gastrointestinal symptoms, the commonest domains, contributed large counts but small excesses; menstrual and sensory symptoms did not join the cluster. Matching the null on amount of text (Table 7, Figure 4) removed most of the

aggregate excess but left the core at 1.6 to 2.5 times chance, and fixing each person-dose's domain count as well (Table 8) left it at 1.4 to 2.7. Summed over the six core pairs (Table 9), the whole excess that survives these nulls sits in the core. Not every core pair carries it: fatigue with dysautonomia lay at or below chance under every null, and the strongest ties all involve cognitive symptoms, the domain with the lowest recall of the four (0.74). The syndrome-like finding is that core, not the aggregate ratio.

Onset. The registered hypothesis assumed that a vaccine-related syndrome would be described first in the weeks after the dose. Instead, 72% of cases first described their case domains at 3 months or later. This has three readings, and our data cannot separate them:

- Late-onset or late-recognised effects. If real, they would be missed by any design that counts only events in a fixed window after the dose, often 1 to 42 days.
- Background illness. Late onset is also what intercurrent illness, including COVID-19 infection, would produce. It weakens the case for attribution to the dose.
- Late reporting. Symptoms present early may be written down only once they have lasted, or once an open-ended long-term survey invites a description that the check-box surveys of the first weeks did not.

Three features of the data bias the measure towards late description: early surveys collected symptoms mainly through check-boxes while long-term surveys were open-ended; early text was classified by keywords only, with lower recall; and daily week-1 check-ins inflate the expected early share to 71%. Classifying early and late text alike raised the early share only to 32.7%, and the check-boxes record reactogenicity that cannot be told apart from the start of a persistent syndrome (Table 10). The registered test therefore records when symptoms were first written down, not when they began; it is uninformative about onset, and the structured data cannot settle it.

Persistence. Persistence in 60% of cases, against 1.5% of group (c), is large, but part of the gap is built in: group (c) by construction reported no worse health or impact at long-term surveys, and cases were selected for naming at least two domains at a long-term survey, which makes naming one later more likely. Two checks specified after the results were seen narrow the gap without closing it (Table 11): same-survey cases persist in 47.1%, and group (d), people with the same worse health or impact who named one domain or none, in 10.5%, a difference of 36.6 points. Group (d) is the comparison the question needs, because it measures how often a single long-term complaint recurs in the same responders. The multi-domain pattern is therefore not a slow resolution of a severe first week, and it recurs more often than other long-term complaints among people who kept answering. Persistence relative to the general population or to those who stopped responding is not estimated, and the risk ratio in Figure 5 should not be read as an effect size.

Detection. Cases were about 6% less likely than people with named acute diagnoses to reach VAERS, and, contrary to our hypothesis, their reports were coded to a narrower spread of

terms, a difference that held under the Miller-Madow correction, rarefaction and a two-sided test. VAERS coding does not appear to scatter this pattern; the few cases who reach VAERS may be those with a more recognisable presentation. The detection problem, if there is one, lies in the 95% of cases who never reach VAERS, which this comparison cannot see. Of three readings of the exclusion criterion, the registered one gives the most conservative rate ratio.

Model sensitivity. The two language models agreed on early text but not on the long-term residue. DeepSeek labelled fatigue there less than half as often as Claude Opus, so fatigue's recall of 0.86, and its passing the gate, depend on which model labelled the residue; three other domains were also model-sensitive, and Claude applied the unrelated-cause exclusion 45% more often, so the case count itself would differ between models.

Relation to earlier work. Case series of PACVS describe fatigue, cognitive impairment, post-exertional malaise, dysautonomia and neuropathic symptoms as the core presentation [4–6, 10, 12], with a clinical and immunological profile shared with post-acute COVID-19 [4]. The cluster found here matches that core; sensory-neuropathic symptoms, prominent in some series [8, 9, 12], did not join it, which may reflect the classifier's recall for that domain (0.25). Published prevalence estimates run from 0.003% to 0.9% [2]; the 1.01% here sits at the top of that range before any adjustment for attrition or classifier error, and is a share of long-term responses, not of recipients. Levi and Lu found that 1.87% of participants reported worse health lasting at least three months that they attributed to vaccination [22]; our definition is stricter, requiring at least two symptom domains in free text and no named alternative cause. Their cohorts were reproduced within 0.5% at doses 1 and 2, and the two definitions overlap only partly (kappa 0.22): a third of text-defined cases attributed worse health to vaccination on a check-box, and a sixth of their cohort met the text definition, so the text-defined phenotype is not a relabelling of self-attribution. Women predominated among cases (77.5%), as in case series [5, 6], but v-safe registrants were also mostly women, and the no-symptoms group was 58% women.

Limitations

- The classifier failed its gate. Eight of nine domains fell short of 0.85 precision or recall, so every analysis rests on exploratory labels. Recall was 0.74 for cognitive and 0.72 for sleep symptoms, so the domains that drive the core are among the least completely counted. Independent coding of a sample large enough for domain-level precision and recall with usable confidence intervals is required before any confirmatory claim.
- The registered null ignores amount of text. The text-matched and curveball nulls removed the aggregate excess and left a specific core; the registered ratio of 1.25 should not be quoted without them.
- The reference is coding, not truth. CDC coded entries for adverse-event reporting, often to the most prominent symptom, and the frozen map misses relevant Preferred Terms, so measured precision is likely too low for some domains. The blind audit (Table 12) supports this reading but comes from a single coder, the author, blind to side but not to the hypotheses, and cannot rescue the gate.

- Unequal classification over time. Layer 3 labelled only long-term text, which favours finding domains at long-term surveys and affects onset most.
- The one gate pass is model-dependent: four PACVS domains, including fatigue, were model-sensitive.
- Self-described text, not diagnosis; no clinical confirmation was available.
- No age and no infection history. Nothing is age-adjusted, and intercurrent COVID-19 infection and other medical events are visible only through the exclusion tags, so a late vaccine effect cannot be told from background illness.
- Attrition and self-selection. Only 49% of person-doses with a check-in completed a long-term survey, so the prevalence bounds are wide; response was higher after a week-1 health impact (60.1% against 42.4%), which may inflate the 1.01%, and v-safe registrants were self-selected.
- Exclusions are imperfect. The unrelated-cause and acute-diagnosis tags had precision of 0.14 and 0.31, and negative COVID-19 tests and chronic conditions were sometimes tagged as exclusions, so criterion 3 and group (b) are leaky and the VAERS rate ratio ranges from 0.78 to 0.94 across readings.
- Comparison groups are partly defined by the outcome, and the strictest persistence contrast was specified after the results were seen.
- VAERS linkage and the detection test. Recorded filing relies on call-centre dispositions and OpenVSafe's probabilistic matches [18], which were not validated independently; Q3 was run per person while groups are defined per dose; and a contrast between two filing rates near 5% is a weak test of under-capture that says nothing about cases who never reach VAERS.
- Implementation choices after registration are listed in Methods; all were made without seeing results.

Conclusions

Of person-doses with a long-term v-safe survey, 1.01% describe a multi-domain symptom pattern. Once the null is matched on how much people wrote and on each person's number of symptoms, its clustering is confined to a core of cognitive symptoms, fatigue, sleep and dysautonomia, in women and men alike. The pattern persists across surveys and reaches VAERS slightly less often than named acute events. It was mostly first described months after vaccination, which is compatible with late vaccine effects, background illness including COVID-19 infection, or late reporting; these data do not establish that vaccination caused the pattern, its prevalence among recipients, or its status as a clinical entity. Because the classifier did not meet its pre-registered accuracy standard, the findings are hypothesis-generating. The next steps are independent multi-coder validation of the classifier, an onset analysis that treats the survey schedule as the observation process, and replication in other free-text or survey sources.

Acknowledgements. The author thanks OpenVAERS for publishing OpenVSafe's v-safe-to-VAERS matches and ICAN for hosting the FOIA releases.

Declarations

Funding. No funding was received for this study.

Conflicts of Interest. MH is Director of the Open Source Medicine Foundation and has previously received funding from the Independent Medical Alliance. He has published on PACVS and vaccine-injury recognition [3, 16]. He has no role with, and has received no payment from, OpenVAERS, ICAN or any litigation body, and the language-model services used were paid for at list price.

Ethics Approval. This secondary analysis used de-identified records released publicly under the US Freedom of Information Act and public VAERS data, and published only aggregate data with at least 10 people per cell. It did not involve contact with participants and did not require ethics review.

Consent to Participate. Not applicable; the study used de-identified public records.

Consent for Publication. Not applicable.

Availability of Data and Material. The v-safe, call-centre and MedDRA files are public FOIA releases hosted by ICAN [19]; VAERS data are public [20]; OpenVSafe's match pages are public [18]. The pre-registration and its two amendments are at osf.io/gnfpd. Only aggregates with at least 10 people per cell are released; no participant-level data or free text are shared.

Code Availability. Analysis code, the frozen taxonomy (v1.0) and the classifier prompts are at github.com/MattH55/vsafe-freetext, tag v1.0.2-submission, archived on Zenodo (doi:10.5281/zenodo.24178319).

Authors' Contributions. Matthew Halma: conceptualisation, methodology, software, validation, formal analysis, investigation, data curation, writing (original draft, review and editing), visualisation, project administration.

AI tools. A large language model (DeepSeek V4.1 Flash) served as the layer-3 classifier and a second (Claude Opus 5.5) as the model-sensitivity comparator, as described in Methods. AI assistants (Anthropic Claude) assisted with code and with drafting and editing this manuscript under the author's direction; the author reviewed all content and takes full responsibility for it. No AI tool is an author.

Data protection. Free text was stored locally, and layer 1 ran locally. For layer 3 it was sent, in batches of 20 entries with no registrant code, date or other field attached, to NanoGPT's API routed to DeepSeek V4.1 Flash with the provider's zero-data-retention and no-data-collection options set; the 2,000-entry comparison sample was sent to the Anthropic API (claude-opus-5-5) under that provider's standard API terms at the time of use. Entries containing a detectable e-

mail address or phone number were not sent. The retention terms as read at the time of use are recorded in the repository (docs/data-use.md). No free text is quoted.

References

1. Bellavite P, Donzelli A, Isidoro C. The WHO Algorithm for Causality Assessment of Adverse Effects Following Immunization with Genetic-Based Anti-COVID-19 Vaccines: Pitfalls and Suggestions for Improvement. J Clin Med. 2024;13:7291. doi:10.3390/jcm13237291
2. Platschek B, Boege F. The Post-Acute COVID-19-Vaccination Syndrome in the Light of Pharmacovigilance. Vaccines. 2024;12:1378. doi:10.3390/vaccines12121378
3. Halma M, Varon J. Restoring trust in vaccination: listening to patients and acknowledging Post-Acute COVID Vaccine Syndrome. Front Med. 2025;12. doi:10.3389/fmed.2025.1688170
4. Dinser J, Anft M, Wiemers L, Skrzypczyk S, Kohut E, Kurek J, et al. Shared Clinical and Immunological Features of Post-Acute COVID-19 and Post-Acute COVID-19 Vaccination Syndromes. medRxiv. 2026. doi:10.64898/2026.09.18.26363380
5. Mundorf AK, Semmler A, Heidecke H, Schott M, Steffen F, Bittner S, et al. Clinical and Diagnostic Features of Post-Acute COVID-19 Vaccination Syndrome (PACVS). Vaccines. 2024;12:790. doi:10.3390/vaccines12070790
6. Krumholz HM, Wu Y, Sawano M, Shah R, Zhou T, Arun AS, et al. Post-Vaccination Syndrome: A Descriptive Analysis of Reported Symptoms and Patient Experiences After Covid-19 Immunization. medRxiv. 2023. doi:10.1101/2023.11.09.23298266
7. Wang W, Yellamsetty A, Edmonds RM, Barcavage SR, Bao S. COVID-19 vaccination-related tinnitus is associated with pre-vaccination metabolic disorders. Front Pharmacol. 2024;15:1374320. doi:10.3389/fphar.2024.1374320
8. Rouleau I, Issa Kana KDN, Zafack JG, Viger YB, De Serres G. New-onset anesthesia/paresthesia following the administration of COVID-19 vaccines in Quebec, Canada. Vaccine. 2025;57:127217. doi:10.1016/j.vaccine.2025.127217
9. Safavi F, Gustafson L, Walitt B, Lehky T, Dehbashi S, Wiebold A, et al. Neuropathic symptoms with SARS-CoV-2 vaccination. medRxiv. 2022. doi:10.1101/2022.05.16.22274439
10. Tokumasu K, Fujita-Yamashita M, Sunada N, Sakurada Y, Yamamoto K, Nakano Y, et al. Characteristics of Persistent Symptoms Manifested after SARS-CoV-2 Vaccination: An Observational Retrospective Study in a Specialized Clinic for Vaccination-Related Adverse Events. Vaccines. 2023;11:1661. doi:10.3390/vaccines11111661
11. Shrestha Y, Venkataraman R. The prevalence of inverse health consequences of COVID-19 vaccines: A post-vaccination study. Vacunas. 2022;23:S67–76. doi:10.1016/j.vacun.2022.03.002
12. Gerhard A, Raeder V, Pernice HF, Boesl F, Schroeder M, Richter J, et al. Neurological symptoms after COVID-19 vaccination: a report on the clinical presentation of the first 50 patients. J Neurol. 2023;270:4673–7. doi:10.1007/s00415-023-11895-9
13. Finterer J, Scorza FA. A retrospective analysis of clinically confirmed long post-COVID vaccination syndrome. J Clin Transl Res. 2022;8:506–8. PMID 36452006
14. Grieshaber V, Strumann C, Holzwarth S, Toepfner N, von Meißner WCG, Konstantopoulos N, et al. The Long-Term Tolerability of BNT162b2 in Children and Adolescents (the CoVacU18 Study). Dtsch Arztebl Int. 2025;122:257–63. doi:10.3238/arztebl.m2025.0032
15. Riad A, Pokorná A, Attia S, Klugarová J, Koščík M, Klugar M. Prevalence of COVID-19 Vaccine Side Effects among Healthcare Workers in the Czech Republic. J Clin Med. 2021;10:1428. doi:10.3390/jcm10071428
16. Halma M, Varon J. DARE-SAFE: Denominator-Adjusted Rate Estimates of Substance Adverse Events Frequency Evaluation in Pharmaceuticals and Vaccines. Pharmacoepidemiology. 2025;4:7. doi:10.3390/pharma4020007

17. Zauche LH, Miller ER, Kim SS, Jazwa A, Shay DK. Monitoring Vaccine Safety: United States Vaccine Safety Surveillance Systems. J Pediatr Health Care. 2026;40:174–86. doi:10.1016/j.pedhc.2025.07.016

18. OpenVAERS. OpenVSafe: v-safe participant records matched to VAERS reports [Internet]. openvsafe.org. Person pages read 2026-11-10 to 2026-11-11; only the VAERS ID, match score (0 to 11), confidence and runner-up score of each match were retained.

19. Informed Consent Action Network. v-safe data [Internet]. icandecide.org/v-safe-data/. Free-text responses released by CDC under FOIA 24-01322 (twelve interim and final releases and an under-3 file, 2024), registrant, vaccination and health check-in files, call-centre records (FOIA 23-00925 and 23-00926) and CDC's MedDRA coding of the free text (November 2024). Downloaded 2026-11-08 to 2026-11-10.

20. US Department of Health and Human Services. Vaccine Adverse Event Reporting System (VAERS) data [Internet]. vaers.hhs.gov/data.html. All-years CSV files dated 2026-06-03.

21. MedDRA Maintenance and Support Services Organization. Medical Dictionary for Regulatory Activities (MedDRA) [Internet]. meddra.org.

22. Levi R, Lu N. Reporting on prolonged adverse health impact among V-SAFE enrollees. Research Square [Preprint]. Posted 25 Aug 2026. doi:10.21203/rs.3.rs-10768889/v1

Supplementary table S1: Q1 by stratum

Stratum	Person-doses	Observed pairs	Expected pairs (null mean)	Ratio	p
All	1,267,509	235,184	188,775.6	1.246	0.001
Dose 1	147,659	30,825	25,068.7	1.230	0.001
Dose 2	827,368	156,367	129,573.6	1.207	0.001
Dose 3	247,784	42,053	30,259.4	1.390	0.001
Dose 4	44,660	5,934	4,383.7	1.354	0.001
Janssen	55,754	12,621	10,320.7	1.223	0.001
Moderna	574,355	105,796	85,728.5	1.234	0.001
Pfizer-BioNTech	637,049	116,722	92,787.0	1.258	0.001

p = 0.001 is the smallest value possible with 1,000 permutations. Strata with fewer than 1,000 person-doses were not tested, as registered.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement.docx