From: "Niu, Manette" <[redacted]>
To: "Menschik, David" <[redacted]>
Cc: "Zinderman, Craig E" <[redacted]>
Subject: RE: a more efficient way to find events of interest
Date: Thu, 29 Apr 2021 11:56:36 +0000
Importance: Normal
Inline-Images: image001.png

No, I haven't requested anything from Ana. I am only passively passing on her data mining runs when she sends them to me.
Sorry for the confusion.
Thank you!

From: Menschik, David <[redacted]>
Sent: Thursday, April 29, 2021 7:25 AM
To: Niu, Manette <[redacted]>
Cc: Zinderman, Craig E <[redacted]>
Subject: RE: a more efficient way to find events of interest

Hi Manette,

Did you request this or anything else (COVID vaccine data mining related) from Ana and/or are you working with Ana on any data mining projects? (if so, please specify)

Thanks,
David

From: Niu, Manette <[redacted]>
Sent: Thursday, April 29, 2021 6:36 AM
To: Zinderman, Craig E <[redacted]>
Cc: Baer, Bethany <[redacted]>; Menschik, David <[redacted]>
Subject: FW: a more efficient way to find events of interest

fyi

From: Szarfman, Ana <[redacted]>
Sent: Thursday, April 29, 2021 12:49 AM
To: Allende, Maria <[redacted]>; Niu, Manette <[redacted]>
Cc: Stockbridge, Norman L <[redacted]>
Subject: a more efficient way to find events of interest

Hi Maria and Manette,

I am sharing an analysis that was requested at your end.

Please refer to the attached audit trail and to the companion 3D data mining analysis displaying all TTP cases reported for COVID-19 vaccines as of April 23, 2021 .

By grouping PTs and HLTs representing TTP into a custom term, and using a 3D display of “vaccine--PT--custom term” it enables the reviewer to focus on every associated single event of interest with each vaccine. The associated reports can be easily grouped and accessed by drilling down techniques.

See highlighted in yellow potential events that may be associated with brain TTP.

Let me know if you need any additional feedback.

--Ana

Ana Szarfman, MD, PhD, FAMIA,
Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), and Fellow of the American Medical Informatics Association (2020)
Medical Officer, Safety Data Mining Developer and Medical Informatics Analyst,
Celebrating nearly a quarter of a century of successful implementation of safety data mining, interactive patient profiles, and other automated analytical tools.
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted] (office)
(personal cell phone and WhatsApp)

FDA U.S. FOOD & DRUG ADMINISTRATION

From: "Niu, Manette" <[redacted]>
To: "Menschik, David" <[redacted]>
Cc: "Zinderman, Craig E" <[redacted]>
Subject: RE: a more efficient way to find events of interest
Date: Thu, 29 Apr 2021 11:57:07 +0000
Importance: Normal
Inline-Images: image001.png

And no, I am not working on anything with her.

From: Menschik, David <[redacted]>
Sent: Thursday, April 29, 2021 7:25 AM
To: Niu, Manette <[redacted]>
Cc: Zinderman, Craig E <[redacted]>
Subject: RE: a more efficient way to find events of interest

Hi Manette,

Did you request this or anything else (COVID vaccine data mining related) from Ana and/or are you working with Ana on any data mining projects? (if so, please specify)

Thanks,
David

From: Niu, Manette <[redacted]>
Sent: Thursday, April 29, 2021 6:36 AM
To: Zinderman, Craig E <[redacted]>
Cc: Baer, Bethany <[redacted]>; Menschik, David <[redacted]>
Subject: FW: a more efficient way to find events of interest

fyi

From: Szarfman, Ana <[redacted]>
Sent: Thursday, April 29, 2021 12:49 AM
To: Allende, Maria <[redacted]>; Niu, Manette <[redacted]>
Cc: Stockbridge, Norman L <[redacted]>
Subject: a more efficient way to find events of interest

Hi Maria and Manette,

I am sharing an analysis that was requested at your end.

Please refer to the attached audit trail and to the companion 3D data mining analysis displaying all TTP cases reported for COVID-19 vaccines as of April 23, 2021 .

By grouping PTs and HLTs representing TTP into a custom term, and using a 3D display of “vaccine--PT--custom term” it enables the reviewer to focus on every associated single event of interest with each vaccine. The associated reports can be easily grouped and accessed by drilling down techniques.

See highlighted in yellow potential events that may be associated with brain TTP.

Let me know if you need any additional feedback.

--Ana

Ana Szarfman, MD, PhD, FAMIA,
Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), and Fellow of the American Medical Informatics Association (2020)
Medical Officer, Safety Data Mining Developer and Medical Informatics Analyst,
Celebrating nearly a quarter of a century of successful implementation of safety data mining, interactive patient profiles, and other automated analytical tools.
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted] (office)
[redacted](personal cell phone and WhatsApp)

From: "Niu, Manette" <[redacted]>
To: "Baer, Bethany" <[redacted]>, "Zinderman, Craig E" <[redacted]>
Cc: "Menschik, David" <[redacted]>
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13
Date: Thu, 22 Apr 2021 18:33:10 +0000
Importance: Normal
Inline-Images: image001.png

Thank you for letting me know. I have not been working with Ana directly, although she has sent me data mining runs that I've forwarded to this group. I will speak to her about this.
Manette

From: Baer, Bethany <[redacted]>
Sent: Thursday, April 22, 2021 2:20 PM
To: Niu, Manette <[redacted]>; Zinderman, Craig E <[redacted]>
Cc: Menschik, David <[redacted]>
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13

I just wanted to let you know that on the biweekly CDER/CBER/Commonwealth Empirica support call today Ana offered to show individuals the interesting VAERS analysis she has been doing with Manette. A couple of the Commonwealth folks expressed interest in meeting with her to see it.
Thanks,
Bethany

From: Niu, Manette <[redacted]>
Sent: Monday, April 19, 2021 6:15 AM
To: Zinderman, Craig E <[redacted]>
Cc: Baer, Bethany <[redacted]>; Menschik, David <[redacted]>
Subject: FW: [EXTERNAL] HLT RUN FOR WEEK 13

fyi

From: Szarfman, Ana <[redacted]>
Sent: Saturday, April 17, 2021 8:12 PM
To: Niu, Manette <[redacted]>
Subject: FW: [EXTERNAL] HLT RUN FOR WEEK 13

FYI

From: Szarfman, Ana
Sent: Saturday, April 17, 2021 8:07 PM
To: 'Bill DuMouchel' <[redacted]>; Rave Harpaz <[redacted]>
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13

Hi,

These are very interesting results Bill. Many thanks for all your work!

The olfactory events are manifestations of the disease, the facial cranial nerve disorders may be cases of Bell's palsy.

There are many interesting cardiac and neurologic events. I highlighted some in your attached spreadsheet.

From: Bill DuMouchel <[redacted]>
Sent: Saturday, April 17, 2021 4:52 PM
To: Rave Harpaz [redacted]; Rob Van Manen <[redacted]>; Steve Bright <[redacted]>; Szarfman, Ana <[redacted]>; Alexander Nip <[redacted]>; Mohammad Al-Ansari <[redacted]>
Cc: Robert Weber <[redacted]>; Bruce Palsulich <[redacted]>
Subject: [EXTERNAL] HLT RUN FOR WEEK 13

I reanalyzed the Week 13 data at the HLT level, with COVID19+MANUFACTURER as the product variable.
There are about 5000 rows if you go to our runID# 335.
The attached spreadsheet only includes 142 rows from Selected SOCs where ER05 > 1. I tried to select SOCs where there would be plausible AEs as opposed to Covid symptoms, etc.

But probably there are still Covid symptoms mixed in, so medical knowledge will still be useful.

As usual, the reports are received from 1/1/2015 and later and stratified by 3 Gender labels and 11 Age groups.
Since the strata are all pretty highly populated because of using HLT and not stratifying by report year, I decided to use stratified versions of PRR and ROR, so that they are now often not too far from RR.

As mentioned above, all the rows in the attachment have ER05 > 1. If ER05 > 2, then the cell is shaded a bit darker. I did the same thing for cells where EB05 > 2.

I'm sort of surprised that there are so many DECs with ER05 greater than 1 and 2. But it provides food for thought, I guess.

Bill

From: Bill DuMouchel
Sent: Friday, April 16, 2021 4:19 PM
To: Rave Harpaz <[redacted]>; Rob Van Manen <[redacted]>; Steve Bright <[redacted]>; Szarfman, Ana <[redacted]>; Alexander Nip <[redacted]>; Mohammad Al-Ansari <[redacted]>
Cc: Robert Weber <[redacted]>; Bruce Palsulich <[redacted]>
Subject: Appendicitis, Bell's Palsy and Thrombotic events with vaccines after Week 13

Run ID#335 showing VCOVID19+Manuf VS PT+SMQ, WITH ER05 > 1 highlighted

COVID19+MANUFACT vs Appendicitis, Bell's Palsy and Thrombotic Events

Vaccine Type + Manufacturer	Event: PT_plus_SMQ	N	E	ER05	ERAM	ER95	EB05	EBGM	PRR
COVID19_PFIZER/BIONTECH	Appendicitis complicated and perforated (Custom Term)	9	2.3	2.52	4.72	7.47	1.43	2.45	8.33
COVID19_MODERNA	Appendicitis complicated and perforated (Custom Term)	3	2.0	0.62	1.97	3.95	0.49	1.12	2.31
COVID19_JANSSEN	Appendicitis complicated and perforated (Custom Term)	0	0.3	0.01	0.66	2.35			
COVID19_PFIZER/BIONTECH	Facial paralysis - Bell's palsy (Custom Term)	360	181.9	2.20	2.40	2.61	1.80	1.97	3.18
COVID19_MODERNA	Facial paralysis - Bell's palsy (Custom Term)	213	151.5	1.52	1.71	1.90	1.25	1.40	2.06
COVID19_JANSSEN	Facial paralysis - Bell's palsy (Custom Term)	14	35.1	0.34	0.53	0.76	0.27	0.42	0.55
COVID19_MODERNA	Thrombectomy	10	2.2	5.38	9.85	15.38	1.65	2.79	14.60
COVID19_PFIZER/BIONTECH	Thrombectomy	7	2.4	3.02	6.33	10.64	1.07	1.94	6.48
COVID19_JANSSEN	Thrombectomy	1	0.3	0.28	3.46	9.64	0.34	1.07	3.56
COVID19_MODERNA	Thrombocytopenia	40	21.9	1.73	2.28	2.90	1.34	1.74	1.23
COVID19_PFIZER/BIONTECH	Thrombocytopenia	42	28.6	1.42	1.86	2.34	1.10	1.42	1.07
COVID19_JANSSEN	Thrombocytopenia	4	4.1	0.49	1.17	2.10	0.43	0.90	0.53
COVID19_MODERNA	Thrombolysis	2	0.6	0.61	2.38	5.11	0.52	1.34	7.30
COVID19_PFIZER/BIONTECH	Thrombolysis	2	0.7	0.57	2.24	4.80	0.50	1.28	6.02
COVID19_JANSSEN	Thrombolysis	0	0.1	0.02	0.90	2.94			
COVID19_PFIZER/BIONTECH	Thrombophlebitis (SMQ) [broad]	121	44.2	4.90	5.75	6.61	2.28	2.65	5.24
COVID19_MODERNA	Thrombophlebitis (SMQ) [broad]	94	39.3	4.17	4.98	5.85	1.95	2.31	4.50
COVID19_JANSSEN	Thrombophlebitis (SMQ) [broad]	11	8.2	1.43	2.47	3.74	0.75	1.22	1.92
COVID19_PFIZER/BIONTECH	Thrombophlebitis (SMQ) [narrow]	7	2.2	1.10	2.17	3.52	1.11	2.01	6.02
COVID19_MODERNA	Thrombophlebitis (SMQ) [narrow]	2	2.0	0.30	0.96	1.94	0.34	0.87	1.54
COVID19_JANSSEN	Thrombophlebitis (SMQ) [narrow]	1	0.4	0.29	1.28	2.84	0.33	1.01	3.38
COVID19_PFIZER/BIONTECH	Thrombosis	64	22.6	3.99	4.95	6.00	2.17	2.67	5.50
COVID19_MODERNA	Thrombosis	41	20.0	2.73	3.59	4.55	1.49	1.93	3.67
COVID19_JANSSEN	Thrombosis	6	3.9	1.05	2.25	3.82	0.67	1.27	2.05
COVID19_PFIZER/BIONTECH	Thrombotic thrombocytopenic purpura	5	1.8	0.98	2.31	4.09	0.86	1.71	4.01
COVID19_JANSSEN	Thrombotic thrombocytopenic purpura	0	0.4	0.01	0.67	2.26			
COVID19_MODERNA	Thrombotic thrombocytopenic purpura	0	1.5	0.01	0.32	1.08			

From: "Niu, Manette" <[redacted]>
To: "Menschik, David" [redacted]>
Subject: FW: VAERS data - HLT RUN FOR WEEK 13
Date: Thu, 29 Apr 2021 13:35:22 +0000
Importance: Normal
Inline-Images: image001.png

I sent Ana this email earlier this week, fyi.
Thank you!
Manette

From: Szarfman, Ana <[redacted]>
Sent: Monday, April 26, 2021 1:44 PM
To: Niu, Manette [redacted]>
Cc: Stockbridge, Norman L <[redacted]>
Subject: RE: VAERS data - HLT RUN FOR WEEK 13

Hi,

I understand that this cannot be your focus now. I appreciate your very valuable insight.

I will still share the outputs with you, to keep you inform about this work.

Warmest regards, Ana

From: Niu, Manette <[redacted]>
Sent: Monday, April 26, 2021 1:26 PM
To: Szarfman, Ana <[redacted]>
Subject: RE: VAERS data - HLT RUN FOR WEEK 13

Ana,
While we are aware that CDER is using the vaccine data to explore new calculations and various deviations of analysis parameters in disproportionality analysis, I haven't been, and are unable to, work as a collaborator with you on this project due to our higher priority work, and because this sort of statistical development work falls outside of my area of expertise.
Thank you!
Manette

From: Szarfman, Ana <[redacted]>
Sent: Sunday, April 25, 2021 10:16 AM
To: Allende, Maria [redacted]; Niu, Manette <[redacted]>
Cc: Stockbridge, Norman L <[redacted]>; Southworth, Mary Ross <[redacted]>; Senatore, Fortunato <[redacted]>
Subject: VAERS data - HLT RUN FOR WEEK 13

Hi Mariaca,

I created a pdf file so you can read the information at your end.

I am forwarding a DM output generated by Bill DuMouchel for the VAERS data up to week 13. He uses the public domain data. All this information is, of course, conveyed to Manette Niu.

ER05-ERAM-ER95 are the results for RGPS, a data mining method that is better at removing false positives and negatives than MGPS.

Note the safety signals for cardiac events with the Pfizer and Moderna vaccines, now in the news, that are better identified by RGPS than by MGPS.

Warmest regards, Ana

From: Szarfman, Ana
Sent: Saturday, April 17, 2021 8:12 PM
To: Niu, Manette <[redacted]>
Subject: FW: [EXTERNAL] HLT RUN FOR WEEK 13

FYI

From: Szarfman, Ana
Sent: Saturday, April 17, 2021 8:07 PM
To: 'Bill DuMouchel' <[redacted]>; Rave Harpaz <[redacted]
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13

Hi,

These are very interesting results Bill. Many thanks for all your work!

The olfactory events are manifestations of the disease, the facial cranial nerve disorders may be cases of Bell's palsy.

There are many interesting cardiac and neurologic events. I highlighted some in your attached spreadsheet.

From: Bill DuMouchel <[redacted]>
Sent: Saturday, April 17, 2021 4:52 PM
To: Rave Harpaz <[redacted]>; Rob Van Manen <[redacted]>; Steve Bright <[redacted]>; Szarfman, Ana <[redacted]>; Alexander Nip <[redacted]>; Mohammad Al-Ansari <[redacted]>
Cc: Robert Weber <[redacted]>; Bruce Palsulich <[redacted]>
Subject: [EXTERNAL] HLT RUN FOR WEEK 13

I reanalyzed the Week 13 data at the HLT level, with COVID19+MANUFACTURER as the product variable.
There are about 5000 rows if you go to our runID# 335.
The attached spreadsheet only includes 142 rows from Selected SOCs where ER05 > 1. I tried to select SOCs where there would be plausible AEs as opposed to Covid symptoms, etc.
But probably there are still Covid symptoms mixed in, so medical knowledge will still be useful.

As usual, the reports are received from 1/1/2015 and later and stratified by 3 Gender labels and 11 Age groups.
Since the strata are all pretty highly populated because of using HLT and not stratifying by report year, I decided to use stratified versions of PRR and ROR, so that they are now often not too far from RR.

As mentioned above, all the rows in the attachment have ER05 > 1. If ER05 > 2, then the cell is shaded a bit darker. I did the same thing for cells where EB05 > 2.

I'm sort of surprised that there are so many DECs with ER05 greater than 1 and 2. But it provides food for thought, I guess.

Bill

From: Bill DuMouchel
Sent: Friday, April 16, 2021 4:19 PM
To: Rave Harpaz <[redacted]>; Rob Van Manen <[redacted]>; Steve Bright [redacted]>; Szarfman, Ana <[redacted]>; Alexander Nip [redacted]>; Mohammad Al-Ansari <[redacted]>
Cc: Robert Weber [redacted]>; Bruce Palsulich <[redacted]>
Subject: Appendicitis, Bell's Palsy and Thrombotic events with vaccines after Week 13

Run ID#335 showing VCOVID19+Manuf VS PT+SMQ, WITH ER05 > 1 highlighted

COVID19+MANUFACT vs Appendicitis, Bell's Palsy and Thrombotic Events

Vaccine Type + Manufacturer	Event: PT_plus_SMQ	N	E	ER05	ERAM	ER95	EB05	EBGM	PRR
COVID19_PFIZER/BIONTECH	Appendicitis complicated and perforated (Custom Term)	9	2.3	2.52	4.72	7.47	1.43	2.45	8.33
COVID19_MODERNA	Appendicitis complicated and perforated (Custom Term)	3	2.0	0.62	1.97	3.93	0.49	1.12	2.31
COVID19_JANSSEN	Appendicitis complicated and perforated (Custom Term)	0	0.3	0.01	0.66	2.35			
COVID19_PFIZER/BIONTECH	Facial paralysis - Bell's palsy (Custom Term)	360	181.9	2.20	2.40	2.61	1.80	1.97	3.18
COVID19_MODERNA	Facial paralysis - Bell's palsy (Custom Term)	213	151.5	1.52	1.71	1.90	1.25	1.40	2.06
COVID19_JANSSEN	Facial paralysis - Bell's palsy (Custom Term)	14	35.1	0.34	0.53	0.76	0.27	0.42	0.55
COVID19_MODERNA	Thrombectomy	10	2.2	5.38	9.83	15.38	1.65	2.79	14.60
COVID19_PFIZER/BIONTECH	Thrombectomy	7	2.4	3.02	6.33	10.64	1.07	1.94	6.48
COVID19_JANSSEN	Thrombectomy	1	0.3	0.28	3.46	9.64	0.34	1.07	3.56
COVID19_MODERNA	Thrombocytopenia	40	21.9	1.73	2.28	2.90	1.34	1.74	1.23
COVID19_PFIZER/BIONTECH	Thrombocytopenia	42	28.6	1.42	1.86	2.34	1.10	1.42	1.07
COVID19_JANSSEN	Thrombocytopenia	4	4.1	0.49	1.17	2.10	0.43	0.90	0.53
COVID19_MODERNA	Thrombolysis	2	0.6	0.61	2.38	5.11	0.52	1.34	7.30
COVID19_PFIZER/BIONTECH	Thrombolysis	2	0.7	0.57	2.24	4.80	0.50	1.28	6.02
COVID19_JANSSEN	Thrombolysis	0	0.1	0.02	0.90	2.94			
COVID19_PFIZER/BIONTECH	Thrombophlebitis (SMQ) [broad]	121	44.2	4.90	5.73	6.61	2.28	2.65	5.24
COVID19_MODERNA	Thrombophlebitis (SMQ) [broad]	94	39.3	4.17	4.98	5.85	1.95	2.31	4.50
COVID19_JANSSEN	Thrombophlebitis (SMQ) [broad]	11	8.2	1.43	2.47	3.74	0.75	1.22	1.92
COVID19_PFIZER/BIONTECH	Thrombophlebitis (SMQ) [narrow]	7	2.2	1.10	2.17	3.52	1.11	2.01	6.02
COVID19_MODERNA	Thrombophlebitis (SMQ) [narrow]	2	2.0	0.30	0.96	1.94	0.34	0.87	1.54
COVID19_JANSSEN	Thrombophlebitis (SMQ) [narrow]	1	0.4	0.29	1.28	2.84	0.33	1.01	3.38
COVID19_PFIZER/BIONTECH	Thrombosis	64	22.6	3.99	4.95	6.00	2.17	2.67	5.50
COVID19_MODERNA	Thrombosis	41	20.0	2.73	3.59	4.55	1.49	1.93	3.67
COVID19_JANSSEN	Thrombosis	6	3.9	1.05	2.25	3.82	0.67	1.27	2.05
COVID19_PFIZER/BIONTECH	Thrombotic thrombocytopenic purpura	5	1.8	0.98	2.31	4.09	0.86	1.71	4.01
COVID19_JANSSEN	Thrombotic thrombocytopenic purpura	0	0.4	0.01	0.67	2.26			
COVID19_MODERNA	Thrombotic thrombocytopenic purpura	0	1.5	0.01	0.32	1.08			

From: "Niu, Manette" <[redacted]>
To: "Baer, Bethany" <[redacted]>, "Zinderman, Craig E" <[redacted]>, "Menschik, David" [redacted]
Subject: RE: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports
Date: Thu, 15 Apr 2021 13:27:27 +0000
Importance: Normal
Inline-Images: image001.png

I'll forward you Ana's email with the attachment. The best person to ask would be Ana as she has close ties with Bill Dumouchel.
Thank you!
Manette

From: Baer, Bethany <[redacted]>
Sent: Thursday, April 15, 2021 9:00 AM
To: Zinderman, Craig E <[redacted]>; Menschik, David <[redacted]>
Cc: Niu, Manette <[redacted]>
Subject: RE: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

Thanks for forwarding this on. I agree that we should consider different approaches as the underlying database is changing significantly due to the high volume of COVID vaccine reports. I think we should welcome any expert input. The spreadsheet that Bill mentioned in the first email is not attached so I can't look at it, but David and I have been discussing and are concerned about the effect of so many COVID reports on the standard system we use. Is there a way that Bill can be more involved in our data mining process and interpretation during this unprecedented reporting time?

Thanks,
Bethany

From: Zinderman, Craig E <[redacted]>
Sent: Wednesday, April 14, 2021 2:02 PM
To: Menschik, David <[redacted]>; Baer, Bethany <[redacted]>
Cc: Niu, Manette <[redacted]>
Subject: FW: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

David, Bethany:
Might be worth considering the below? I don't pretend to understand it, but sounds like they are suggesting an analysis not stratified by year. Thoughts?

Thanks,
Craig

From: Niu, Manette <[redacted]>
Sent: Wednesday, April 14, 2021 6:24 AM
To: Zinderman, Craig E <[redacted]>
Subject: FW: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

fyi

From: Szarfman, Ana <[redacted]>
Sent: Tuesday, April 13, 2021 9:17 PM
To: Niu, Manette <[redacted]>
Cc: Stockbridge, Norman L [redacted]>
Subject: RE: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

Thanks Manette.

Exactly. As DuMouchel pinpointed, there is a need to extend the stratification brackets by the fact that 99% of the results for FY2021 are for COVID-19 vaccines this indeed affects the results.

From: Niu, Manette <[redacted]>
Sent: Monday, April 12, 2021 7:01 AM
To: Szarfman, Ana <[redacted]>
Subject: FW: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

Ana,
Does this effect the data mining results we are receiving in 2021? As you know, there is a backlog in VAERS reports with the contractor due to the high volume of reports we are receiving for the COVID-19 vaccines and the prioritization of those vaccine reports.
Thank you!
Manette

From: Szarfman, Ana <[redacted]>
Sent: Saturday, April 10, 2021 1:22 PM
To: Niu, Manette <[redacted]>
Cc: Vega, Amarilys <[redacted]>; Stockbridge, Norman L <[redacted]>; Quinn, John <[redacted]>; bill.dumouchel <[redacted]>; Rave Harpaz <[redacted]>; Pease-Fye, Meg <[redacted]>; Weichold, Frank <[redacted]>; Callahan, Lawrence <[redacted]>; Paredes, Antonio <[redacted]>; Temple, Robert <[redacted]>; Blum, Michael <[redacted]>; Dal Pan, Gerald <[redacted]>; Zander, Judith <[redacted]>; Munoz, Monica <[redacted]>; Diak, Ida-Lina <[redacted]>
Subject: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

Hello all,

Please refer to the message from Bill DuMouchel that I am forwarding and to his attached spreadsheet.

Notice how Bill discovered the need to eliminate the stratification by year when the reports for the COVID-19 vaccine in VAERS are 99% of all reports for a year (2021).

I think that we need to invite him to talk with us about the effect of adjustment factors, given the data, so we can all learn from his knowledge.

Warmest regards to all,

--Ana

Ana Szarfman, MD, PhD, FAMIA,
Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), and Fellow of the American Medical Informatics Association (2020)
Medical Officer, Safety Data Mining Developer and Medical Informatics Analyst,
Celebrating nearly a quarter of a century of successful implementation of safety data mining, interactive patient profiles, and other automated analytical tools.
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted] (office)
[redacted] (personal cell phone and WhatsApp)

From: Bill DuMouchel <[redacted]>
Sent: Saturday, April 10, 2021 2:25 AM
To: Rave Harpaz <[redacted]>; Steve Bright <[redacted]>; Rob Van Manen <[redacted]>
Cc: Szarfman, Ana <[redacted]>; Mohammad Al-Ansari <[redacted]>; Robert Weber [redacted]>; Bruce Palsulich [redacted]>
Subject: [EXTERNAL] Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

The attached spreadsheet shows some COVID19 results for the three-year period 2019-2021
2019 has no COVID19 reports
2020 has a few
2021 consists of almost all (33929/34256 > 99%) COVID99 reports

Look at the values of A, B, C, D ... A+C is much greater than B+D in 2021.

The years 2020 and 2021 are shown as separate analyses. Note that RR as well as the Bayesian estimates are almost equal to 1.
They stay almost equal to one if the run is stratified by year, because the 2021 results dominate.
The next two sets of results show the full 3-year estimates with and without including year as one of the stratification covariates.
Only if you mix in more non-covid reports within each stratum can you get enough diversity to allow larger disproportionalities.

-Bill

From: "Niu, Manette" <[redacted]>
To: "Zinderman, Craig E" <[redacted]>, "Baer, Bethany" <[redacted]>, "Menschik, David" <[redacted]>
Subject: FW: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports
Date: Thu, 15 Apr 2021 13:33:02 +0000
Importance: Normal
Attachments: CompareStratifications.xls
Inline-Images: image001.png

---

fyi

From: Szarfman, Ana <[redacted]>
Sent: Saturday, April 10, 2021 1:22 PM
To: Niu, Manette <[redacted]>
Cc: Vega, Amarilys <[redacted]>; Stockbridge, Norman L <[redacted]>; Quinn, John <[redacted]>; bill.dumouchel <[redacted]>; Rave Harpaz <[redacted]>; Pease-Fye, Meg <[redacted]>; Weichold, Frank <[redacted]>; Callahan, Lawrence <[redacted]>; Paredes, Antonio <[redacted]>; Temple, Robert <[redacted]>; Blum, Michael <[redacted]>; Dal Pan, Gerald <[redacted]>; Zander, Judith <[redacted]>; Munoz, Monica <[redacted]>; Diak, Ida-Lina <Ida-[redacted]>
Subject: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

Hello all,

Please refer to the message from Bill DuMouchel that I am forwarding and to his attached spreadsheet.

Notice how Bill discovered the need to eliminate the stratification by year when the reports for the COVID-19 vaccine in VAERS are 99% of all reports for a year (2021).

I think that we need to invite him to talk with us about the effect of adjustment factors, given the data, so we can all learn from his knowledge.

Warmest regards to all,

--Ana

Ana Szarfman, MD, PhD, FAMIA,
Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), and Fellow of the American Medical Informatics Association (2020)
Medical Officer, Safety Data Mining Developer and Medical Informatics Analyst,
Celebrating nearly a quarter of a century of successful implementation of safety data mining, interactive patient profiles, and other automated analytical tools.
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted] (office)
[redacted] (personal cell phone and WhatsApp)

U.S. FOOD & DRUG ADMINISTRATION

From: Bill DuMouchel <[redacted]>
Sent: Saturday, April 10, 2021 2:25 AM
To: Rave Harpaz <[redacted]>; Steve Bright <[redacted]>; Rob Van Manen <[redacted]>
Cc: Szarfman, Ana <[redacted]>; Mohammad Al-Ansari <[redacted]>; Robert Weber <[redacted]>; Bruce Palsulich <[redacted]>
Subject: [EXTERNAL] Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

The attached spreadsheet shows some COVID19 results for the three-year period 2019-2021
2019 has no COVID19 reports
2020 has a few
2021 consists of almost all (33929/34256 > 99%) COVID99 reports

Look at the values of A, B, C, D ... A+C is much greater than B+D in 2021.

The years 2020 and 2021 are shown as separate analyses. Note that RR as well as the Bayesian estimates are almost equal to 1.
They stay almost equal to one if the run is stratified by year, because the 2021 results dominate.
The next two sets of results show the full 3-year estimates with and without including year as one of the stratification covariates.
Only if you mix in more non-covid reports within each stratum can you get enough diversity to allow larger disproportionalities.

-Bill

From: "Zinderman, Craig E" <[redacted]>
To: "Menschik, David" <[redacted]>
Subject: FW: [EXTERNAL] HLT RUN FOR WEEK 13
Date: Mon, 26 Apr 2021 13:55:06 +0000
Importance: Normal
Inline-Images: image001.png

Just fyi…

Thanks,
Craig

From: Niu, Manette <[redacted]>
Sent: Monday, April 26, 2021 9:52 AM
To: Zinderman, Craig E [redacted]
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13

Craig,
I'm fine with the data mining runs she is sending, but the main issue is the collaborator one, since I've not been actively working with her.
Thank you so much for your suggestions, much appreciated!
Manette

From: Zinderman, Craig E <[redacted]>
Sent: Monday, April 26, 2021 9:32 AM
To: Niu, Manette <[redacted]>
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13

Seems reasonable to ask her to please stop describing you as a collaborator.

I would say something like this: while we are aware that CDER is using the vaccine data to explore new calculations and various deviations of analysis parameters in disproportionality analysis, you haven't been, and are unable to, work as a collaborator with her on this project to our higher priority work, and because this sort of statistical development work falls outside of your area of interest/expertise.

Just a suggestion; fail free to revise or not use at all, as you see fit.

Are you asking her to stop sending you updates/results? That would present a bigger problem for us I think.

Thanks,
Craig

From: Niu, Manette <[redacted]>
Sent: Monday, April 26, 2021 9:06 AM
To: Zinderman, Craig E <[redacted]>
Subject: FW: [EXTERNAL] HLT RUN FOR WEEK 13

Craig, Bethany told me of this situation, to which I was trying to respond. How best to proceed? Is there anyone in our group who may be willing to work with her? Thank you! Manette

From: Baer, Bethany <[redacted]>
Sent: Thursday, April 22, 2021 2:56 PM
To: Niu, Manette <[redacted]>
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13

I realize it is a difficult situation and I defer to Craig and you about how you want to handle this, but I thought you should be aware of what was being said. She used your name as her CBER collaborator, which from our earlier discussions, I didn't think was quite accurate.
Thanks,
Bethany

From: Niu, Manette <[redacted]>
Sent: Thursday, April 22, 2021 2:33 PM
To: Baer, Bethany <[redacted]>; Zinderman, Craig E [redacted]

Cc: Menschik, David <[redacted]>
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13

Thank you for letting me know. I have not been working with Ana directly, although she has sent me data mining runs that I've forwarded to this group. I will speak to her about this.
Manette

From: Baer, Bethany <[redacted]>
Sent: Thursday, April 22, 2021 2:20 PM
To: Niu, Manette <[redacted]>; Zinderman, Craig E [redacted]
Cc: Menschik, David <[redacted]>
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13

I just wanted to let you know that on the biweekly CDER/CBER/Commonwealth Empirica support call today Ana offered to show individuals the interesting VAERS analysis she has been doing with Manette. A couple of the Commonwealth folks expressed interest in meeting with her to see it.
Thanks,
Bethany

From: Niu, Manette <[redacted]>
Sent: Monday, April 19, 2021 6:15 AM
To: Zinderman, Craig E <[redacted]>
Cc: Baer, Bethany <[redacted]>; Menschik, David <[redacted]>
Subject: FW: [EXTERNAL] HLT RUN FOR WEEK 13

fyi

From: Szarfman, Ana <[redacted]>
Sent: Saturday, April 17, 2021 8:12 PM
To: Niu, Manette <[redacted]>
Subject: FW: [EXTERNAL] HLT RUN FOR WEEK 13

FYI

From: Szarfman, Ana
Sent: Saturday, April 17, 2021 8:07 PM
To: 'Bill DuMouchel' [redacted]; Rave Harpaz <[redacted]>
Subject: RE: [EXTERNAL] HLT RUN FOR WEEK 13

Hi,

These are very interesting results Bill. Many thanks for all your work!

The olfactory events are manifestations of the disease, the facial cranial nerve disorders may be cases of Bell's palsy.

There are many interesting cardiac and neurologic events. I highlighted some in your attached spreadsheet.

From: Bill DuMouchel <[redacted]>
Sent: Saturday, April 17, 2021 4:52 PM
To: Rave Harpaz <[redacted]>; Rob Van Manen [redacted]; Steve Bright [redacted]; Szarfman, Ana <[redacted]>; Alexander Nip <[redacted]>; Mohammad Al-Ansari <[redacted]>
Cc: Robert Weber <[redacted]>; Bruce Palsulich <[redacted]>
Subject: [EXTERNAL] HLT RUN FOR WEEK 13

I reanalyzed the Week 13 data at the HLT level, with COVID19+MANUFACTURER as the product variable.
There are about 5000 rows if you go to our runID# 335.
The attached spreadsheet only includes 142 rows from Selected SOCs where ER05 > 1. I tried to select SOCs where there would be plausible AEs as opposed to Covid symptoms, etc.
But probably there are still Covid symptoms mixed in, so medical knowledge will still be useful.

As usual, the reports are received from 1/1/2015 and later and stratified by 3 Gender labels and 11 Age groups.
Since the strata are all pretty highly populated because of using HLT and not stratifying by report year, I decided to use stratified versions of PRR and ROR, so that they are now often not too far from RR.

As mentioned above, all the rows in the attachment have ER05 > 1. If ER05 > 2, then the cell is shaded a bit darker. I did the same thing for cells where EB05 > 2.

I'm sort of surprised that there are so many DECs with ER05 greater than 1 and 2. But it provides food for thought, I guess.

Bill

From: Bill DuMouchel
Sent: Friday, April 16, 2021 4:19 PM
To: Rave Harpaz <[redacted]>; Rob Van Manen [redacted]>; Steve Bright <[redacted]>; Szarfman, Ana <[redacted]>; Alexander Nip <[redacted]>; Mohammad Al-Ansari <[redacted]>
Cc: Robert Weber <[redacted]>; Bruce Palsulich <[redacted]>
Subject: Appendicitis, Bell's Palsy and Thrombotic events with vaccines after Week 13

Run ID#335 showing VCOVID19+Manuf VS PT+SMQ, WITH ER05 > 1 highlighted

COVID19+MANUFACT vs Appendicitis, Bell's Palsy and Thrombotic Events

Vaccine Type + Manufacturer	Event: PT_plus_SMQ	N	E	ER05	ERAM	ER95	EB05	EBGM	PRR
COVID19_PFIZER/BIONTECH	Appendicitis complicated and perforated (Custom Term)	9	2.3	2.52	4.72	7.47	1.43	2.45	8.33
COVID19_MODERNA	Appendicitis complicated and perforated (Custom Term)	3	2.0	0.62	1.97	3.93	0.49	1.12	2.31
COVID19_JANSSEN	Appendicitis complicated and perforated (Custom Term)	0	0.3	0.01	0.66	2.35			
COVID19_PFIZER/BIONTECH	Facial paralysis - Bell's palsy (Custom Term)	360	181.9	2.20	2.40	2.61	1.80	1.97	3.18
COVID19_MODERNA	Facial paralysis - Bell's palsy (Custom Term)	213	151.5	1.52	1.71	1.90	1.25	1.40	2.06
COVID19_JANSSEN	Facial paralysis - Bell's palsy (Custom Term)	14	35.1	0.34	0.53	0.76	0.27	0.42	0.55
COVID19_MODERNA	Thrombectomy	10	2.2	5.38	9.83	15.38	1.65	2.79	14.60
COVID19_PFIZER/BIONTECH	Thrombectomy	7	2.4	3.02	6.33	10.64	1.07	1.94	6.48
COVID19_JANSSEN	Thrombectomy	1	0.3	0.28	3.46	9.64	0.34	1.07	3.56
COVID19_MODERNA	Thrombocytopenia	40	21.9	1.73	2.28	2.90	1.34	1.74	1.23
COVID19_PFIZER/BIONTECH	Thrombocytopenia	42	28.6	1.42	1.86	2.34	1.10	1.42	1.07
COVID19_JANSSEN	Thrombocytopenia	4	4.1	0.49	1.17	2.10	0.43	0.90	0.53
COVID19_MODERNA	Thrombolysis	2	0.6	0.61	2.38	5.11	0.52	1.34	7.30
COVID19_PFIZER/BIONTECH	Thrombolysis	2	0.7	0.57	2.24	4.80	0.50	1.28	6.02
COVID19_JANSSEN	Thrombolysis	0	0.1	0.02	0.90	2.94			
COVID19_PFIZER/BIONTECH	Thrombophlebitis (SMQ) [broad]	121	44.2	4.90	5.73	6.61	2.28	2.65	5.24
COVID19_MODERNA	Thrombophlebitis (SMQ) [broad]	94	39.5	4.17	4.98	5.85	1.95	2.31	4.50
COVID19_JANSSEN	Thrombophlebitis (SMQ) [broad]	11	8.2	1.43	2.47	3.74	0.75	1.22	1.92
COVID19_PFIZER/BIONTECH	Thrombophlebitis (SMQ) [narrow]	7	2.2	1.10	2.17	3.52	1.11	2.01	6.02
COVID19_MODERNA	Thrombophlebitis (SMQ) [narrow]	2	2.0	0.30	0.96	1.94	0.34	0.87	1.54
COVID19_JANSSEN	Thrombophlebitis (SMQ) [narrow]	1	0.4	0.29	1.28	2.84	0.33	1.01	3.38
COVID19_PFIZER/BIONTECH	Thrombosis	64	22.6	3.99	4.95	6.00	2.17	2.67	5.50
COVID19_MODERNA	Thrombosis	41	20.0	2.73	3.59	4.55	1.49	1.93	3.67
COVID19_JANSSEN	Thrombosis	6	3.9	1.05	2.25	3.82	0.67	1.27	2.05
COVID19_PFIZER/BIONTECH	Thrombotic thrombocytopenic purpura	5	1.8	0.98	2.31	4.09	0.86	1.71	4.01
COVID19_JANSSEN	Thrombotic thrombocytopenic purpura	0	0.4	0.01	0.67	2.26			
COVID19_MODERNA	Thrombotic thrombocytopenic purpura	0	1.5	0.01	0.32	1.08			

12 Confidential – © 2021 Oracle

From: "Baer, Bethany" <[redacted]>
To: "Menschik, David" <[redacted]>
Subject: RE: Data mining question
Date: Fri, 15 Mar 2024 17:49:51 +0000
Importance: Normal
Attachments: Almenoff_data_mining_Drug_Safety_2005.pdf; Pharmacoepidemiology_and_Drug_-_2013_-_Maignen_-_Assessing_the_extent_and_impact_of_the_masking_effect_of.pdf
Inline-Images: image001.png; image002.jpg; image003.jpg; image004.jpg; image005.jpg; image006.jpg

---

Hi David,
Here are 2 more articles that were used as references in the Harpaz article. They don't focus on vaccines, but they both mention the risk of masking due to lack of diversity of the products in the background database used for data mining. I will let you choose if you feel these and/or the Harpaz article are relevant to pass on in response to Narayan's question.
Thanks,
Bethany

From: Baer, Bethany
Sent: Friday, March 15, 2024 1:30 PM
To: Menschik, David <[redacted]>
Subject: FW: Data mining question

Hi David,
I am just responding to you so you can decide if you want to use this article as an example or not. It goes back to the discussions about Ana's involvement in VAERS data mining and her interest in updating data mining methods.
Bethany

From: Zinderman, Craig [redacted]
Sent: Friday, March 15, 2024 1:22 PM
To: Nair, Narayan <[redacted]>; Menschik, David <[redacted]>; Baer, Bethany <[redacted]>
Subject: RE: Data mining question

I'm not aware of literature articles (although I can't say I've looked for it either). I recall Anna talking about masking in the few interactions we had with her, but I don't remember there being references.

Thanks,
Craig

Craig Zinderman, MD, MPH

Associate Director for Medical Policy
Office of Biostatistics and Pharmacovigilance
FDA/Center for Biologics Evaluation and Research
[redacted]

From: Nair, Narayan <[redacted]>
Sent: Friday, March 15, 2024 1:04 PM
To: Menschik, David <[redacted]>; Zinderman, Craig [redacted]; Baer, Bethany

<[redacted]>
Subject: Data mining question

Good afternoon,
I know in the past we have discussed one of the possible limitations of data mining currently is the vast number of VAERS reports from the COVID vaccines may limit our ability to detect statistical alerts because disproportionality scores may be driven towards the null. Do you know if there is a public reference that discusses this limitation? I have found some references that discuss general limitations for data mining but not sure if there is one that talks about how a large volume of reports from a single class of products could masks results.

Narayan Nair, MD (he/him/his)
Division Director

Division of Pharmacovigilance
Office of Biostatistics and Pharmacovigilance
Center for Biologics Evaluation and Research
U.S. Food and Drug Administration

[redacted]

FDA U.S. FOOD & DRUG ADMINISTRATION

LEADING ARTICLE

Drug Safety 2005; 28 (11): 981-1007
0114-5916/05/0011-0981/$34.95/0

Perspectives on the Use of Data Mining in Pharmacovigilance

June Almenoff,[1] *Joseph M. Tonning*,[2] *A. Lawrence Gould*,[3] *Ana Szarfman*,[2] *Manfred Hauben*,[4,5,6] *Rita Ouellet-Hellstrom*,[2] *Robert Ball*,[2] *Ken Hornbuckle*,[7] *Louisa Walsh*,[8] *Chuen Yee*,[9] *Susan T. Sacks*,[10] *Nancy Yuen*,[1] *Vaishali Patadia* *,[11] *Michael Blum*,[12] *Mike Johnston* **,[2] *Charles Gerrits* ***,[13] *Harry Seifert*[1] and *Karol LaCroix*[1]

1 GlaxoSmithKline, Research Triangle Park, North Carolina, USA
2 US Food & Drug Administration, Rockville, Maryland, USA
3 Merck Research Laboratories, West Point, Pennsylvania, USA
4 Pfizer Inc., New York, New York, USA
5 Department of Medicine, NYU School of Medicine, New York, New York, USA
6 Departments of Pharmacology and Community and Preventive Medicine, New York Medical College, Valhalla, New York, USA
7 Eli Lilly and Company, Indianapolis, Indiana, USA
8 AstraZeneca LP, Wilmington, Delaware, USA
9 Johnson & Johnson Pharmaceutical Research & Development L.L.C., Titusville, New Jersey, USA
10 Hoffmann-La Roche Inc., Nutley, New Jersey, USA
11 Allergan Inc., Irvine, California, USA
12 Wyeth Research, Collegeville, Pennsylvania, USA
13 Schering-Plough Research Institute, Springfield, New Jersey, USA

Abstract

In the last 5 years, regulatory agencies and drug monitoring centres have been developing computerised data-mining methods to better identify reporting relationships in spontaneous reporting databases that could signal possible adverse drug reactions. At present, there are no guidelines or standards for the use of these methods in routine pharmacovigilance. In 2003, a group of statisticians, pharmacoepidemiologists and pharmacovigilance professionals from the pharmaceutical industry and the US FDA formed the Pharmaceutical Research and Manufacturers of America-FDA Collaborative Working Group on Safety Evaluation Tools to review best practices for the use of these methods.

In this paper, we provide an overview of: (i) the statistical and operational attributes of several currently used methods and their strengths and limitations; (ii) information about the characteristics of various postmarketing safety databases with which these tools can be deployed; (iii) analytical considerations for using safety data-mining methods and interpreting the results; and (iv) points to consider in integration of safety data mining with traditional pharmacovigilance methods. Perspectives from both the FDA and the industry are provided.

* Currently with Amylin Pharmaceuticals.
** Retired from the US FDA.
*** Currently with Takeda Global Research and Development.

Data mining is a potentially useful adjunct to traditional pharmacovigilance methods. The results of data mining should be viewed as hypothesis generating and should be evaluated in the context of other relevant data. The availability of a publicly accessible global safety database, which is updated on a frequent basis, would further enhance detection and communication about safety issues.

The term 'data mining' refers to the use of computerised algorithms to discover hidden patterns of associations or unexpected occurrences (i.e. 'signals') in large databases. These signals can then be evaluated for intervention as appropriate. Information gained from data-mining analyses can generate hypotheses that can be validated by other means.

Large postmarketing drug safety databases are the key data source currently used for drug safety data mining. Analysing these data is challenging because these voluntary reporting systems are subject to the problems of under-reporting, various reporting biases and incomplete, unverified data. The number of drug safety databases is also growing rapidly, with some databases containing millions of records. The application of computerised algorithms offers the opportunity to analyse these large databases in a timely and consistent manner. This paper will discuss the role of data mining in pharmacovigilance.

1. History and Mission of the Working Group

In the last 5 years, regulatory agencies and drug monitoring centres have been developing computerised data-mining methods to better identify reporting relationships in spontaneous reporting databases that could signal possible adverse drug reactions. Some pharmaceutical manufacturers are now using these methods via several commercial applications that have become available. However, at present there are no guidelines or standards for the use of these methods in routine pharmacovigilance.

In 2003, we formed a collaborative working group of statisticians, pharmacoepidemiologists and pharmacovigilance professionals from both the US pharmaceutical industry and the US FDA. Individuals from the industry serving on the Working Group are not official representatives of their organisations. The mission and goals of the Pharmaceutical Research and Manufacturers of America-FDA Collaborative Working Group on Safety Evaluation Tools are:

- to develop a consensus view of best practices to optimise the use of data mining in pharmacovigilance and risk management;
- to better understand the databases used for data mining, including data quality issues, and the optimal configurations and specifications for various uses;
- to better understand the possibility of assessing the performance characteristics of various data-mining methods in the drug safety arena for which no true and established gold standards exist;
- to understand the strengths and limitations of these methods, particularly as they affect the interpretation of results;
- to create opportunities for the FDA and industry to develop a common language, to share systematic approaches to the detection and assessment of safety signals from postmarketing adverse event (AE) data and to improve communication regarding data-mining issues;
- to communicate this information to industry and regulatory colleagues.

The use of data mining in pharmacovigilance is a complex topic and the organisations represented on the Working Group are at different stages of use and acceptance of these methods. Data mining in pharmacovigilance is also an evolving science and there was often lack of agreement among group members regarding preferred methodologies, signal definitions and even whether some of the references cited had adequate data to support the claims that were made. For these reasons, developing a consensus view of best practices was not always possible. Hence, this paper will present a spectrum of views on the uses of these techniques and how they fit into the pharmacovigilance 'toolbox'.

2. The Role of Data Mining in Pharmacovigilance

The role of data-mining methodologies in pharmacovigilance is evolving. Evaluating the value and utility of these methods to the pharmaceutical industry and regulators is a work in progress. The Working Group believes that *potential* values of safety data mining include the following.

- Systematic, automated and practical means of screening large datasets.
- Better utilisation of the large safety databases maintained by the FDA, the WHO and other organisations.
- Improved efficiency by focusing pharmacovigilance efforts on key reporting associations.
- Positive contributions to public health by identifying potential safety issues more quickly and/or more accurately than traditional pharmacovigilance methods.
- Better decision support for the pharmaceutical industry and regulators because of broader insight/knowledge of drug safety.

It is important to state at the outset of this discussion that all of the data-mining methods discussed in this paper identify *observed reporting relationships* between drugs and events in large safety databases. These reporting relationships are based solely on the frequency with which drugs and events are reported and thus cannot prove or refute causal relationships between drugs and events. Reporting relationships identified by data-mining methods must be viewed as *hypotheses* regarding *possible* causal relationships between the drugs and events of interest, when observed in the appropriate clinical contexts. Subsequent detailed clinical case reviews and other investigations, as appropriate, are necessary to explore hypotheses generated from data mining.

Data mining has the potential to clarify the many complex interdependent factors (e.g. concomitant drugs and/or diseases) that can play a role in the development of AEs in a clinical setting. Traditional methods may not always be able to detect these complex relationships. Drug exposure data and background rates of AEs of interest are often difficult to obtain systematically;[1-3] thus, it is often difficult for a safety evaluator to put counts of reported events in context. For some serious events, the reporting of one or two such events should prompt a review, regardless of context. For non-serious events and for serious events known to occur in the patient population for a variety of reasons (including exposure to drugs), it is less straightforward to define a threshold for the number of reports that should necessitate a review. For a pharmacovigilance department with many drugs to monitor, a comparative measure of reporting frequency (as provided by data mining) may be seen as an improvement over crude frequency counts and may aid in identifying potential safety issues and prioritising work. Data mining may also add value by detecting disproportionalities involving multiple drugs or multiple events that would be too difficult to detect by traditional methods.

Potential limitations of data mining include those inherent to postmarketing safety databases (e.g. under-reporting, reporting biases) that no signal detection method is likely to overcome. There are published examples of known safety issues that are not retrospectively identified by data-mining methods using predefined thresholds; this is not surprising since not all safety issues emerge from spontaneous reports.[2,4,5] There are concerns that, in some situations, data mining may generate more signals than can be followed up effectively with available resources. In this case, focus might be directed to signals with the greatest public health impact and seriousness. There is also concern about the lack of systematic, objective validation of the methods, a problem that also exists for traditional pharmacovigilance methods. Unfortunately, efforts to validate data-mining methods (and traditional methods) are complicated by the absence of a gold standard for identifying true drug toxicities, although various imperfect reference standards may be used to obtain useful insights on the performance of any method (see section 4.2.2). For this and other reasons, it has not yet been practical to evaluate data-mining methods or traditional methods using performance criteria generally accepted for screening and diagnostic tests.

The Working Group believes that data mining has a place in the pharmacovigilance toolbox but acknowledges that more work is needed before that place is fully defined. Systematic evaluation using traditional and data-mining methods with large

databases will be needed to determine if the promise of the methods actually pays off in practice. Hence, the intent of this paper is to review the current use of these methods for quantitative signal detection, to briefly highlight uses other than signal detection, to share the insights and experiences of Working Group members and to stimulate further discussion and investigation into the utility of data mining in pharmacovigilance.

3. The Need for Consistent Terminology

There is a lack of consistency of terminology in the quantitative signal detection literature. The WHO defines a signal as "reported information on a possible causal relationship between an adverse event and a drug, the relationship being unknown or incompletely documented previously. Usually, more than a single report is required to generate a signal, depending on the seriousness of the event and the quality of the information."[6,7] More recently, the report of the Council for International Organizations of Medical Sciences (CIOMS) VI project offered the following definition for signal: "a report or reports of an event with an unknown causal relationship to treatment that is recognised as worthy of further exploration and continued surveillance".[8]

Since these definitions do not specify the *type* of information that constitutes a signal, it is reasonable to view a signal as any information, qualitative or quantitative, that prompts further investigation of the relationship between a drug and an event.

In the context of data mining, some authors use the terms 'association' and 'signal' interchangeably. The term 'signal' is often defined in terms of the quantitative association alone. Others distinguish a signal as an association that has additional supportive clinical information.[1,2,7,9]

The Working Group believes that the consistent use of terminology related to data mining would facilitate communications among pharmacovigilance practitioners. We encourage those who generate, report, publish and/or present data-mining analyses to provide clear, unambiguous definitions of such terms, as these definitions are critical to understanding and evaluating the results. The definitions for terms used in this paper are given in table I.

Table I. Definitions of terms used in this paper

Term	Definition
Drug-event pair	Refers to the co-reporting of a drug and an event in a case report
Association	A relationship between a drug and an event, irrespective of the strength of the relationship. The presence of a *reporting* association between a drug and an event is *purely statistical* and by no means implies a direct, or even an indirect, causal relationship
Signal	A relationship between a drug and event that is strong enough, using a predefined threshold or criteria set by an analyst, to warrant further evaluation
Signal 'score'	A number reflecting the 'strength' of a reporting association, i.e. by how much the observed frequency differs from 'expected'. 'Expected' can be defined in various ways, depending on the criteria that are set for the analysis. There are several methods for computing a signal score
Quantitative signal detection	Refers to computational or statistical methods used to identify drug-event pairs (or higher-order combinations of drugs and events) that occur with disproportionately high frequency in large safety databases
Reporting proportion	The reporting proportion for a specific time period is defined as the number of adverse event reports containing both the target drug and the target event divided by the total number of adverse event reports for the target drug over the same period of time (see also section 4.1)
Reporting ratio	The reporting ratio corresponding to the target drug and the target event over a defined time period is equal to the reporting proportion for the target drug and target event divided by the marginal reporting proportion for the target event. The marginal reporting proportion is equal to the total number of reports for the target event divided by the total number of reports in the database. The marginal reporting proportion for the target event may be computed using all of the reports or using only those reports that do not mention the target drug (see also section 4.1)
Safety data mining/ disproportionality analysis	The application of computer-assisted computational and statistical methods to large safety databases for the purpose of systematically identifying drug-event pairs reported at disproportionately high frequencies, relative to what a statistical independence model would predict. 'Safety data mining' and 'disproportionality analysis' are used interchangeably in this paper

4. Overview of Data-Mining Methods used for Quantitative Signal Detection

4.1 Statistical Principles

Many descriptions of data-mining methods, as applied to pharmacovigilance, are available.[2,10-13] The most commonly used methods with the greatest published experience are the proportional reporting ratio (PRR)[9,14] and the reporting odds ratio (ROR),[15-17] as well as Bayesian[18,19] and empirical Bayesian[20,21] methods that account for the variability associated with small report counts. All of these methods identify statistical *associations* between drugs and events in the reports contained in the spontaneous reporting databases. These associations are based solely on the frequency with which drugs and events are reported together and thus must be viewed as *hypotheses* regarding possible causal relationships between the drugs and events of interest, recognising that there are many possible reasons other than a direct causal relationship for the observed association.

The quantitative evaluation of the relative frequency of reports in spontaneous reporting databases that mention both a particular target drug and a particular target AE is based primarily on the entries in table II.

Thus, of a total of T reports in the database, M mention the target drug, N mention the target AE, A mention both the target drug and the target AE, C mention the target AE but not the target drug, and D mention neither the target drug nor the target AE. The reporting ratio (RR) [equation 1]

$$RR = \frac{A}{\text{Expected}(A)} = \frac{A}{MN/T} \quad \text{(Eq. 1)}$$

measures relatively how much more or less often reports in the database actually mention the target drug and target AE than would be expected if the mention of either was statistically independent of whether the other was mentioned or not. Under this assumption of independence, MN/T reports would be expected to mention the target drug and target AE.

The expected value of A can be expressed in other ways, giving rise to statistics similar to the RR.

Table II. Number of reports mentioning a target drug and target adverse event (AE)

No. of reports	Target AE	All other AEs	All AEs
Target drug	A	B	M
All other drugs	C	D	T – M
All drugs	N	T – N	T

The PRR is obtained when Expected(A) = MC/(T − M). The expected likelihood that a report mentioning the target drug also mentions the target AE [i.e. Expected(A)/M = C/(T − M)] for the PRR is based on the target AE reporting proportion among reports not mentioning the target drug. In contrast, the expected likelihood for the RR [Expected(A)/M = N/T] is based on all reports, including those mentioning the target drug.

If many reports mention the target drug and many reports mention the target AE, then there will not be much uncertainty associated with Expected(A). However, if the target drug has not been on the market long (i.e. M is small) or if the target AE is rare (i.e. N is small), then there may be considerable uncertainty about Expected(A) that should be accounted for when any of the statistics are interpreted. Methods have been described for doing so, based on Bayesian[18] and empirical Bayesian[20,21] principles. Software is available to carry out the calculations. Gould[10] provides a detailed comparison of the two approaches.

4.2 Performance Characteristics

4.2.1 Overview

Pharmacovigilance professionals and institutions contemplating the use of data-mining methods to supplement their traditional[22] or manual methods of signal detection should consider a number of factors, including which method to use, which database(s) to use and the choice of variable configurations that can be specified for each data-mining run. Research on data-mining algorithms is dynamic, with new methods under development. This report focuses on the most commonly cited disproportionality methods: PRR, ROR, Bayesian methods and empirical Bayesian methods. Other methods that have been or are being developed based on various statistical algorithms/techniques include probability filtering ('PROFILE'),[23] fuzzy

logic,[24,25] sequential probability ratio testing[26,27] and a tree-based scan statistic.[28]

4.2.2 Challenges in Assessing Performance

Many methodological issues complicate a systematic and comprehensive assessment of the performance of the methods discussed in this paper. These issues include: the variety and volume of data populating spontaneous reporting databases; variations in database environments/architectures; the lack of standards for adjudicating causality and expectedness; the lack of a gold standard with which to calculate traditional screening or diagnostic metrics (i.e. predictive values, sensitivity and specificity); and the lack of clear guidelines on desirable performance characteristics in pharmacovigilance.

A major difficulty arises in trying to evaluate how well any pharmacovigilance method identifies toxicities that truly are causally associated with drugs. The language of diagnostic evaluation is intuitively appealing for this purpose. The truth table (table III) provides explicit definitions of terms used in diagnostic evaluation.

In the context of signal detection, a 'false-positive' finding would be a disproportionately high frequency of drug-event reports that is shown, by other means, to represent an artifactual relationship. Similarly, a 'false-negative' finding would be a drug-event association that does reflect a causal relationship but is not disproportionately reported or is reported less frequently than expected, based on all other drug-event associations in the database.

Unfortunately, the lack of an objective measure of 'truth' (a gold standard) or evidence of a true causal relationship makes the evaluation and validation of all signal detection methods in terms of sensitivity, specificity and predictive value difficult. Given that there are no perfect gold standards, some authors have attempted to validate these methods using imperfect gold standards ranging from selected events in product labelling,[9,19] to selected published information from epidemiological studies and/or reports of positive rechallenge,[29] to labelled events updated by information from large clinical trials.[2]

Table III. Truth table for assessing signalling

Signal	True causality	
	Yes	No
Yes	a	b
No	c	d

Negative predictive value = true negative signal/negative signa = d/(c + d).

Positive predictive value = true positive signal/positive signal = a/(a + b).

Sensitivity = true positive signal/true causal = a/(a + c).

Specificity = true negative signal/true noncausal = d/(b + d).

One disproportionality measure cannot be judged better (or worse) than another because it yields a 'signal' (i.e. exceeds its arbitrary critical value) more often in the absence of a well accepted gold standard. Sensitivity can always be increased at the expense of specificity and *vice versa*. More experience is necessary over a broad spectrum of potential drug-event relationships, algorithms and threshold metrics in 'real life' pharmacovigilance settings to provide a better idea of the diagnostic potential of disproportionality measures. However, the internal validity of these methods is suggested by their ability to reliably detect relationships that are already known. This is reassuring, because failure to recognise these relationships would suggest the possibility of a high false-negative rate and would seriously compromise the value of exploring spontaneous reporting databases for early detection of potential toxicity issues.

Brief summaries follow in section 4.2.3 of published efforts to validate or describe the utility of the most commonly used methods. The Working Group believes that much work remains to be done in this area.

4.2.3 Published Evaluations of Performance

Proportional Reporting Ratio

Evans et al.[9] used the Adverse Drug Reactions On-line Information Tracking (ADROIT) database, the postmarketing safety database of the Medicines and Healthcare products Regulatory Agency (MHRA, formerly the MCA) in the UK. They evaluated 481 'signals' for 15 drugs in the database that were identified using the PRR and found that 339 (70%) were recognised adverse reactions (per labelling), 62 (13%) were events considered likely to be related to the underlying disease and 80 (17%) required further evaluation. Of the 80 events requiring evaluation, 22 warranted a detailed review (leading

to requests for labelling changes for three events), 22 were to be kept under continuing review and no further action was planned for the remaining 36 events. The MHRA determines which signals identified by the PRR need further follow-up in terms of four factors (called SNIP criteria): Strength of signal; whether the signal is New or not; the clinical Importance; and the potential for Preventive measures. More recently, the MHRA has piloted a scoring system to assess which signals require detailed evaluation.[30] In this system, signal strength (PRR value) is one of three factors used to compute an 'evidence score' that is plotted against a 'public health score' to assess the potential importance of the signal. Preliminary evidence suggests that this innovation may be useful for systematising the evaluation process and aiding scientific discussion.[31,32]

Reporting Odds Ratios

The ROR (A/B C/D or AD/BC, see table II) has been described in the pharmacovigilance literature as an additional analytical approach for disproportionality analysis of spontaneous data.[16,17] The ROR, like the traditional odds ratio, is an estimate of the incidence rate ratio; it estimates the odds of the AE in those exposed to a particular drug divided by the odds of the AE occurring in those not exposed to drug. The ROR is not affected by general underreporting for a specific drug or a specific event.[16] Rothman et al.[33] have proposed that the ROR may, in theory, be a less biased methodology than other disproportionality methods in that a series of spontaneous reports can be viewed as cases and controls; the 'cases' are those experiencing a specific AE, the 'controls' are those that do not experience the AE (in a spontaneous reporting database that would be those with 'other AEs') and the 'exposure' is exposure to the specific drug under study. However, others believe that in practice, both the PRR and the ROR yield similar results and that there is no benefit in using the ROR instead of the PRR.[17,34]

Bayesian Confidence Propagation Neural Network

The Uppsala Monitoring Centre (UMC) uses the Bayesian Confidence Propagation Neural Network (BCPNN) software to identify drug-event pairs that stand out statistically from the background of all reports in the database. Members of the UMC international expert assessment panel then decide which of these constitute potential drug-AE relationships that should undergo more detailed evaluation. Bate et al.[18] described how the system could have detected the relationship between captopril and cough before it was widely reported in the literature and provided an example of false-positive signal avoidance. Lindquist et al.[7] checked case reports for critical terms published in *Reactions Weekly* from January to June 1998 against the WHO database for the same time period. They found that 12 of 43 pairs appearing as 'first reports' in *Reactions Weekly* fulfilled the criteria of association in the WHO database using the BCPNN system at the same time as, or before, appearing in the publication. Lindquist et al.[19] also described a retrospective evaluation where the 'gold standard' was whether or not the signalled association was described in the reference literature (*Martindale's Extra Pharmacopoeia* and the *Physicians' Desk Reference*) at a given point in time or whether the association was confirmed or strengthened over a specified period. In this evaluation, the BCPNN detected signals in the WHO database with a 'positive predictive value' of 44% and a 'negative predictive value' of 85%. The BCPNN identified six of the ten signals produced by the former system used at the UMC, four of the six being detected earlier than with the former system.

Recently, the UMC has introduced triage logic to further filter the large number of associations that are generated by the BCPNN and sent to reviewers for evaluation.[35] The filters are applied to the combinations database produced by the BCPNN scan to reduce the number of combinations highlighted for assessment and to help focus on the areas of greatest importance. The filters currently in use highlight rapid increases in reporting, serious reactions with new drugs and reactions of special interest such as those very likely to be drug related. After using these filtering strategies for some time, the UMC plans to evaluate how successful they have been in finding 'potential signals' and to examine whether early detection of important signals has been enhanced.

Empirical Bayes (Gamma Poisson Shrinker, Multi-Item Gamma Poisson Shrinker)

Several retrospective studies in which empirical Bayesian methods detected early signals of AEs

have been published. An analysis utilising multi-item gamma Poisson shrinker (MGPS) showed a number of signals for rhabdomyolysis and renal failure for cerivastatin several years before this drug was removed from the market.[2] MGPS also showed important signals of various adverse drug events in paediatric and adult patients.[1,3] In a previous study, the gamma Poisson shrinker (GPS, a precursor to MGPS) was applied to 30 drug-event combinations declared as signals by FDA epidemiologists using traditional methods applied to the Spontaneous Reporting System (SRS) [now known as the Adverse Event Reporting System (AERS)] database.[2] The GPS method signalled all 30 of these selected drug-event combinations, with 20 signalled by GPS in the data collected 1–5 years before index cases were detected by traditional methods, nine signalled by GPS the same year and one signaled by GPS a year after the data were detected by traditional methods.

The GPS was also used to analyse the differences in time in detecting 160 drug-event combinations involving 85 drugs. These 160 drug-event combinations were coded as signals between 1985 and 1996 by FDA safety evaluators and collected in the FDA Monitoring Adverse Reports Tracking (MART) system. These 160 drug-event signals were selected for data-mining analysis because the drug-event pair names in the MART matched the drug-event pair names in the SRS. GPS signaled 97 drug-event combinations in the SRS data collected 1–4 years before they were entered as signals in the MART system, with 36 signaled by the GPS the same year and 27 signaled by the GPS 1–3 years later. One-half of the 27 signals detected later by GPS included designated medical events (such as severe liver events, Stevens-Johnson syndrome, aplastic anaemia and anaphylaxis) that are easier to characterise with fewer reports.[2,36]

These studies illustrate the potential for the GPS/MGPS to detect signals of drug-event combinations that have been declared to be signals by traditional methods.

Comment on the Comparative Performance of Data-Mining Methods

There are no published, large-scale, systematic comparisons of data-mining methods currently used for pharmacovigilance and the published performance characteristics vary, depending on criteria selected for signal detection. Although the precise statistical approaches of the methods differ, they all involve some assessment of disproportionality and would therefore be expected to provide overlapping results. Some investigators have shown concordance among methods when the number of reports exceeds four.[17] It has also been observed that Bayesian and empirical Bayesian methods generate fewer signals than the PRR when commonly cited thresholds are used. This is to be expected since both Bayesian and empirical Bayesian methods make adjustments for the increased variability associated with small actual and expected report counts. It should be noted that the number of drug-event pairs signaled by any of the available methods depends in large part on selection of empirical signal thresholds that involve subjective judgements. By itself, the number of signals flagged by a particular method is an inappropriate criterion for comparing the performance of data-mining algorithms. As discussed previously, all methods incur tradeoffs between sensitivity and specificity, especially when varying criteria for eliciting signals are used.[2-4,37-41] When examining the literature on performance analyses of drug safety data mining, it should be borne in mind that sensitivity and specificity are highly dependent upon the definition of signal thresholds used, the minimum number of reports required before a signal is declared, the number of relevant event codes analysed, the type of data configurations utilised (reports from manufacturers versus reports from all other sources, etc.) and many other factors. There is no basis at present for recommending any of these methods and signal thresholds as superior for all users and situations.

5. Postmarketing Safety Databases used for Quantitative Signal Detection

Databases utilised in drug safety data mining include large postmarketing databases maintained by regulators, pharmaceutical manufacturers and various consortia, each with their own reporting criteria, coding dictionaries and data entry rules. Extracting meaningful data from these databases is often challenging because voluntary reporting systems are subject to the problems of under-reporting, various reporting biases and incomplete, unverified

data. Some of these databases are quite large, containing hundreds of thousands and even millions of AE reports. Despite their inherent limitations, the size and scope of these databases make them appealing for pharmacovigilance.

The Working Group acknowledges that this situation of differing databases is far from ideal, as results may vary between databases. Ideally, there would be a single 'canonical' database available to industry and regulators in real time; such a database would contain worldwide data on all products, have no duplicate reports and employ consistent conventions for drug naming, event coding and data entry. The reports submitted to such a database would be complete and include treatment indication, past medical history and co-medications.

Until such a 'canonical' database exists, essentially three databases are available to pharmaceutical manufacturers for signal detection activities: (i) their own internal safety database(s); (ii) the FDA's public-release safety databases (SRS/AERS and the Vaccine Adverse Event Reporting System [VAERS]); and (iii) the database of the WHO International Drug Monitoring Programme. These databases are described and discussed in sections 5.1–5.4. Regulators typically rely on their own agency databases for signal detection activities.

5.1 US FDA Adverse Event Reporting System (AERS) and Spontaneous Reporting System (SRS)

AERS is the FDA's postmarketing safety database and is used herein to refer to the combined datasets of SRS (1968 to October 1997) and AERS (November 1997 to present). The public-release version of AERS is available for purchase from the National Technical Information Service (NTIS) on a quarterly basis [1] and, as of December 2004, from the FDA's website beginning with the January 2004 quarterly data (http://www.fda.gov/cder/aers/default.htm). AERS is a surveillance system that relies on voluntary reporting of AEs to the FDA by healthcare professionals and consumers, as well as required reporting by pharmaceutical manufacturers. AERS includes spontaneous reports from US sources, serious and unlabelled spontaneous reports from non-US sources and serious, unlabelled and attributable postmarketing clinical trial reports from all sources. As of December 2004, AERS contained approximately 2.6 million reports. The size and diversity of this database are its primary advantages.

At present, there are several important limitations in using the public-release version of AERS data. Although historically there has been a lag time of 9–12 months for release of data through the NTIS, the FDA anticipates that this interval will shorten.

Another limitation of the database is the presence of duplicate and multiple reports for some cases. NTIS data contain all reports received by the FDA in AERS. Multiple reports of the same case are generated from updates by the manufacturers to previously submitted original reports. Potential duplicate reports of the same case are generated from reports by multiple manufacturers and 'direct' reports received from healthcare providers and consumers via the FDA's MedWatch programme. The multiple reports are linked by the manufacturer's control number in the internal FDA database only, leaving the public-release database with potential duplicates and multiple reports for a case. Although there are no plans at present to remove duplicate or multiple reports from the public-release version of AERS, commercial vendors provide versions of AERS that have been 'cleaned' by consolidating multiple and duplicate reports. However, there may be differences between datasets provided by various vendors because of the use of different duplicate detection and removal algorithms.

The AERS database also lacks standardisation of drug names. The FDA attempts to link the reported verbatim drug name to an 'active ingredient'. Correction or standardisation of names of combination products, different formulations, herbal products, foreign drug products and spelling errors is necessary to consolidate spellings and product names to

1 Access to the entire AERS database (public-release version) is available from commercial vendors on a subscription basis; these vendors offer the service of collecting all of the updates issued by the NTIS into one repository. Each vendor uses its own rules and algorithms to 'clean' the database by standardising drug names, accommodating changes to coding dictionaries and removing duplicate reports. Selection of a vendor may require an evaluation period to ensure that the methods used to format and clean the public data are acceptable to users.

improve the fidelity of data mining. The FDA is participating in efforts to develop a global coding dictionary with ICH M5 (Data Elements and Standards for Drug Dictionaries) [http://www.ich.org]. Commercial vendors provide access to datasets in which drug names have been organised according to their respective methods.

Reports from outside the US that are present in AERS are likely to be serious, unlabelled events, whereas both labelled and unlabelled events, regardless of seriousness, are present in AERS for reports from within the US. Therefore, it is possible that signals could be generated for drugs with a high proportion of foreign to domestic reports because most of the foreign reports received for these drugs are only for serious, unlabelled events. This can be addressed by stratifying on reports by foreign/domestic submissions, using separate analyses or analysing company databases.

The types of case reports that are being entered into AERS are changing over time. Since data-mining algorithms derive the frequency of 'expected' drug-event pairs used as the denominator from the total AERS database, understanding the impact of changes in the composition of AERS is vital. The electronic submission of reports and the availability of waivers for submission of non-serious, expected events are two examples of changes that have influenced the content of AERS. Since 1998, non-serious, expected reports for drugs marketed for ≥3 years have not been entered into AERS if they were submitted on paper. However, electronic submissions may be directly imported into AERS, although at this time most are still undergoing the FDA quality control process before being entered into AERS. As electronic submissions increase, this 'shift' in the content of the AERS database has the potential to modify the database in a favourable way so that all reports, regardless of their labelledness, will be entered, thus creating a more complete and coherent dataset. The effect of these changes on the results of data-mining analyses is unknown.

5.1.1 Working Group Recommendations Regarding AERS

The members of the Working Group, although understanding the financial constraints of the FDA, believe that improvements to the public-release version of AERS would enhance the utility of the database. Toward this end, the Working Group has made several recommendations concerning the NTIS AERS product, some of which have already been addressed by the FDA.

- Decrease the lag time between report receipt by the FDA and the public release of AERS.
- Publish the entry specifications and coding conventions to enhance understanding of the data.
- Make available as many data fields (including narratives) as possible without infringing on patient privacy.

5.2 US FDA Vaccine Adverse Event Reporting System Database

In the US, surveillance of AEs after vaccination is undertaken by the government using VAERS, which the FDA and the Centers for Disease Control and Prevention (CDC) jointly manage. VAERS is the national system for surveillance of AEs after vaccination. It was initiated by the 1986 National Childhood Vaccine Injury Act and was established in 1990. The uses of VAERS include detecting novel AEs, monitoring the frequency and severity of known AEs, identifying possible risk factors, and vaccine lot surveillance.

VAERS is substantially smaller than AERS, receiving 10 000–15 000 reports per year on top of approximately 160 000 existing reports. AE data in these reports are coded using the Coding Symbols for Thesaurus of Adverse Reaction Terms (COSTART) dictionary. Some reports are submitted directly to VAERS and are therefore not in manufacturers' databases. Because vaccines are more likely to be given to children than adults, and to healthy rather than ill people, the VAERS database contains a predominance of reports involving children.

5.2.1 Systems of Administration and Reporting

The system for administering vaccines and reporting AEs is different than the system used for drugs. Drugs are primarily administered by licenced practitioners by prescription in a healthcare system focused on treating illness. While vaccines are sometimes individually prescribed by practitioners, they are often given based on public health guidelines as part of a systematic disease prevention programme, which might include physician's offices,

public health clinics, and the military. Similarly, AEs are reported from each part of this system at different rates that might influence the disproportionality calculated by various algorithms. For example, the recent smallpox vaccination campaign was limited to the military and certain public health workers. Both the military and the CDC had safety surveillance systems in place that went beyond VAERS, although all reports of AEs were submitted to VAERS. Interpretation of data-mining analyses that compared the smallpox vaccine with other adult vaccines would need to take this differing reporting mechanism, and the possibility of higher reporting rates, into consideration.

5.3 WHO Safety Database

The WHO safety database is a large, global database with >3.4 million individual case reports spanning >30 years (1968 to present). AE reports are contributed by national centres participating in the WHO International Drug Monitoring Programme.[42]

There are currently 78 member countries that submit domestic AE reports to the WHO database, ideally on a quarterly basis. A significant proportion of the WHO database comprised reports from the AERS (US) database. The top five contributors by number of reports received since joining the programme are the US (1 314 525), UK (391 868), Germany (160 648), Australia (146 116) and Canada (136 192). Only domestic cases from the US are entered. Differences in reporting requirements between countries should be considered in an analysis of this database. Some of the differences between countries relate to whether reports are voluntary or mandatory, or whether consumer reports are accepted. Furthermore, reporting rates and profiles may also be influenced by differences in medical practice and societal factors.

The report sources include healthcare professionals, consumers and marketing authorisation holders. Most consumer reports are from the US. Duplicate cases are identified by a systematic check for the same case ID and by analysis of case series. AEs are coded using the WHO-Adverse Reaction Terminology (WHO-ART) coding dictionary and drugs are coded using the WHO Drug Dictionary, which offers indexing and retrieval of drugs by the hierarchical Anatomical Therapeutic Chemical (ATC) classification.

Strengths of the WHO database include the capability to evaluate drugs by generic or trade name, the capability to identify between-country differences and the capability to identify well documented reports via a quality grading system. As with AERS, limitations of the WHO database include a limited systematic process for identification of duplicates, many empty data fields and the unavailability of case narratives.

Access to the WHO database is available by subscription either directly from the WHO or through commercial vendors. As with other databases, selection of a vendor may require an evaluation period to ensure that the vendor's methods for formatting the data are acceptable to users.

5.4 European Medicines Agency EudraVigilance Database

The EMEA created and maintains a pharmacovigilance database management system and data processing network known as EudraVigilance. EudraVigilance was created for the electronic exchange and processing of AE reports involving medicinal products authorised in the European Economic Area (EEA). It offers remote access to registered partners and their administrative and scientific users in the European Commission, the EMEA, Competent Authorities in the EEA and pharmaceutical companies via a secure connection over the internet. EudraVigilance contains both a clinical trial (EVCTM) and a post-authorisation module (EVPM). The EVPM was established in December 2001 to support reporting requirements for spontaneous reports and adverse reaction reports originating from organised data collection systems (e.g. registries and post-authorisation safety studies). As of July 2005, 118 791 Individual Case Safety Reports (ICSRs) corresponding to 70 901 cases were reported from outside the EEA to the EVPM and 60 326 ICSRs corresponding to 35 649 cases were reported to the EVPM from within the EEA. The latter reporting activity originated from 47 market MAHs and 15 member states. Retrospective electronic population of EVPM with legacy data is underway. EudraVigilance includes data analytical capabilities and quantitative signal detection func-

tionality based on PRRs and RORs. At present, pharmaceutical companies have restricted access to EudraVigilance in that each can only view AE reports that they have submitted to EMEA.[43]

5.5 Company Safety Databases

Although it is technically feasible to use data mining with in-house company safety databases, there are a number of caveats to consider. Although there are no precise guidelines, the database should be of sufficient size and diversity to serve as a suitable 'background' for evaluating disproportionate reporting. Among the potential limitations of company databases are a relative lack of diversity of events or drugs, which leads to a greater likelihood of masking (see section 6.6).[10] One way to measure diversity in a safety database is to examine the number and distribution of reports in the database by therapeutic area or drug product. It may be prudent to also compare the results of analyses using the proprietary database with those obtained using AERS or WHO for several 'well characterised' products. However, interpretation of the clinical impact of such 'diversity' or lack thereof is complicated by the often cited lack of gold standards.

Each institution's proprietary database may have strengths that can be exploited, for example companies with a global dataset, rather than one weighted toward US cases (as AERS is) may find this useful. If the company's database started earlier than 1968, when SRS/AERS started, it may be possible to explore relative reporting frequencies for older drugs. There may be more data elements with consistent data quality and coding, which may allow for further exploration of relative reporting frequencies among demographic subsets. Importantly, because the data are not subject to the delays associated with the public databases, company databases may allow for earlier detection of safety signals, particularly for new products.

6. Analytical Considerations

6.1 Overview

The essential first step in undertaking an exploration of a spontaneous reporting database with data mining is to specify the purpose of the analysis. Depending on the questions specified, technical/analytical options that might be considered include:

- whether to include all drug-event pairs in the analysis or only those pairs where the role of the drug of interest was considered 'suspect';
- whether to base the calculations on counts of drug-event pairs or counts of reports;
- whether to perform the analysis using specific AE terms or groups of related AE terms that are aggregated under a 'higher-level term' with hierarchical AE dictionaries such as the Medical Dictionary for Regulatory Activities (MedDRA);
- whether to stratify the calculation of expected counts (see section 4.1) and, if so, by which variables.

These and other considerations are discussed briefly in the remainder of section 6.

6.2 Role of the Drug (Suspect Only versus Suspect and Concomitant)

Spontaneous AE reports originate from individuals who suspect they have experienced, observed or heard about an adverse drug reaction. A typical report will cite a drug(s) and an event(s) that the reporter believes are related. The reporter may mention other medications, but the reason for the report is the belief that an event is related to a particular drug or drugs. In order to capture this distinction, safety databases such as AERS and proprietary databases maintained by pharmaceutical manufacturers typically classify each drug cited in a report as 'suspect' or 'concomitant'.

The drug and event information in a safety database can be thought of as a very large two-way table composed of many cells. The number in any given cell is the number of reports containing the drug and the event that define that cell. Obviously, the value in a particular cell will be different depending on the choice to only count reports where a drug is coded as suspect (S only) versus all reports containing the drug irrespective of coding as suspect or concomitant (S + C).

The experience of some Working Group members, using empirical Bayesian methods, is that computed relative RRs are slightly higher with 'S only' compared with 'S + C' and that a greater number of drug-event pairs meet or exceed a numerical 'signal'

threshold with 'S only' compared with 'S + C'. However, there are instances where the reverse is true. The Working Group is not aware of any data to date that demonstrates that the differences are clinically important. It would be reasonable to expect that other methods would produce similar results.

There is no reason to believe that either strategy is superior with respect to identifying or not identifying reporting relationships that may turn out to reflect causal relationships. As users gain familiarity with the performance of these methods, they may need to adjust data-mining strategies accordingly.

6.3 Counts of Drug-Event Pairs versus Counts of Reports

Another factor to consider in the implementation of these algorithms is whether the statistical parameters should be calculated with respect to the total number of drug-event pairs or the total number of reports in a given database. Calculations appear to be based on numbers of drug-event pairs in the paper by Evans et al.[9] describing PRR and in the paper by DuMouchel[20] on empirical Bayesian methods. Bate et al.[18] and DuMouchel and Pregibon[21] base their calculations for Bayesian and empirical Bayesian methods, respectively, on numbers of reports. Any of these methods can be executed either way. Statisticians in the Working Group note that either approach is acceptable, although counting reports probably provides a more intuitively appealing estimate of sample size, since reports often contain multiple events and multiple drugs that are not independent of each other. Presently, there are no data demonstrating that the choice of denominator is clinically important.

6.4 Combining Drug and Event Terms

The validity and utility of combining (pooling, 'lumping', collapsing) drug and/or event terms in the setting of safety data mining is not well studied. Combining drugs of the same class or medically related AE terms may allow earlier detection of safety issues by increasing the power of the analysis through larger numbers. This is of particular interest for databases that encode AE data using highly granular dictionaries such as MedDRA. Combining drug or event terms must be done carefully because, as the following example shows, relative RR values can decrease when terms are combined.

Table IV provides the number of reports mentioning the target drug and either of two synonyms for the target AE.

The RR for the target drug using the first synonym is (equation 2):

$$RR_1 = A_1 \div \frac{(M \times N_1)}{T} = A_1 \div E(A_1) \quad \text{(Eq. 2)}$$

The RR for the target drug when the counts for the two synonyms are combined (assuming, of course, that no report ever mentions both synonyms) is (equation 3):

$$RR_c = (A_1 + A_2) \div M \times \frac{(N_1 + N_2)}{T} \quad \text{(Eq. 3)}$$

It is easy to show that the RR value when the synonyms are combined is greater than the value using only the first synonym ($RR_c > RR_1$) if and only if (equation 4):

$$\frac{A_2}{N_2} > \frac{A_1}{N_1} \quad \text{(Eq. 4)}$$

that is, if and only if the target drug is mentioned more often among the reports that mention the second target AE synonym than among the reports that mention the first synonym.

Another consequence of this demonstration is that the increase of RR with the combination over the reporting with the first synonym implies a decrease of RR with the combination related to the RR with the second synonym. This effect on RRs is not necessarily a bad thing. A few reports of a rare event can lead to a very large, but very imprecise, RR value. A high RR value is not meaningful by itself. It takes whatever meaning it might have only when considered in the right context. Combining synonyms will decrease the value obtained for some of

Table IV. Number of reports mentioning a target drug and either of two terms for a target event, assuming no report mentions both adverse event (AE) terms

No. of reports	Target AE 1	Target AE 2	Other AEs	Total
Target drug	A_1	A_2	$M - A_1 - A_2$	M
Total	N_1	N_2	$T - N_1 - N_2$	T

the synonyms, but the ratio based on the combination will become more precise and perhaps more medically relevant.

Combined terms should be highly similar or synonymous to minimise the risk of distorting a result (e.g. QT prolonged and corrected QT [QT_c] prolonged). The driving considerations must be the medical meaning and coding practices, not the statistical consequences. For example, the specific term 'torsade de pointes' should not be combined with the general term 'arrhythmia', because torsade de pointes is a highly specific type of arrhythmia and has different pathophysiological implications than most other arrhythmias. The same point can be made for combining an event such as torsade de pointes with other specific arrhythmia terms that are more likely than torsade to result from non-drug causes.

When planning an analysis with combined terms, the terms should be specified *a priori* and with careful consideration to the medical/scientific meaning of the combination and historical coding practices in the database. Repeated searching for a combination that 'works' may increase the false-positive rate because of multiplicity considerations.

6.5 Stratification

Stratification is a statistical procedure for mitigating the effects of confounding by adjusting for associations between a drug and a variable and an event and the same variable. For example, suppose that drug A is frequently prescribed for men aged >60 years and event B is common in men aged >60 years. Disproportionality analysis might detect a strong association between drug A and event B when the true associations are between the drug and men aged >60 years and between the event and men aged >60 years. In this example, stratification of the computation of expected counts by age and sex removes the effect of confounding. Another commonly used stratification variable is year of report. Stratification by year of report reduces the chance of detecting spurious associations because of temporal factors that may influence the reporting of specific drugs and/or specific events. Some members of the Working Group have also found it useful, when concerned about effects from publicity that stimulates consumer reporting, to stratify on report source (i.e. consumer, healthcare provider). However, stratification will not adjust for over-reporting of a specific drug-event pair. One should be aware that many factors can stimulate reporting and these factors may extend across report sources.[44]

The effect of stratification can be illustrated with an example suggesting that sensible stratification generally should be used. Suppose that one has counts as in table V.

The value of the stratified RR is (equation 5):

$$RR_{str} = A \div \left(M_1 \times \frac{N_1}{T_1} + M_2 \times \frac{N_2}{T_2}\right) \quad \text{(Eq. 5)}$$

The value of the RR ignoring stratification is (equation 6):

$$RR_{uns} = \frac{AT}{NM} \quad \text{(Eq. 6)}$$

The difference between the unstratified and stratified ratios is proportional to (equation 7):

$$\left(\frac{M_1}{T_1} \quad \frac{M_2}{T_2}\right) \times \left(\frac{N_1}{T_1} \quad \frac{N_2}{T_2}\right) \quad \text{(Eq. 7)}$$

The stratified ratio is not necessarily greater than the unstratified ratio, nor is it necessarily less. If the target drug and the target AE are both mentioned more frequently in stratum 1 than in stratum 2, then the stratified ratio will be less than the unstratified ratio regardless of the values of A_1 and A_2. The stratified ratio also will be less than the unstratified ratio if the target drug and the target AE are both

Table V. Number of reports mentioning a target drug and a target event in each of two distinct subgroups (strata) of the patients providing reports

No. of reports	Stratum 1		Stratum 2		Combined	
	target AE	total	target AE	total	target AE	total
Target drug	A_1	M_1	A_2	M_2	$A = A_1 + A_2$	$M = M_1 + M_2$
Total	N_1	T_1	N_2	T_2	$N = N_1 + N_2$	$T = T_1 + T_2$

AE = adverse event.

mentioned less frequently in stratum 1 than in stratum 2. If the target drug is mentioned more frequently in stratum 1 than in stratum 2, but the target AE is mentioned less frequently in stratum 1 than in stratum 2 (or *vice versa*), then the stratified ratio will exceed the unstratified ratio.

If the within-stratum RR is actually the same in both strata, then the stratified ratio will equal the common within-stratum value. However, the unstratified ratio will usually differ from the common within-stratum value. Consequently, since unstratified estimates may present a distorted picture of reporting relationships, especially when RRs differ little among strata, it seems generally advisable to stratify sensibly.

Many potential stratification factors can affect the values of disproportionality measures based on data from spontaneous reporting datasets. The dataset may contain values for some of these, but will not contain values for many others because the information was not captured on the report form or is not recorded in an easily recoverable form. No analysis can stratify by all of the recognised factors, let alone the unrecognised ones.[45] The fact that the value of the RR (or any disproportionality measure) could be increased or decreased by stratification should be borne in mind during any analysis. Disproportionality measures should be computed for important subsets of the patients when there is reason to believe that potential toxicity risks may be particularly elevated in some, but not all, of these subsets (e.g. for elderly, but not non-elderly, patients).

The general consensus of the Working Group is that routine use of stratification for computing expected counts is a reasonable approach. Software programmes should be designed to provide alerts when unusual strata or data distributions exist.

6.6 Masking of Drug-Event Relationships by Experience with Related Drugs

The terms 'masking' and 'cloaking' have been used to describe the effects that experience with related drugs may have on the observed reporting relationships between a drug and various AEs. Masking is possible in any database but because the pharmacovigilance databases held by pharmaceutical companies are generally smaller and less diverse than regulatory databases, the former may be more vulnerable to these effects.

Table VI. Number of reports mentioning either of two drugs and a target event. The rows are not mutually exclusive because a report could mention both drug A and drug B

No. of reports	Target AE	All AEs
Drug A	A	M_A
Drug B	B	M_B
All drugs	N	T

AE = adverse event.

For example, if drug A is an angiotensin-2 antagonist and drug B is an ACE inhibitor that has been on the market longer than drug A, then the information accumulated about drug B may affect relative RR values for drug A.[10] Let us suppose that the reports can be summarised as shown in table VI.

The ratio of the RRs for drug A with and without reports mentioning drug B (which we assume never mention drug A) is (equation 8):[10]

$$RR_A^{(\text{excl B})} \div RR_A^{(\text{incl B})} = 1 + \frac{M_B}{N - B}\left(B/M_B - N/T\right) \quad \text{(Eq. 8)}$$

Clearly, if the reporting proportion for the target AE on drug B (B/M_B) is greater than the overall reporting proportion for the target AE (N/T), then the RR for drug A based on all of the reports will be less than the RR for drug A calculated after removing all of the reports mentioning drug B. Conversely, if the reporting proportion for the target AE on drug B is less than the overall reporting proportion, then the RR for drug A based on all of the reports will be greater than the ratio after removing the reports mentioning drug B. Because of this, and because the value of N/T may largely be determined by reports involving drugs other than drug A or drug B, no blanket recommendation can be given about whether the RR should be calculated including or excluding drug B. When most of the target AEs can be identified with drugs like drug A and drug B, then it may be advisable to compute the RRs both ways.

6.7 Signal 'Absorption' ('Innocent Bystander' Phenomenon)

Signal 'absorption', also known as the 'innocent bystander' phenomenon, occurs when a drug that is

commonly co-prescribed with another drug appears to be associated with an event that is actually associated with the other drug. This is a problem in polypharmacy scenarios. Currently, this phenomenon is identified by case review and is difficult to quantify. Regression techniques may be used to untangle the relative contributions of individual drugs to the high relative RR.[46]

The following example illustrates how the 'innocent bystander' phenomenon can arise. Suppose among a total of T reports in the database, M_B mention drug B and, of these, M_{AB} mention drug A and drug B. Let π_A denote the true likelihood that an AE is mentioned among the reports mentioning drug A (equation 9),

$$\pi_A = \text{Prob}(\text{AE} \mid \text{A}) \quad \text{(Eq. 9)}$$

and let (equation 10)

$$\pi_B = \text{Prob}(\text{AE} \mid \text{B}) \quad \text{(Eq. 10)}$$

denote the true likelihood that the AE is mentioned among the reports mentioning drug B. Suppose that $\pi_A > \pi_B$, and that the true probability that a report mentions the AE if it mentions drug A is π_A regardless of whether drug B is mentioned or not, that is (equation 11),

$$\pi_{AB} = \text{Prob}(\text{AE} \mid \text{A and B}) = \text{Prob}(\text{AE} \mid \text{A}) = \pi_A \quad \text{(Eq. 11)}$$

Let p_B denote the fraction of the reports mentioning drug B that are observed also to mention the AE. The quantity p_B is what one observes if only information about the mention of drug B and the AE in the reports is used. Since some of the reports that mention drug B also mention drug A, the observed reporting proportion for drug B, p_B, will not exceed the true reporting proportion, π_B because (equation 12):

$$p_B = \text{Prob}(\text{AE} \mid \text{A and B})\ \text{Prob}(\text{A and B} \mid \text{B}) + \text{Prob}(\text{AE} \mid \text{B and not A})\ \text{Prob}(\text{B and not A} \mid \text{B})$$
$$= \pi_B + (M_{AB}/M_B)(\pi_A - \pi_B) > \pi_B \text{ when } \pi_A > \pi_B \quad \text{(Eq. 12)}$$

If the AE is mentioned in N of the T reports, then the RR for the combination of the AE and drug B will be the reporting proportion divided by the overall reporting proportion, N/T. The true RR for drug A is (equation 13)

$$RR_A^{True} = \pi_A \div \frac{N}{T} \quad \text{(Eq. 13)}$$

and the true RR for drug B is (equation 14)

$$RR_B^{True} = \pi_B \div \frac{N}{T} \quad \text{(Eq. 14)}$$

If $\pi_A > \pi_B$, then the *observed* RR for drug B is (equation 15)

$$RR_B^{Obs} = RR_B^{True} + \frac{M_{AB}}{M_B}\left(RR_A^{True} - RR_B^{True}\right) \quad \text{(Eq. 15)}$$

If s $\pi_A \leq \pi_B$, then because we assume (equation 16)

$$\pi_{AB} = \pi_A,\ RR_B^{Obs} = RR_B^{True} \quad \text{(Eq. 16)}$$

In other words, if the true RR for drug A exceeds that for drug B, then the observed RR for drug B will be greater than the true ratio and the degree to which it is increased will depend on how many of the reports mention drug A as well as drug B. Otherwise, the RR for drug B will not be affected. In particular, this means that the RR for drug A will not be affected by the presence of drug B even though the converse is not true, at least under the assumptions used for the argument. The fact that the RR for drug B might be inflated by the effect of drug A does not mean that the true RR for drug B necessarily reflects no association; i.e. drug B might not be an 'innocent bystander'.

7. Issues in Interpreting Data-Mining Outputs

7.1 Overview

On the surface, interpreting the results of a disproportionality analysis is straightforward. Regardless of the method used, the numeric result, coupled with a defined 'threshold', indicates whether or not a drug-event pair of interest has been reported more frequently than 'expected' considering the background (usually the entire database) to which it was compared. The magnitude of the numeric result describes the degree of disproportionality, often referred to as the magnitude or 'strength' of the

'score'. Most methods also return some measure of confidence in the result (e.g. confidence limits or p-values).

Beyond this straightforward interpretation of reporting frequency, much caution is needed. One must understand the strengths and limitations of the method, the configuration details and the database in order to begin to understand the results. Even more importantly, one must understand the product, the event, the treatment of disease, complications of the underlying disease(s), the known pharmacotoxicology of the product and the external reporting environment in order to place the result in context. Apparent associations can occur for many reasons other than causal relationships between drugs and events. As stated previously, associations identified via data mining must be viewed as *hypotheses* regarding *possible* causal relationships between the drugs and events of interest. Causality can be established, if at all, only by careful medical follow-up of the clues about possible associations that are provided by quantitative signal detection methods. In some instances, epidemiological investigation of the issue may be valuable. It is also critically important to remember that the absence of a reporting relationship in a spontaneous reporting database does not rule out the existence of a safety problem and cannot be used to refute a signal detected by other means. Data-mining algorithms assist but do not replace the acuity of knowledgeable medical reviewers.[1,2]

The remainder of this section will discuss some of the issues related to databases, products, and the external environment that one must consider when interpreting the results of quantitative signaling methods. Most, if not all, of these issues are relevant to any method used to evaluate observational data.

7.2 Adverse Event Coding

As with any analysis of AE data, knowledge of the coding dictionary and the conventions used to code event terms is key. Most pharmaceutical manufacturers use MedDRA, which was introduced in the FDA's safety database AERS in November 1997 (replacing COSTART), although some of the new MedDRA terms were introduced in the 1995 version of COSTART. At its introduction, MedDRA had ten times more preferred term (PT) codes than COSTART and was designed with a hierarchical structure to allow the inclusion of more specific terms. Consequently, direct comparison of COSTART codes and MedDRA codes can be problematic.

Considerations related to event coding and quantitative signal detection include the following. (The WHO database uses a different dictionary, WHO-ART, to which many of the same considerations apply.)

- Although potentially important safety events cannot always be anticipated, prospectively grouping AE terms and developing case definitions whenever possible could be beneficial. Prospective grouping might be particularly important for syndromes involving multiple body systems such as serotonin syndrome and drug withdrawal.
- Generating results at the high-level term (HLT) level is generally not helpful as MedDRA HLTs often contain non-homogeneous medical concepts. Non-homogeneous groupings can contain disparate medical concepts, such as both high and low blood pressure PTs under the same HLT or can be groupings of very important and specific terms with less important and less specific terms under the same HLT. Examples of potentially misleading results at the HLT level include:[47]

(i) a high relative RR for the HLT 'ventricular arrhythmia', which is an event that is often of high clinical significance, may raise concern when the majority of reports are for the PT 'extrasystoles', which is an event that is often of minor clinical significance;

(ii) a low relative RR for the HLT 'febrile disorders' may not raise concern, but hidden under this grouping may be a high score for the PT 'neuroleptic malignant syndrome', which is a clinically significant event.

- A single medical concept may be represented by more than one PT and related medical concepts may be distributed in different system organ classes (SOCs). As an example, consider the PT 'hyperkalaemia' under the SOC 'metabolism and nutrition disorders' and the PT 'blood potassium increased' under the SOC 'investigations'. Advantages of this granularity are improved likelihood of capturing the actual event and reduced likelihood of misclassification resulting from the lack of a coding match. It is important to let

signal detection methods identify all disproportionately reported drug-event pairs and then to further investigate all related terms.
- Signals may be generated for an event for which a new MedDRA term was recently introduced. These situations can result in very high relative RRs because the expected value is very low because of the recent inclusion of the term in the database.
- Although both AERS and internal company databases use MedDRA, each organisation has its own coding rules that allow for consistency in data retrieval and data analysis. Different coding rules can profoundly affect signal detection characteristics (also see section 6.4).

7.3 Product Age (Time on Market)

When a drug first receives market authorisation, there is generally a steep increase in spontaneous reporting of AEs that plateaus after a number of years and eventually declines. The chance of a given event ever being reported increases as more data are accrued. There is evidence that as a drug matures, higher proportions of the reported AEs include known reactions and disease-related events.[48] Based on this evidence it is intuitively plausible, but not proven, that the number of new signals detected is likely to reach a peak over time, with a subsequent decline. However, a new dosage regimen or indication for a mature product, or its introduction into a new market, may result in a new pattern of reporting. Therefore, one should be aware of the lifecycle status of the drug in question, as well as the years of introduction to new markets and significant changes in its use.

It is also important to note that the number of spontaneous reports to AERS has increased dramatically since the start of the MedWatch programme in 1993. Although SRS/AERS began in 1968, Working Group members have noted that one-half of the reports in AERS were reported after 1997, with approximately 90% reported since 1990.

7.4 Targeted Surveillance and Stimulated Reporting

AE reporting for a given product may be influenced by many factors, including the initiation or intensity of targeted surveillance activities, selective prescribing (channeling) by physicians and publicity resulting from regulatory activities, litigation or highly publicised studies. Awareness of targeted surveillance and stimulated reporting situations is important when using spontaneous reporting databases for signal detection, regardless of the methods used.

7.4.1 Targeted Surveillance

Targeted surveillance activities include postmarketing epidemiological studies, product registries and surveillance requirements imposed by risk-management programmes. An example of targeted surveillance is the encouraged reporting of inadvertent exposure during pregnancy to two pregnancy category X drugs over several years following their approval[49] (category X is a designation in US product labelling that denotes the potential for fetal harm and contraindicates use during pregnancy). In situations such as this, relative RRs generated via data mining for adverse pregnancy outcomes or complications of pregnancy would need to be viewed in light of the targeted surveillance.

7.4.2 Selective Prescribing (Channeling)

Physicians' prescribing decisions are influenced by a number of factors. A patient's disease characteristics, including severity and prognosis, can influence prescribing, creating the potential for confounded drug-effect associations.[50-55] Prescribing also may be influenced by third-party payer formulary restrictions and by a patient's level of insurance coverage, again creating the potential for confounded drug-event associations.[56,57]

7.4.3 Stimulated Reporting

Publicity resulting from advertising, litigation or regulatory actions (e.g. 'Dear healthcare provider' letters and product withdrawals) may result in increased reporting and can generate higher-than-expected relative RRs.[56,58] Relative RRs should be examined over time in hopes of detecting these influences, although there are no definitive criteria for using data-mining techniques to reliably identify such effects.

8. Vaccines

8.1 Importance of Vaccine Safety

The ethical principle 'first do no harm' (*primum non nocere*) is the basis for the imperative for continuous evaluation of the safety of pharmaceutical products. Several features of vaccination add to this universal principle. Vaccinees are generally healthy and a large number of people are vaccinated, compared with drugs that are generally given to targeted groups of ill individuals. Paediatric vaccinations are often universally recommended or mandated by law, and children are a vulnerable population that needs special protection. Delivering the benefits of vaccination depends on maintaining the public's confidence in vaccine safety with both monitoring for previously unknown adverse effects or increases in known effects and careful analysis of hypothesised vaccine adverse effects.

There are established methods for ensuring the safety of vaccines postlicensure[59] that are beyond the scope of this section to review. Data-mining methods are a relatively new addition to these approaches, so there is a need to carefully consider how vaccines might require special consideration when applying these methods. Although the operational aspects of applying data-mining methods to vaccine AE databases and drug AE databases are identical, interpretation of the outputs of these methods might vary because of intrinsic differences between drugs and vaccines, as well as differences between the vaccine and drug AE databases.

8.2 Product Differences

Data mining in vaccine safety databases is likely to have different characteristics than data mining in drug safety databases because of intrinsic differences between drugs and vaccines. Because of their large preregistration trials, vaccines may be less likely than drugs to have common novel adverse effects emerge in the marketplace. However, the broad populations, with widely varying health profiles and co-morbidities, to which vaccines are administered postmarketing may increase the potential risks of developing rare AEs when compared with drugs, which are often administered as therapy for a single (or relatively narrow) set of conditions. The prophylactic use of most vaccines, versus the therapeutic use of most drugs, imparts a different benefit-risk profile. The pathophysiology of vaccine adverse effects is not as well defined as most drug adverse effects (e.g. hepatotoxicity after paracetamol [acetaminophen]), making it more difficult to use basic toxicological information and clinical judgement to interpret data mining results.

8.3 Concomitantly Administered Vaccines

There are fewer marketed vaccines than drugs, but many vaccines are given in specific combinations, especially during childhood. While technically it is possible to evaluate specific combinations of vaccines and AEs using data-mining methods, the fact that some vaccines are rarely administered or reported to VAERS alone may make it more difficult to distinguish AE associations for individual vaccines than for individual drugs. For example, intussusception is accepted as being caused by rotavirus vaccine, but rotavirus vaccine was usually administered simultaneously with diphtheria, tetanus and acellular pertussis (DTaP) vaccine, leading to DTaP being falsely signaled as being associated with intussusception (see section 6.7 regarding 'signal absorption'). Additional information from traditional methods of safety surveillance is needed to resolve such issues.

In data-mining analyses of vaccine safety data, we are attempting to identify associations between vaccines and AE coding terms. We know that vaccines are administered according to patient age (e.g. children receive 7-valent pneumococcal vaccines, whereas adults generally do not) and that the spectrum of AEs that occur in children is different than in adults (e.g. sudden infant death syndrome [SIDS] is limited by definition to infants, adults develop lung cancer). These patterns will influence the vaccine-event pairs that are reported to VAERS. Similarly, vaccines may be administered disproportionately by sex (e.g. more women may receive hepatitis B vaccine because of their status as healthcare workers) and disease patterns may differ in men and women (e.g. women experience autoimmune conditions more often than men). It is important that estimates of disproportionality be calculated based on a comparison in groups that have a similar likelihood of receiving similar vaccines and experiencing

similar AEs. This approach helps to prevent vaccine-AE pairs from being signaled because of differences in the underlying populations, rather than true differences in reporting of the AE from similar populations (e.g. DTaP-SIDS PRR misleadingly elevated because the comparison products are given to adults who do not routinely receive DTaP or die from SIDS). In the analysis, one can control for such confounding either by partitioning the data into like groups (e.g. only adults) or by stratification.

8.4 Validation of Data-Mining Methods

Although there are no 'gold standards' for the detection of vaccine-AE associations that can be used to precisely calculate sensitivity and specificity of a particular method, several surrogate measures of adverse effects have been proposed. In addition to product labelling, the Institute of Medicine (IOM) has conducted systematic reviews of vaccine AEs since the late 1980s and provided a list of AEs that they determined to be caused or not caused by vaccination.[60] In the absence of a gold standard, the IOM reviews might provide useful surrogate vaccine-AE pairs on which to retrospectively gauge the performance of various data-mining methods. However, the large number of vaccine-AE pairs for which a determination of causality has not been made and the continual improvement of knowledge about vaccine adverse effects limit our ability to precisely define sensitivity and specificity of these methods. New vaccines are continually introduced, older vaccines are used in new ways (e.g. smallpox vaccine to counter bioterrorism) and reporting patterns change over time; therefore, validation of the usefulness of these methods will ultimately depend on prospective application and successful early detection of an important new signal or signals.

9. Integrating Quantitative Signal Detection and Traditional Pharmacovigilance Methods

9.1 Overview

The process of signal detection in the postmarketing environment is both qualitative (e.g. clinical and scientific judgement) and quantitative. Traditional methods of signal detection and evaluation involve literature searching and case-by-case analysis, as well as crude frequency counts and calculation of reporting rates. The newer quantitative methods involving data-mining techniques are reviewed in section 4. Institutions considering the use of these newer methods should consider how to integrate them with traditional pharmacovigilance methods.[61]

Traditional methods alone are generally satisfactory when the volume of data is manageable. When the number of reports exceeds traditional signal-evaluating resources, combining traditional and data-mining methods may be considered. The choice of whether or not to employ data-mining methods should be evaluated by each institution since the added value of these methods is likely to be highly situation dependent. Among the many factors to consider are: (i) the rigor of existing signal detection practices/protocols based on clinical and scientific judgement; (ii) timelines; (iii) internal domain expertise of drugs and databases; (iv) availability and validation status of newer signal-detection methods; (v) availability and quality of comparative databases; and (vi) the uncertainties that remain about the predictive performance of these methods and databases through time.

As shown in figure 1, the process of signal detection can be initiated by selecting one or more traditional and/or data-mining methods. For example, one can begin the process by using the PRR method and complementing it with case-by-case analysis for signal detection. If the choice is made to use one or more data-mining methods, they should be used as a supplement to traditional methods. It should be noted that traditional methods can reveal safety signals that are otherwise not detected by data-mining methods and thus data mining should not be relied upon as a substitute for traditional methods, particularly with rare events or designated medical events.

If a signal is detected, it is important to evaluate the signal by conducting cumulative case review, literature review, assessment of preclinical and pharmacological data and, if appropriate, pharmacoepidemiological and clinical studies to assess causality. If a signal is not detected or is detected but not verified, then one needs to monitor and periodically repeat the process of safety signal detection or refute and 'close out' the signal if appropriate. Regardless

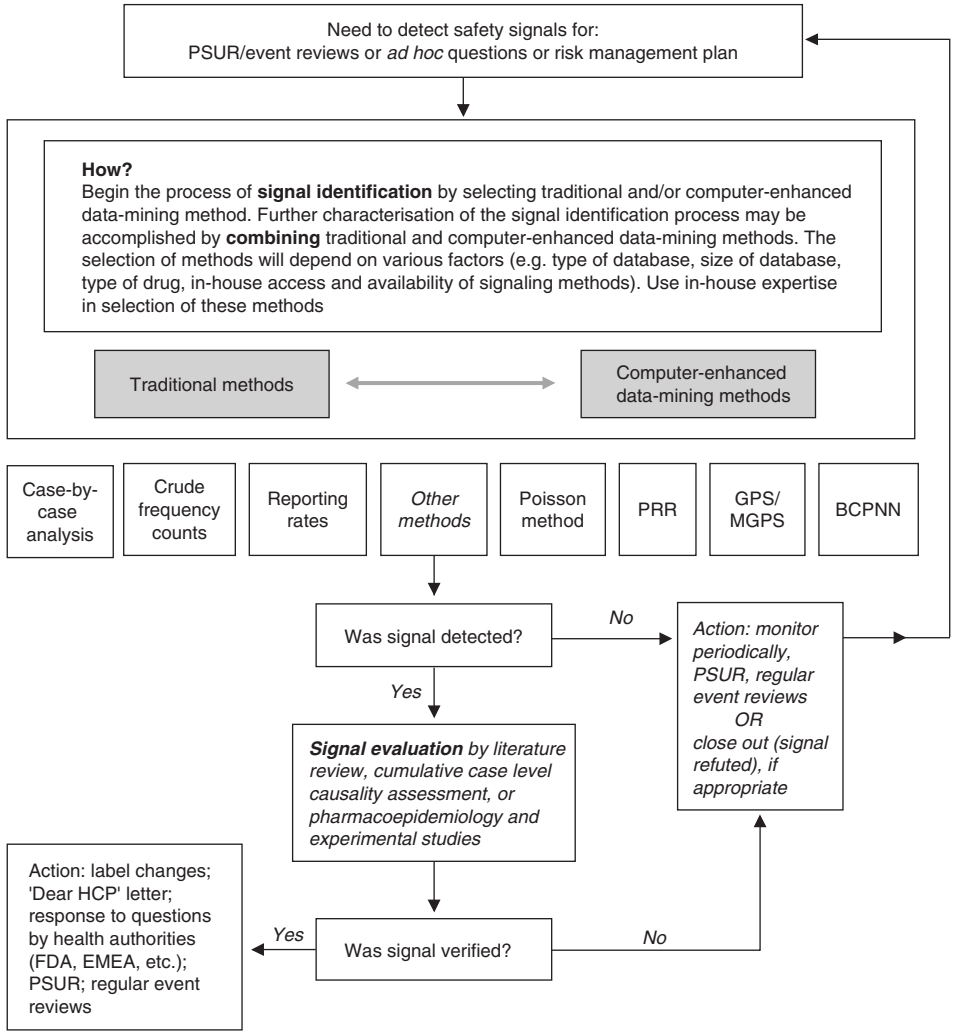


Fig. 1. Integrating computer-enhanced data-mining methods and traditional pharmacovigilance methods in process for signal detection. **BCPNN** = Bayesian confidence propagation neural network; **EMEA** = European Medicines Agency; **GPS** = gamma Poisson shrinker; **HCP** = healthcare provider; **MGPS** = multi-item gamma Poisson shrinker; **PRR** = proportional reporting ratio; **PSUR** = periodic safety update report.

of the method(s) used, the frequency of conducting proactive signal detection is highly dependent on the product and the potential safety issues involved.

In summary, disproportionality methods are not intended to be used in isolation. When these methods are appropriately incorporated into a comprehensive pharmacovigilance programme, clinical judgement and domain expertise should significantly mitigate the impact of false-positive and false-negative signals.

9.2 US FDA Perspective

Traditionally, a signal is generated by a question from the reviewing division, by a publication or by a safety reviewer's judgement based on the number and/or seriousness of reports of an AE for a particular drug. To confirm the observation, safety evaluators use a variety of approaches. Initially, the reviewers retrieve 'raw' numbers of reported cases to provide some perspective on the number of times an

event has been reported for a specific drug. A review of the medical literature may also be done. A 'hands on' review of each report is necessary to eliminate duplicate reports. A crude reporting rate can then be calculated by counting the number of reports of the AE in individual patients exposed to the drug and then dividing by the estimated number of prescriptions for the drug. The reporting rate (which should not be confused with an incidence rate) may be compared with an expected rate in the general population, but often such expected rates are difficult to ascertain.[3]

The empirical Bayesian data mining algorithm was initially implemented in February 1998.[1] In 2002, the FDA entered into a formal Cooperative Research and Development Agreement (CRADA) with a private advanced computer technology firm to collaborate in the development of a data-mining software application (MGPS) for use by safety evaluators, epidemiologists and medical officers at the FDA. When piloting of this system began in March 2003, various issues, including the following, were raised.

- Validity of approach: it was initially thought by some evaluators accustomed to a case-by-case review approach that applying empirical Bayesian methods to a database containing spontaneously submitted reports would not provide an accurate representation of a drug's potential association for AEs. Although some scepticism has diminished among these evaluators, the absolute interpretation of these results continues to pose challenges. These challenges in interpretation served as an important impetus in the formation of this Working Group.
- Lack of guidelines for interpretation: some safety evaluators and epidemiologists stated that they had difficulty in interpreting data-mining outputs because there were no standard guidelines for interpretation. For example, the definition of a signal may be dependent on several factors, including the AE(s) in question, the indication of a drug and the data being analysed (e.g. fatal outcomes). The signal threshold for a drug indicated for a serious disease with few, if any, treatment options may be higher than the threshold for a drug indicated for non-serious conditions with many treatment options. Thus, thresholds for action may be variable.[2,62]
- Added value of data mining: when considering the addition of a data-mining component to an already existing postmarketing surveillance group, questions naturally arose about whether the benefits associated with data mining outweigh its costs (e.g. economic, impact on public health). Indeed, the benefits of data mining can be difficult to quantify in any objective way. For example, the use of data mining is presumed to make postmarketing safety surveillance more efficient. As previously discussed, it is difficult to establish positive or negative predictive values with data mining. It is also difficult to define prospectively what a success with data mining would be and the quantity and quality of evidence needed to formulate a decision on whether data mining should be incorporated into an organisation's pharmacovigilance practices. Nonetheless, these methods do have some advantages over conventional clinical and epidemiological techniques. Because they are computer based, many analyses that would be difficult or impossible to do by standard methods can be carried out conveniently. This includes subsetting of the data, stratification, examination of the evolution of a signal over time and efficiently drilling down to individual reports.
- Added use of data mining: as part of its regulatory responsibility to monitor the potential toxicity of all marketed products, the FDA periodically examines drugs with similar chemical structures but different indications (e.g. α_1-blockers), drugs with different structures but the same indications (e.g. analgesics) and drugs with (or without) a specific, well established toxicity discovered by traditional methods (e.g. hepatotoxicity) to assure that no drug presents a previously unsuspected risk of important AEs noticeably higher than that presented by other drugs. The FDA is currently evaluating a screening approach that compares the confidence interval for the empirical Bayesian estimate of the RR for a suspect compound with the confidence intervals for multiple other 'control' drugs over successive intervals of time. Although the comparisons are formally equivalent to hypothesis tests, in fact the

results are not (and cannot be) interpreted as comparisons between treatments for reasons that are articulated in section 7.1. Instead, they are intended to provide a way to identify compounds for which further follow-up is needed to elucidate the reasons for the apparently elevated RRs. Experience so far is limited, but this approach may have merit as a regulatory screening tool, recognising the inherent problems in comparing RRs.

In 1999, researchers at the FDA/Center for Biologics Evaluation and Research retrospectively applied the empirical Bayesian method to determine when intussusception (a documented adverse reaction to rotavirus vaccine) showed association with rotavirus vaccine in the VAERS database.[63] They found that the empirical Bayesian method was able to detect the signal when only four cases of intussusception had been reported, which suggested the potential usefulness of this method to enhance vaccine safety. Evaluation of the empirical Bayesian and PRR methods in VAERS showed that both methods could contribute to vaccine safety data mining.[64] Application of the PRR method for surveillance of AEs after typhoid vaccines contributed to the detection of atypical allergic reactions after typhim Vi vaccine[65] and photophobia after smallpox vaccine.[66] The CDC also applied PRR methods in their evaluation of Bell's palsy after influenza vaccine,[67] although this was more controversial. In this study, the signal for Bell's palsy was generated independently of the VAERS data and the investigators used the increased PRR for Bell's palsy after influenza vaccines in VAERS, among other lines of evidence, to support the need for further evaluation. In an accompanying editorial, Shapiro[68] criticised this use of a data-mining method because, in his opinion, traditional clinical and epidemiological evaluation of the underlying case reports revealed sufficient limitations to undermine the conclusion.[69] The integration of traditional safety surveillance and data-mining methods for vaccine safety is an area that requires refinement, and delineating key concepts in applying data-mining methods to vaccine AE databases is an important step in this process.

9.2.1 Overall Lessons Learned

Data mining of surveillance systems may assist in identifying possible signals, but additional review and scientific investigations are always required to validate the signal and establish or rule out a causal relationship between a product and an AE. The absence of elevated relative RRs does not rule out a safety problem. Electronic pharmacovigilance systems assist but do not replace the acuity of knowledgeable safety evaluators and medical reviewers.

9.3 Industry Perspective from Pharmaceutical Research and Manufacturers of America Working Group Members

There are currently no regulatory or scientific requirements to use data mining for signal detection nor is there a single recommended approach to signal detection by regulatory authorities and pharmaceutical companies. However, the FDA recently issued guidances describing good practices for pharmacovigilance and pharmacoepidemiological assessment that discuss, among other things, potential roles for data mining in evaluating drug safety based on spontaneous postmarketing reports.[62] Before implementing any of these data-mining methods, a company should take a critical look at their current pharmacovigilance practices to determine what complementary methods might be needed.

If a decision is made to employ data-mining methods, it is very important to educate all members of a drug safety organisation, as well as others outside of drug safety, as to the strengths and limitations of the methods and of the spontaneous reporting databases themselves. People tend to want to draw very broad conclusions from outputs of data-mining analyses. It is important to emphasise that these methods are intended for screening databases, generating hypotheses and helping set priorities for review of reported AEs.

It is also advisable to develop transparent processes for data-mining activities that are consistent with company standard operating procedures (SOPs) and used consistently no matter what methods are implemented. At present, Working Group members are not aware of local or international regulations that cover data-mining processes. As data-mining methods evolve, SOPs may need to be updated. From a legal perspective, signal detection and follow-up of signals are likely to be discovera-

ble in litigation. It is therefore advisable to emphasise their preliminary and non-conclusive status, as well as to follow prudent document management guidelines once final decisions are made regarding signal validity.

Pharmacovigilance practitioners should periodically evaluate the effectiveness of their current procedures and carefully consider whether any additional methods, such as those discussed in this paper, could enhance their pharmacovigilance practices. Potential users should be encouraged to perform their own evaluations, not only to identify potential areas for improvement but also to contribute to further understanding of these methods, thereby promoting optimum use and minimising misuse or misapplication.

10. Other Uses of Data Mining

10.1 Comparing 'Signal Scores' Across Products

It is tempting to compare signal scores at some level and it is easy to construct various statistics for this purpose. However, differences between RRs do not imply differences in risk because spontaneous reporting databases are biased in ways that cannot be measured or controlled. It is not legitimate to infer that differences between scores imply differences between treatments without carefully considering the mechanisms that generate reports, including the known and unknown biases.

10.2 Evaluation of Potential Drug Interactions and Medical Syndromes

The principles of disproportionality may be applied to the detection of drug-drug interactions. Two approaches to analysing the effect of specific drug combinations on a predefined adverse effect of interest have been described. Both methods were tested using well known examples of drug-drug interactions. One approach involves the use of logistic regression modelling to evaluate statistical interactions amongst various therapies.[70] A second approach involves executing disproportionality analysis on therapy combinations of interest, where each combination is analysed as a unique drug variable.[71] In the latter approach, two potentially interacting drugs are analysed by creating three drug category variables: one category corresponds to all reports that mention both of the potentially interacting drugs and the other two categories correspond, respectively, to reports that mention either the first drug or the second, but not both. The use of multidimensional data-mining strategies that simultaneously screen for frequent 'drug-event-event' combinations may provide a fruitful approach to identifying syndromes with multiple AEs that are associated with the use of a drug.[72]

10.3 Evaluation of Demographic and Treatment-Related Factors

Some members of the Working Group have found it useful to conduct disproportionality analyses on subsets of a database, where the subset is defined by variables such as age, sex, report source or year of report. Such partitioning of the database may facilitate analyses involving specific populations of interest (e.g. females, paediatric patients).[1,73,74] For example, if an analysis is to be done on females, a subset of the database is created that only contains cases describing female patients and the data-mining algorithm is run on this data subset. Database subsets can also be used to examine the evolution of relative RRs over time. Subsets of the database are created based on report date and reporting statistics (e.g. RR or Empirical Bayes Geometric Mean) are computed for discrete or cumulative periods of time, typically year or quarter, depending on the age of the drug.[2,18] It may also be possible to compute signal scores for a particular drug according to different doses or dose ranges (if the data are available) by configuring a single drug name variable as multiple drug name variables, each of which represents a unique dose or dose range. Repeated analyses with multiple subsets generally should be avoided.

10.4 Restriction or Customisation of Database Backgrounds

It is possible to perform disproportionality analyses using customised or restricted 'backgrounds' rather than the entire database. One use of this technique is possibly to enhance the detection of signals specific to a particular drug within a drug

class. For example, if a drug class is generally associated with a particular event, a disproportionality algorithm could be run for all drugs in the class using only those drugs as the background. Another use is to assess the relative reporting of a drug-event pair with respect to a population of interest.

Changing the background alters the expected counts and therefore the relative RRs for events of interest. These approaches have not been studied in a systematic way; thus, there is no information on minimum background size or predictive utility. The Working Group agrees with the recommendation of Gogolak[75] that such analyses should be done in parallel with analyses that use the entire background dataset.

11. Summary and Recommendations

It was the goal of the Working Group to provide insights into both the potential utilities and limitations of data-mining methodologies and their role in pharmacovigilance. The following is a summary of key points and recommendations.

- Quantitative signal detection is a potentially useful adjunct to traditional pharmacovigilance practices.
- Data-mining methods analyse *relative reporting* of AEs. Hence, they cannot replace careful medical and scientific evaluation and should be considered as potential supplements to, not substitutes for, traditional signal-detection strategies.
- Data-mining results are hypothesis generating. They should be evaluated only in the context of other relevant data.
- The Working Group is not aware of any regulatory requirements for the use of data-mining methods in pharmacovigilance. Although there is evidence that data mining may be useful, the evidence is not sufficient to fully judge the value of data mining in pharmacovigilance. Time and experience will reveal its value and utility. Individual institutions may wish to evaluate the available methods and determine if data mining could provide value to their pharmacovigilance efforts.
- Statistical refinements that improve the capabilities of data-mining methods (e.g. adjusting for polytherapy and signal-masking effects) should further enhance their usefulness.
- The importance of vaccine safety as well as differences between vaccine and drug safety surveillance warrant special attention to data mining in vaccine AE databases. Continued efforts should be made to determine the best data-mining practices to enhance vaccine safety surveillance.
- A universal 'canonical' database, containing up-to-date information, for use in monitoring drug safety is highly desirable.

Acknowledgements

The Working Group acknowledges the participation and help of Rosanne Ososki, Lesley-Anne Furlong, Cheryl Watton, Dionigi Maladorno and Min Chu Chen. The Working Group also thanks Miles Braun and Paul Seligman for their support of this collaboration and for their helpful reviews of the manuscript.

We acknowledge PhRMA for funding technical support for preparation of this manuscript. A number of the authors are employed by pharmaceutical companies, as described in their respective affiliations.

References

1. O'Neill RT, Szarfman A. Some FDA perspectives on data mining for pediatric safety assessment. Curr Ther Res Clin Exp 2001; 62 (9): 650-63
2. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. Drug Saf 2002; 25 (6): 381-92
3. Szarfman A, Tonning JM, Doraiswamy PM. Pharmacovigilance in the 21st century: new systematic tools for an old problem. Pharmacotherapy 2004; 24 (9): 1099-104
4. Hauben M. Early postmarketing drug safety surveillance: data mining points to consider. Ann Pharmacother 2004; 38: 1625-30
5. Wolkenstein P, Latarget J, Roujeau J, et al. Randomised comparison of thalidomide versus placebo in toxic epidermal necrolysis. Lancet 1998; 352: 1586-9
6. Edwards IR, Biriell C. Harmonization in pharmacovigilance. Drug Saf 1994; 10 (2): 93-102
7. Lindquist M, Edwards IR, Bate A, et al. From association to alert: a revised approach to international signal analysis. Pharmacoepidemiol Drug Saf 1999; 8: S15-25
8. Report of CIOMS Working Group VI. Management of safety information from clinical trials. Geneva: Council for International Organization of Medical Sciences (CIOMS), 2005
9. Evans SJW, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf 2001; 10: 483-6
10. Gould AL. Practical pharmacovigilance analysis strategies. Pharmacoepidemiol Drug Saf 2003; 12: 559-74
11. Hauben M. A brief primer on automated signal detection. Ann Pharmacother 2003; 37: 1117-23
12. Hauben M, Zhou X. Quantitative methods in pharmacovigilance: focus on signal detection. Drug Saf 2003; 26 (3): 159-86

13. Wilson AM, Thabane L, Holbrook A. Application of data mining techniques in pharmacovigilance. Br J Clin Pharmacol 2003; 57 (2): 127-34
14. Evans SJ. Pharmacovigilance: a science or fielding emergencies? Stat Med 2000; 19 (23): 3199-209
15. Egberts AC, Meyboom RH, van Puijenbroek EP. Use of measures of disproportionality in pharmacovigilance: three Dutch examples. Drug Saf 2002; 25 (6): 453-8
16. van der Heijden PGM, van Puijenbroek EP, van Buuren S, et al. On the assessment of adverse drug reactions from spontaneous reporting systems: the influence of under-reporting on odds ratios. Stat Med 2002; 21: 2027-44
17. van Puijenbroek EP, Bate A, Leufkens HGM, et al. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiol Drug Saf 2002; 11: 3-10
18. Bate A, Lindquist M, Edwards IR, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol 1998; 54: 315-21
19. Lindquist M, Stahl M, Bate A, et al. A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. Drug Saf 2000; 23 (6): 533-42
20. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am Stat 1999; 53 (3): 177-202
21. DuMouchel W, Pregibon D. Empirical Bayes screening for multi-item associations. In: Conference on Knowledge Discovery in Data. Proceedings of the seventh ACM SigKDD International Conference on Knowledge Discovery and Data Mining. 2001 Aug 26–29; San Francisco (CA). New York: ACM Press, 2001: 67-76
22. Council for International Organizations of Medical Sciences (CIOMS). Guidelines for Preparing Core Clinical-Safety Information on Drugs. 2nd ed. Report of CIOMS Working Groups III and V. Geneva: Council for International Orgnizations of Medical Sciences (CIOMS), 1999: 27-33
23. Purcell P, Barty S. Statistical techniques for signal generation: the Australian experience. Drug Saf 2002; 25 (6): 415-21
24. Mamedov MA, Saunders GW. Fuzzy set analysis of Australian drug safety data. Proceedings of HIC 2002: Tenth National Health Informatics Conference; 2002 Dec 4–6, Melbourne
25. Mamedov MA, Saunders GW, Yearwood J. A fuzzy derivative approach to classification of outcomes from the ADRAC database. International Transactions in Operational Research 2004; 11 (2): 169-79
26. Spiegelhalter D, Grigg O, Kinsman R, et al. Risk adjusted sequential probability ratio tests: application to Bristol, Shipman and adult cardiac surgery. Int J Qual Health Care 2003; 15: 7-13
27. Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. Stat Methods Med Res 2003; 12 (2): 147-70
28. Kulldorf M, Fang Z, Walsh SJ. A tree-based scan statistic for database disease surveillance. Biometrics 2003; 59: 323-31
29. Hauben M, Reich L. Drug-induced pancreatitis: lessons in data mining [letter]. Br J Clin Pharmacol 2004; 58 (5): 560-2
30. Waller P, Heeley E, Moseley J. Impact analysis of signals detected from spontaneous adverse drug reaction reporting data. Drug Saf 2005; 28 (10): 843-50
31. Waller PC, Heeley EL, Moseley JNS. Impact analysis of signals detected from spontaneous adverse reaction reporting data [abstract]. Pharmacoepidemiol Drug Saf 2004; 13: S323
32. Heeley E, Waller P, Moseley J. Testing and implementing signal impact analysis in a regulatory setting: results of a pilot study. Drug Saf 2005; 28 (10): 901-6
33. Rothman KJ, Lanes S, Sacks ST. The reporting odds ratio and its advantages over the proportional reporting ratio. Pharmacoepidemiol Drug Saf 2004; 13: 519-23
34. Waller P, van Puijenbroek E, Egberts A, et al. The reporting odds ratio versus the proportional reporting ratio: 'deuce' [letter]. Pharmacoepidemiol Drug Saf 2004; 13: 525-6
35. Stahl M, Lindquist M, Edwards IR, et al. Introducing triage logic as a new strategy for the detection of signals in the WHO drug monitoring database. Pharmacoepidemiol Drug Saf 2004; 13: 355-63
36. Begaud B, Moride Y, Tubert-Bitter P, et al. False-positive in spontaneous reporting: should we worry about them? Br J Clin Pharmacol 1994; 38 (5): 401-4
37. Hauben M. Application of an empiric Bayesian data mining algorithm to reports of pancreatitis associated with atypical antipsychotics. Pharmacotherapy 2004; 24 (9): 1122-9
38. Hauben M. Trimethoprim-induced hyperkalaemia-lessons in data mining [letter]. Br J Clin Pharmacol 2004; 58 (3): 338-9
39. Hauben M, Reich L. Safety related drug-labelling changes: findings from two data mining algorithms. Drug Saf 2004; 27 (10): 735-44
40. Levine JG, Tonning JM, Szarfman A. Reply: the evaluation of data mining methods for the simultaneous and systematic detection of safety signals in large databases. Lessons to be learned [letter]. Br J Clin Pharmacol. In press
41. Hauben M, Reich L. Response to letter by Levine et al. [letter]. Br J Clin Pharmacol. In press
42. Lindquist M. The WHO adverse reaction database: basic facts [online]. Available from URL: http://www.who-umc.org/pdfs/WHO%20Adverse%20Reaction%20Database%20basic%20facts.pdf [Accessed 2004 Sep 14]
43. European Medicines Agency. EudraVigilance [online]. Available from URL: http://www.eudravigilance.org/highres.htm [Accessed 2005 Sep 17]
44. Cosentino M, Leoni O, Michielotto D, et al. Increased reporting of adverse reactions to ACE inhibitors associated with limitations to drug reimbursement for angiotensin-II antagonists. Eur J Clin Pharmacol 2001; 57: 509-12
45. Bate A, Edwards RI, Lindquist M, et al. The authors' reply [letter]. Drug Saf 2003; 26 (5): 364-6
46. Szarfman A, DuMouchel W, Fram D, et al. Lactic acidosis: unraveling the individual toxicities of drugs used in HIV and diabetes polytherapy by hierarchical Bayesian logistic regression data mining [abstract]. 11th Annual FDA Science Forum, 2005 Apr 27-28 [online]. Available from URL: http://www.cfsan.fda.gov/~frf/forum05/H-30.htm [Accessed 2005 Sep 14]
47. Brown EG. Effects of coding dictionary on signal generation: a consideration of use of MedDRA compared with WHO-ART. Drug Saf 2002; 25 (6): 445-52
48. Haramburu F, Begaud B, Moride Y. Temporal trends in spontaneous reporting of unlabelled adverse drug reactions. Br J Clin Pharmacol 1997; 44: 299-301
49. Manson JM, Freyssinges C, Ducrocq MB, et al. Postmarketing surveillance of lovastatin and simvastatin exposure during pregnancy. Reprod Toxicol 1996; 10 (6): 439-46
50. Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of beta 2-agonists. Am J Epidemiol 1996; 15 (12): 1161-9
51. Blais L, Ernst P, Suissa S. The authors' reply [letter]. Am J Epidemiol 1997; 146 (10): 886-7
52. Leufkens HG. Pharmacoepidemiology and gastroenterology: a close couple. Scand J Gastroenterol Suppl 2000; 232: 105-8
53. Leufkens H, Urquhart J. Variability in patterns of drug usage. J Pharm Pharmacol 1994; 46 Suppl. 1: 433-7
54. Meijer WEE, Heerdink ER, Pepplinkhuizen LP, et al. Prescribing patterns in patients using new antidepressants. Br J Clin Pharmacol 2001; 51: 181-3

55. Pearce N, Beasley R, Crane J, et al. Confounding by indication and channeling over time: the risks of beta-2 agonists [letter]. Am J Epidemiol 1997; 146 (10): 885-6

56. deBruin ML, van Puijenbroek EP, Egberts ACG, et al. Non-sedating antihistamine drugs and cardiac arrhythmias: biased risk estimates from spontaneous reporting systems? Br J Clin Pharmacol 2002; 53: 370-4

57. Tisonova J, Szalayova A, Kriska M. Factors influencing the spontaneous reporting of adverse drug reactions: the experience of the Slovak Republic. Pharmacoepidemiol Drug Saf 2003; 13: 333-7

58. Coster TS, Szarfman A, Tonning J. The application of data mining to analyze pre-publicity psychiatric signals with the use of mefloquine [abstract]. ASCPT Annual Meeting; 2004 Mar 4; Miami (FL)

59. Varricchio F, Iskander J, Destefano F, et al. Understanding vaccine safety information from the vaccine adverse event reporting system (VAERS). Pediatr Infect Dis J 2004; 23: 287-94

60. Institute of Medicine. Immunization safety review [online]. Available from URL: http://www.iom.edu/project.asp?.id=4-705 [Accessed 2005 Sep 26]

61. Yee CL, Klincewicz SL, Knight JF, et al. Practical considerations in developing an automated signaling program within a pharmacovigilance department. Drug Inf J 2004; 38: 293-300

62. US FDA. Guidance for industry: good pharmacovigilance practices and pharmacoepidemiologic assessment. US Food and Drug Administration Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research, March 2005 [online]. Available from URL: http://www.fda.gov/cder/guidance/6359OCC.htm#_Toc48124197 [Accessed 2005 Sep 27]

63. Niu MT, Erwin DE, Braun MM. Data mining in the US vaccine adverse event reporting system: early detection of intussusception and other events after rota virus vaccine. Vaccine 2001; 19: 4627-34

64. Banks D, Woo EJ, Burwen D, et al. Comparison of 4 data mining methods in the US Vaccine Adverse Event Reporting System (VAERS) [abstract]. Pharmacoepidemiol Drug Saf 2003; 12 Suppl. 1: S138

65. Begier EM, Burwen D, Haber P, et al. Post-marketing safety surveillance for typhoid fever vaccines from the Vaccine Adverse Event Reporting System, July 1990-June 2002. Clin Infect Dis 2004; 38: 771-9

66. McMahon AW, Bryant-Genevier MC, Woo EJ, et al. Photophobia following smallpox vaccination [letter]. Vaccine 2005; 23: 1097-8

67. Zhou W, Pool V, DeStefano F, et al. A potential signal of Bell's palsy after parenteral inactivated influenza vaccines: reports to the vaccine adverse event reporting system (VAERS): United States, 1991-2001. Pharmacoepidemiol Drug Saf 2004; 13: 505-10

68. Shapiro S. Clinical judgment, common sense and adverse reaction reporting. Pharmacoepidemiol Drug Saf 2004; 13: 511-3

69. Zhou W, Pool V, DeStefano F, et al. Reply to the editorial. Pharmacoepidemiol Drug Saf 2004; 13: 515-7

70. van Puijenbroek EP, Egberts ACG, Heerdink ER, et al. Detecting drug-drug interactions using a database for spontaneous adverse drug reactions: an example with diuretics and non-steroidal anti-inflammatory drugs. Eur J Clin Pharmacol 2000; 56: 733-8

71. Almenoff JS, DuMouchel W, Kindman A, et al. Disproportionality analysis using empirical Bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting. Pharmacoepidemiol Drug Saf 2003; 12 (6): 517-21

72. Szarfman A. Syndromic surveillance and risk management using multi-item gamma Poisson shrinker. Journal of Urban Health: bulletin of the New York Academy of Medicine 2003; 80 (2 Suppl. 1): i133 [online]. Available from URL: http://www.syndromic.org/syndromicconference/2002/Supplement-pdf/Abstracts_SectionIV.pdf [Accessed 2005 Sep 15]

73. Yuen NA, Almenoff JS, DuMouchel W, et al. Disproportionality analysis to explore patient and treatment related factors associated with adverse events [abstract]. Pharmacoepidemiol Drug Saf 2004; 13: S259

74. Szarfman A. Gender-related 'higher-than-expected' drug-event combinations in spontaneous adverse drug event reports [abstract no. D05]. 2000 FDA Science Forum – FDA and the science of safety: new perspectives; 2000 Feb 14-15; Washington, DC [online]. Available from URL: http://vm.cfsan.fda.gov/~frf/forum00/d05.htm [Accessed 2004 Oct 12]

75. Gogolak VV. The effect of backgrounds in safety analysis: the impact of comparison cases on what you see. Pharmacoepidemiol Drug Saf 2003; 12: 249-52

Correspondence and offprints: Dr *June Almenoff*, Global Clinical Safety and Pharmacovigilance, GlaxoSmithKline, 5 Moore Drive, Mail Stop 5.4214.4C, PO Box 13398, Research Triangle Park, NC 27709-3398, USA.
E-mail: june.s.almenoff@gsk.com

PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2014; **23**: 195–207
Published online 15 November 2013 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3529

ORIGINAL REPORT

Assessing the extent and impact of the masking effect of disproportionality analyses on two spontaneous reporting systems databases

Francois Maignen[1]*, Manfred Hauben[2,3,4,5], Eric Hung[2], Lionel Van Holle[6] and Jean-Michel Dogne[7]

[1] *European Medicines Agency, London, UK*
[2] *Pfizer Inc., New York, NY, USA*
[3] *School of Medicine, New York University, New York, NY, USA*
[4] *New York Medical College, New York, NY, USA*
[5] *Brunel University, West London, UK*
[6] *GlaxoSmithKline Biologicals SA, Wavre, Belgium*
[7] *Department of Pharmacy-NTHC-NARILIS, FUNDP, University of Namur, Namur, Belgium*

ABSTRACT

Background Masking is a statistical issue by which signals are hidden by the presence of other medicines in the database. In the absence algorithm, the impact of the masking effect has not been fully investigated.
Objective Our study is aimed at assessing the extent and the impact of the masking effect on two large spontaneous reporting databases.
Study design Cross sectional study using a set of terms of importance for public health in two spontaneous reporting databases.
Setting The analyses were performed on EudraVigilance (EV) and the Pfizer spontaneous reporting database (PfDB).
Main outcome measure Using the masking ratio, we have identified and removed the products inducing the highest masking effect.
Results Studying a total of almost 50 000 drug-event combinations masking had an impact on approximately 60% of drug-event combinations were masked by another product with a masking ratio >1 in EV and 84% in PfDB. The prevalence of important masking was quite rare (0.003% of the DECs) and mainly affected events rarely reported in EV. The products involved in the highest masking effects are products known to induce the reaction. The removal of the masking effect of the highest masking product has revealed 974 signals of disproportionate reporting in EV including true signals. The study shows that the original ranking provided by the quantitative methods included in our study is marginally affected by the removal of the masking product.
Conclusion Our study suggests that significant masking is rare in large spontaneous databases and mostly affects events rarely reported in EV.

KEY WORDS—disproportionality analysis; masking; EudraVigilance; signal detection; public health; proportional reporting ratio; pharmacoepidemiology

Received 15 October 2012; Revised 15 May 2013; Accepted 14 August 2013

INTRODUCTION

The masking effect is a collateral effect of the quantitative methods based on disporportionality analysis.[1] Masking is an effect by which a signal of disproportionate reporting (SDR)[2] for a given drug-event pair might be suppressed by the presence of another product in the same database. The impact of the masking effect on the detection of new signals of public health relevance is not fully understood. The effect has been studied using mathematical simulations or an empirical approach.[3,4] Two studies have suggested that the prevalence of masking might be low in two spontaneous reporting databases (a vaccine safety database and Adverse Event Reporting System [AERS]).[3,5] However, a recent study has confirmed previous findings made by Gould that the removal of the masking effect can unravel new signals of public health relevance (isotretinoin and gastrointestinal haemorrhages, methylprednisolone

*Correspondence to: F. Maignen, Pharmacovigilance and Risk Management sector, European Medicines Agency, London, UK. E-mail: francois.maignen@ema.europa.eu

Table 1. Contingency table for the computation of measures of disproportionality in the presence of a masking medicinal product (the contingency table includes two products, one product for which the disproportionality analysis is performed, product A, and a second, masking product, product B)

	Event of interest	Other events		Total
Product A	n_{11}	n_1	$n_{11}=n_{12}$	$n_{1.}$
Product B (masking)	n_{21}	$n_{2.}$	$n_{21}=n_{22}$	$n_{2.}$
All other products in the database (excluding both A and B)	n_{31}	n_3	$n_{31}=n_{32}$	$n_{3.}$
Total	$n_{.1}$	$n_{..}$	$n_{.1}=n_{.2}$	$n_{..}$

and cerebrovascular accidents and haemorrhages).[1,6] Finally, the removal of the masking effect may also be associated with a gain of time in detecting new signals as its removal results in a decrease of the number of reports needed to detect a signal.[5]

At the moment, the only method to detect and quantify a possible masking effect is to run a disproportionality analysis excluding a subset of reports suspected to induce a masking effect and compare the results with the original measure performed on the entire dataset.[1] The identification of candidate masking products still relies on empirical approaches. There is currently no algorithm that can detect and quantify the presence, direction and magnitude of a masking effect. We have developed and validated against the reference method a mathematical algorithm by using simulated and real spontaneous reporting data.[7] We have subsequently conducted a study aimed at assessing the extent and the impact of the masking effect on two large spontaneous reporting databases EudraVigilance (EV) and Pfizer spontaneous reporting database (PfDB). We have focused our study on the masking induced by a single product responsible for the highest effect for a given event.

MATERIALS AND METHODS

We have conducted our analyses in two large spontaneous reporting databases (both databases contained more than two million records at the time of the analysis), EV[8,9] and PfDB, separately and independently on the spontaneous reports (duplicate reports not systematically excluded) received until end of April 2011. We have used our mathematical algorithm to identify and remove the effect associated with the highest masking product for each event included in our study. The analyses were performed using the proportional reporting ratio (PRR)[10,11] (method currently used in EV)[8] and its corresponding masking ratios (MR) (Tables 1 and 2). All our computations were performed at the report level.[8,10]

The selection of MedDRA terms was based on a set of hypotheses underlying the mathematical development of our algorithm. Firstly, we wanted to assess the impact of the masking effect on events of public health importance (because signal detection activities primarily focuses on such events). For that purpose, we have used the 23 adverse events identified by the EU-ADR group considered to be important in pharmacovigilance.[12] Secondly, we have added a set of designated medical events associated with the use of biological, vaccines or new chemical entities, which are currently not present in the EU-ADR list (e.g. progressive multifocal leukoencephalopathy, anti-erythropoietin antibody positive, polyomavirus associated nephropathy). Finally, we have added a set of Medical Dictionary for Regulatory Activities (MedDRA) terms rarely reported in EV as the mathematical expression of the exact masking ratio suggested that events rarely reported could be preferentially affected by the masking (i.e. total number of reports in EV ranging from approximately less than 10 cases to 1500 reports in the entire database) (Table 3).

Table 2. mathematical expressions of the masking ratio computed at the report level and its proposed approximations (*valid for detecting the product inducing the highest masking effect*) for the three main measures of disproportionality analyses used on spontaneous reporting system databases. The corresponding definitions of the n_{ij} included in this Table are given in Table I, the masking ratio studies in influence of a masking product (B) on the product which the analysis is performed (% n_{21} denotes the number of reports containing the masking product but not the product for which the analysis is conducted). *The reports containing both A and B allocated to both products, the proportion of reports containing these two products is assumed to be low (less than 50%)

Measure of disproportionality	Exact masking ratio	Exact masking ratio (alternative expression)	Proposed approximations of the masking ratio*	Assumptions
Proportional reporting ratio	$MR=\frac{PRR_{A(without\ B)}}{PRR_{A(with\ B)}}=\frac{n_{3.}(\%n_{21}+n_{31})}{n_{31}(\%n_{2.}+n_{3.})}=\frac{n_{3.}}{\%n_{2.}+n_{3.}}\left(1+\frac{\%n_{21}}{n_{31}}\right)$	$MR=\frac{1+\frac{\%n_{21}}{n_{31}}}{1+\frac{\%n_{2.}}{n_{3.}}}$	$MR_{PRRApprox.1}=1+\frac{n_{21}}{n_{11}+n_{31}}$	Assumption 1: $1+\frac{n_{2.}}{n_{3.}}$ close to 1
			$MR_{PRRApprox.3}=\frac{1+\frac{n_{21}}{n_{11}+n_{31}}}{1+\frac{n_{2.}}{n_{3.}}}$	Assumption 2: $n_{11}<<n_{31}$ $n_{11}<<n_{31}$
			$\%B=\frac{\%n_{21}}{n_{.1}}$ or $B=\frac{n_{21}}{n_{.1}}$	$1+\frac{n_{2.}}{n_{3.}}$ close to 1 (i.e. $n_{2.}<<n_{3.}$) and $n_{31}>>n_{11}$

DOI: 10.1002/pds

Table 3. List of Medical Dictionary for Regulatory Activities preferred terms included in the study. This list includes terms from the list of EU-ADR list of important events. The other terms were chosen in EudraVigilance either for their public health importance or because of the low number of reports in the whole database. The figures included in the table gives the approximate number of reports involving this preferred term in the database. The EU-ADR events are commonly reported to EudraVigilance

Reaction PT	Event
Acute hepatic failure	EU-ADR
Acute myocardial infarction	EU-ADR
Amnesia	EU-ADR
Anaphylactic shock	EU-ADR
Anterograde amnesia	EU-ADR
Anti-erythropoietin antibody positive	Less 500
Aplastic anaemia	EU-ADR
Bone debridement	Less 500
Bone marrow reticulin fibrosis	Less 100
Bronchiolitis	Less 1500
Cardiac valve disease	EU-ADR
Confusional state	EU-ADR
Convulsion	EU-ADR
Craniopharyngioma	Less 100
Depression	EU-ADR
Dermatitis bullous	EU-ADR
Drug specific antibody present	Less 1500
Dupuytren's contracture	Less 100
Electrocardiogram QT prolonged	EU-ADR
Epiphysiolysis	Less 500
Extrapyramidal disorder	EU-ADR
Factor IX inhibition	Less 100
Factor VIII inhibition	Less 1500
Fanconi syndrome	Less 500
Fanconi syndrome acquired	Less 500
Gambling	Less 100
Haemolytic anaemia	EU-ADR
Intussusception	Less 1500
Jaw operation	Less 500
Mania	EU-ADR
Mitochondrial toxicity	Less 100
Nephrogenic systemic fibrosis	Less 1500
Neuropathy peripheral	EU-ADR
Neutropenia	EU-ADR
Ovarian hyperstimulation syndrome	Less 1500
Pancreatitis acute	EU-ADR
Pancytopenia	EU-ADR
Pathological gambling	Less 1500
Polyomavirus-associated nephropathy	Less 500
Pregnancy with contraceptive patch	Less 500
Progressive external ophthalmoplegia	Less 100
Progressive multifocal leukoencephalopathy	Less 1500
Rapid correction of hyponatraemia	Less 100
Rash maculo-papular	EU-ADR
Renal failure acute	EU-ADR
Retrograde amnesia	EU-ADR
Rhabdomyolysis	EU-ADR
Rosai-Dorfman syndrome	Less 100
Stevens-Johnson syndrome	EU-ADR
Sudden onset of sleep	Less 500
Suicidal behaviour	EU-ADR
Suicide attempt	EU-ADR
Thrombocytopenia	EU-ADR
Toxic epidermal necrolysis	EU-ADR
Upper gastrointestinal haemorrhage	EU-ADR
Venous thrombosis	EU-ADR

PT = preferred term.

RESULTS

Results obtained in EudraVigilance

The disproportionality analysis conducted on the MedDRA preferred terms (PT) included in our study, yielded a total of 30 645 drug-event combinations (DECs). Of these 29 245 DECs involved the events considered to be important in pharmacovigilance (EU-ADR events; these events are commonly reported in EV). A total of 1400 DECs involved our additional set of events that have been rarely reported in EV. Our study confirms results from previous studies that the highest masking effect is induced by products for which the reaction is known or has been extensively or over reported (Table 4). Masking was consistently associated with products for which a very high PRR is observed for the event. The inverse is not necessarily true. We have identified a potential masking effect (the approximate MR was greater than 1) for 18 599 masking DECs (MECs), that is, 60.85% of the DECs. The MR was greater than 1.1 for only 87 MECs (0.5% of MECs for which the MR is above 1), above 1.5 for only 28 MECs (0.15%) and above 2 for only 20 MECs (0.1%). All the DECs actually affected by a consequential masking effect involved events rarely reported in EV (Table 5).

The distribution of the masking ratio shows that events rarely reported in EV are mostly affected by a possible masking effect of the PRR (Figures 1 and 2). A significant carry-over masking effect can still be observed with products that have been withdrawn from the market (subject of a stimulated reporting).

The removal of the masking effect has revealed 974 new SDRs (SDRs, defined by a new PRR above or equal to 2). The number of SDRs before the removal of the highest masking product was 12 861 (i.e. 42% of the DECs included in our study), the number after removal increased to 13 835 (i.e. 45% of the DECs, increase by approximately 3%). The respective proportion of SDRs revealed for each of the PTs included in our study was heterogeneous (Table 6). We could not find any clear correlation between the number of unmasked SDRs and the MR of the highest masking product removed (Figure 3). Some of the new SDRs revealed by the removal of the masking correspond to true signals including signal of public health importance (e.g. progressive multifocal leukoencephalopathy and natalizumab) (Table 5). We show examples of new SDRs identified for the MedDRA PT 'anaphylactic shock' (Table 7).

We have studied the changes in the ranking of the PRR before and after the removal of the masking effect induced by the highest masking drug on all the events included in our study (using $MR_{PRRApprox3}$).

DOI: 10.1002/pds

Table 4. Highest masking product for each of the reaction term (Medical Dictionary for Regulatory Activities preferred terms) included in the study conducted in EudraVigilance. The table shows that the masking effect is mainly induced by known associations. ApproxMR denotes $MR_{PRRApprox1}$ (Table 2)

Reaction PT	Product inducing the highest masking effect	Approximate MR
Acute hepatic failure	Paracetamol	1.652825
Acute myocardial infarction	Rofecoxib	1.743396
Amnesia	Zolpidem	1.12965
Anaphylactic shock	Ceftriaxone	1.042165
Anterograde amnesia	Zolpidem	1.65486
Anti-erythropoietin antibody positive	Epoetin alfa	6.518472
Aplastic anaemia	Carbamazepine	1.046274
Bone debridement	Zoledronic acid	5.236867
Bone marrow reticulin fibrosis	Romiplostim	8.558506
Bronchiolitis	Palivizumab	1.765868
Cardiac valve disease	Phentermine	1.224371
Confusional state	Tramadol	1.020954
Convulsion	Bupropion	1.03647
Craniopharyngioma	Somatropin	6.264051
Depression	Isotretinoin	1.125742
Dermatitis bullous	Ketoprofen	1.047514
Drug specific antibody present	Heparin	1.687291
Dupuytren's contracture	Rofecoxib	1.1931
Electrocardiogram QT prolonged	Cisapride	1.176308
Epiphysiolysis	Somatropin	4.819412
Extrapyramidal disorder	Risperidone	1.256385
Factor IX inhibition	Nonacog alfa	2.624811
Factor VIII inhibition	Octocog alfa	2.062682
Fanconi syndrome	Tenofovir	1.630408
Fanconi syndrome acquired	Tenofovir	1.630408
Gambling	Pramipexole	2.692829
Haemolytic anaemia	Ribavirin	1.081483
Intussusception	Rotavirus vaccine, live, oral, pentavalent	2.522654
Jaw operation	Zoledronic acid	4.881738
Mania	Paroxetine	1.089134
Mitochondrial toxicity	Lamivudine	1.574607
Nephrogenic systemic fibrosis	Gadodiamide	2.521013
Neuropathy peripheral	Bortezomib	1.104694
Neutropenia	Docetaxel	1.067381
Ovarian hyperstimulation syndrome	Chorionic gonadotrophin	1.57331
Pancreatitis acute	Quetiapine	1.064298
Pancytopenia	Methotrexate	1.101807
Pathological gambling	Pramipexole	2.97808
Polyomavirus-associated nephropathy	Tacrolimus	3.326415
Pregnancy with contraceptive patch	Ethinylestradiol, Norelgestromin	4.036374
Progressive external ophthalmoplegia	Didanosine	11.99059
Progressive multifocal leukoencephalopathy	Rituximab	1.647151
Rapid correction of hyponatraemia	Tolvaptan	21.99909
Rash maculo-papular	Carbamazepine	1.049695
Renal failure acute	Furosemide	1.034391
Retrograde amnesia	Midazolam	1.067364
Rhabdomyolysis	Simvastatin	1.289431
Rosai-Dorfman syndrome	Canakinumab	1.166657
Stevens-Johnson syndrome	Carbamazepine	1.111347
Sudden onset of sleep	Pramipexole	1.908177
Suicidal behaviour	Varenicline	1.143097
Suicide attempt	Lorazepam	1.061408
Thrombocytopenia	Heparin	1.045518
Toxic epidermal necrolysis	Allopurinol	1.085932
Upper gastrointestinal haemorrhage	Ibuprofen	1.104194
Venous thrombosis	Bevacizumab	1.065716

PT = preferred term; MR = masking ratio.

The plot of the PRR before and after the removal of the product inducing the highest masking effect show that the removal did not (or marginally) affected the ranking provided by the PRR. The removal of the masking effect is an indirect way to lower the thresholds commonly used to define SDRs. This effect was observed whether the event is commonly or rarely represented in the database (Figures 4 and 5).

Results obtained in a company database compared to EudraVigilance

The events selected in the study included 18 982 DECs in PfDB. Of these, there was perfect agreement between approximate and the exact calculation in terms of the drug identified as having the highest MR in the company database. The results obtained on PfDB showed that the MR was distributed slightly differently compared to the distribution observed in EV. Specifically, there was a greater prevalence of masking on the same selected set of events (84.25% of DECs vs 60.85% in EV). Masking was not only more common but MRs in the company database were skewed towards higher values (Table 8). Figure 6 is a frequency histogram of the exact MRs for the event under study.

Table 9 Displays the differences in the nature of masking observed between PfDB and EV (events affected and magnitude of the masking). The differences in maximum MR between the company database and EV are very small for most of the terms included in the study. Discrepancies were observed for structural differences (absence of DECs reported to the database), events connected with products marketed by the company (higher masking in PfDB compared to EV; these involved the events cardiac valve disease, retrograde amnesia and epiphysiolysis) or for products subject of an important reporting in EV (Bone

Table 5. Top 50 drug-event combinations mostly affected by the masking effect of other products present in the database identified in our study (EudraVigilance [EV]). The table contains the preferred terms and the active substance affected by the masking effect. Most of these events were rarely reported in EV (84% of these drug-event pairs affected by the most important masking effect involve terms, for which less than 500 reports have been reported in EV). ApproxMR denotes $MRPRR_{Approx1}$ (Table 2)

Reaction PT	Masked active substance	Approximate MR
Progressive external ophthalmoplegia	Stavudine	11.98783925
Progressive external ophthalmoplegia	Lamivudine	11.96701266
Progressive external ophthalmoplegia	Abacavir	5.99432888
Progressive external ophthalmoplegia	Ritonavir	5.988677247
Progressive external ophthalmoplegia	Indinavir	3.995437552
Progressive external ophthalmoplegia	Zidovudine	3.992756816
Progressive external ophthalmoplegia	Etravirine	2.999223713
Progressive external ophthalmoplegia	Raltegravir	2.998501018
Progressive external ophthalmoplegia	Darunavir	2.998361351
Progressive external ophthalmoplegia	Saquinavir	2.997781572
Progressive external ophthalmoplegia	Nevirapine	2.996704839
Progressive external ophthalmoplegia	Tenofovir	2.995400741
Polyomavirus-associated nephropathy	Mycophenolate mofetil	2.678459766
Progressive external ophthalmoplegia	Lopinavir	1.999850589
Progressive external ophthalmoplegia	Nelfinavir	1.999064557
Progressive external ophthalmoplegia	Efavirenz	1.997851947
Nephrogenic systemic fibrosis	Gadopentetate dimeglumine	1.903688291
Jaw operation	Pamidronate disodium	1.890500639
Bone debridement	Pamidronate disodium	1.74253722
Mitochondrial toxicity	Stavudine	1.536902468
Progressive external ophthalmoplegia	Delavirdine	1.499938287
Mitochondrial toxicity	Tenofovir	1.49770037
Anti-erythropoietin antibody positive	Epoetin beta	1.429034084
Ovarian hyperstimulation syndrome	Menotrophin	1.401388192
Nephrogenic systemic fibrosis	Gadoversetamide	1.40008464
Factor IX inhibition	Human coagulation factor IX	1.399795372
Mitochondrial toxicity	Didanosine	1.39425507
Mitochondrial toxicity	Zidovudine	1.392822145
Intussusception	Oral rotavirus vaccine (live, attenuated)	1.383351232
Factor VIII inhibition	Factor VIII	1.333650334
Progressive external ophthalmoplegia	Enfuvirtide	1.332723419
Mitochondrial toxicity	Abacavir	1.332073084
Mitochondrial toxicity	Nevirapine	1.331868817
Progressive external ophthalmoplegia	Lopinavir, Ritonavir	1.331308707
Anti-erythropoietin antibody positive	Darbepoetin alfa	1.327737659
Ovarian hyperstimulation syndrome	Follitropin alfa	1.30848219
Polyomavirus-associated nephropathy	Prednisone	1.266286619
Progressive multifocal leukoencephalopathy	Natalizumab	1.258910773
Polyomavirus-associated nephropathy	Ciclosporin	1.254165319
Fanconi syndrome acquired	Emtricitabine, Tenofovir	1.25237055
Polyomavirus-associated nephropathy	basiliximab	1.250724525
Mitochondrial toxicity	Saquinavir	1.249075655
Mitochondrial toxicity	Indinavir	1.248574235
Nephrogenic systemic fibrosis	Gadobenic acid	1.248138318
Mitochondrial toxicity	Ritonavir	1.247641093
Nephrogenic systemic fibrosis	Gadoteridol	1.241194428
Polyomavirus-associated nephropathy	Prednisolone	1.231949178
Progressive external ophthalmoplegia	Delavirdine mesilate	1.199992854
Progressive external ophthalmoplegia	Atazanavir	1.198684534
Mitochondrial toxicity	Lopinavir, Ritonavir	1.198177836

PT = preferred term; MR = masking ratio.

debridement, Jaw operation). Most importantly, some events commonly reported were affected by a consequential masking in PfDB.

DISCUSSION

The mechanism by which signals can be masked under some circumstances has not been elucidated. Specifically, we demonstrated the following, which are of practical value in pharmacovigilance. Masking to some degree is common in both a large company and EV databases. Our findings confirmed issues that masking involves products for which the given adverse event is known. Even though this finding highlights a relation between a masking effect and an important disproportionality of reporting, because the masking effect is directly associated with a reported association, an important masking effect could potentially be induced by spurious associations resulting from stimulated reporting.

Masking of important magnitude was rare in both databases and primarily affected events that are rarely reported. As observed for drug-event pairs subject of important public awareness and attention, stimulated reporting contributes to a carry-over effect of the

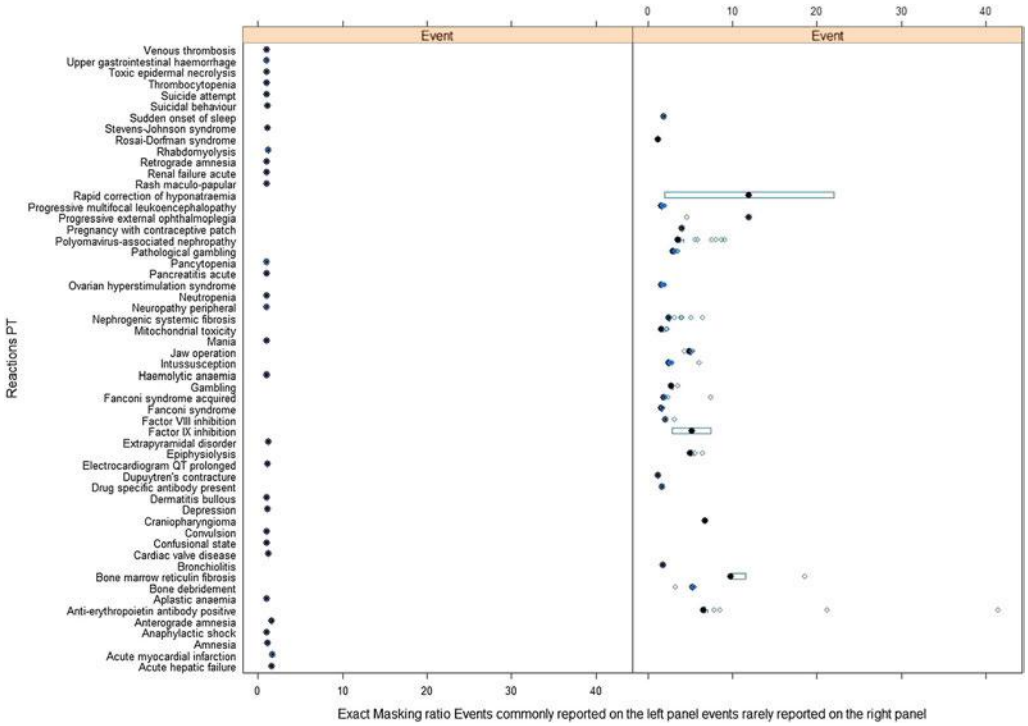


Figure 1. Distribution of the exact masking ratio computed the highest masking products for the set of events included in the study in EudraVigilance (very similar results have been obtained for the approximate masking ratio—graph not shown). The EU-ADR events are displayed on the left panel, the results obtained with the additional events included in the study in particular those events rarely reported to EudraVigilance are displayed in the right panel. The x-axis represents the value of the masking ratio

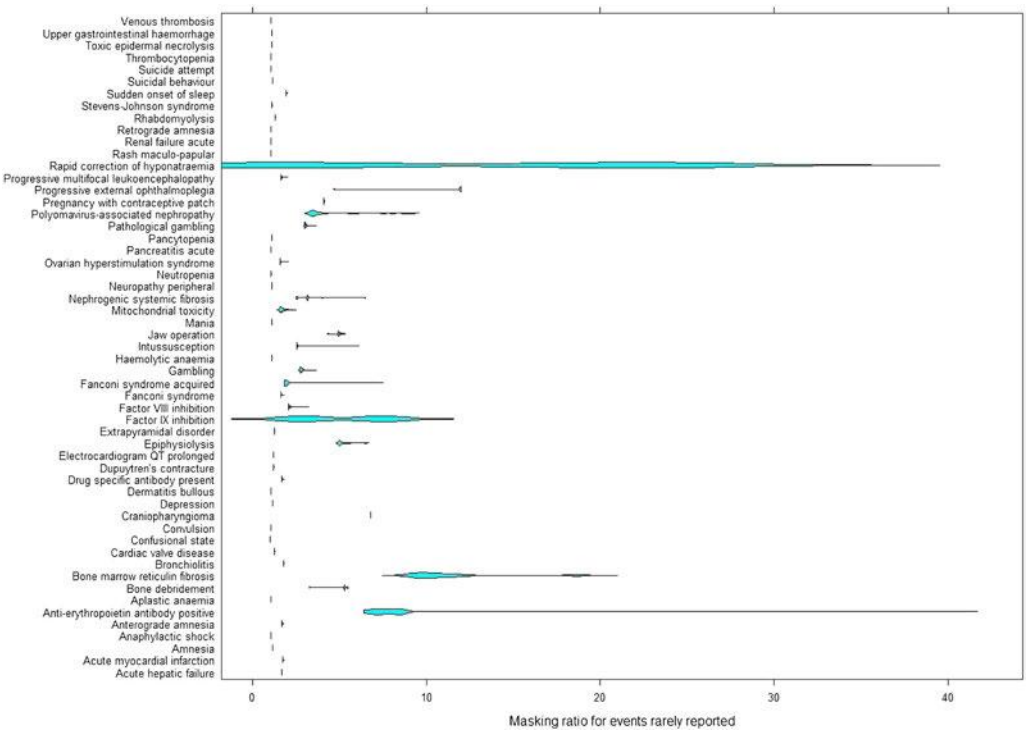


Figure 2. Violin plot of the exact masking ratios observed in EudraVigilance. The figure shows that the masking plays a significant role for the events rarely reported in the database. The x-axis shows the value of the masking ratios

DOI: 10.1002/pds

Table 6. The number of new signals of disproportionate reporting (SDRs) identified by the removal of the masking effect of the highest masking product (ranked by decreasing value of ratio of new SDRs revealed by unmasking). The table also includes the value of the masking ratio of this product for the preferred term of interest

Reaction PT	Nbr SDRs	New SDRs	Ratio	MR drug
Bone debridement	15	17	1.13	5.236867
Jaw operation	29	22	0.76	4.881738
Pregnancy with contraceptive patch	3	2	0.67	4.036374
Acute myocardial infarction	191	112	0.59	1.743396
Anti-erythropoietin antibody positive	21	9	0.43	6.518472
Bronchiolitis	93	33	0.35	1.765868
Intussusception	47	16	0.34	2.522654
Factor VIII inhibition	16	5	0.31	2.062682
Acute hepatic failure	292	88	0.3	1.652825
Cardiac valve disease	124	29	0.23	1.224371
Craniopharyngioma	13	3	0.23	6.264051
Sudden onset of sleep	89	18	0.2	1.908177
Rhabdomyolysis	470	80	0.17	1.289431
Pathological gambling	33	5	0.15	2.97808
Drug specific antibody present	61	9	0.15	1.687291
Bone marrow reticulin fibrosis	7	1	0.14	8.558506
Extrapyramidal disorder	246	35	0.14	1.256385
Fanconi syndrome	87	11	0.13	1.630408
Depression	280	35	0.13	1.125742
Amnesia	317	37	0.12	1.12965
Epiphysiolysis	19	2	0.11	4.819412
Electrocardiogram QT prolonged	476	47	0.1	1.176308
Nephrogenic systemic fibrosis	21	2	0.1	2.521013
Neuropathy peripheral	320	29	0.09	1.104694
Anterograde amnesia	57	5	0.09	1.65486
Polyomavirus-associated nephropathy	47	3	0.06	3.326415
Venous thrombosis	190	12	0.06	1.065716
Progressive multifocal leukoencephalopathy	100	6	0.06	1.647151
Gambling	17	1	0.06	2.692829
Ovarian hyperstimulation syndrome	38	2	0.05	1.57331
Mania	219	11	0.05	1.089134
Pancytopenia	538	27	0.05	1.101807
Pancreatitis acute	388	19	0.05	1.064298
Stevens-Johnson syndrome	862	41	0.05	1.111347
Thrombocytopenia	598	28	0.05	1.045518
Fanconi syndrome acquired	44	2	0.05	1.630408
Haemolytic anaemia	387	16	0.04	1.081483
Suicide attempt	572	23	0.04	1.061408
Neutropenia	408	16	0.04	1.067381
Dupuytren's contracture	53	2	0.04	1.1931
Dermatitis bullous	450	15	0.03	1.047514
Rash maculo-papular	543	18	0.03	1.049695
Suicidal behaviour	106	3	0.03	1.143097
Renal failure acute	727	18	0.02	1.034391
Convulsion	575	14	0.02	1.03647
Retrograde amnesia	90	2	0.02	1.067364
Confusional state	596	13	0.02	1.020954
Anaphylactic shock	631	12	0.02	1.042165
Upper gastrointestinal haemorrhage	181	3	0.02	1.104194
Toxic epidermal necrolysis	796	13	0.02	1.085932
Aplastic anaemia	306	2	0.01	1.046274

PT = preferred term; SDR = signal of disproportionate reporting; MR = masking ratio.

masking in the database (e.g. acute myocardial infarction). We have also observed an important masking effect for very specific drug-event associations involving events rarely reported (e.g. intussusception with rotavirus vaccines). Recent studies[6,13] relied on a repository of products likely to induce an important masking effect or a set of control events.[13] The evolution of the masking effect varies in an unpredictable way over time, as new reports that affect the value of the variables of the masking ratio are transmitted to the database. Therefore, the main risk associated to the use of a repository is to miss the presence of an emerging or underestimate the magnitude of an existing masking effect because of incoming reports. It is therefore safer from a public health perspective to run our proposed algorithm, which does not rely on any prior knowledge, each time a disproportionality analysis is conducted.

The two databases used in our study are structurally different: EV mostly contains serious reports reported by health care professionals for all medicinal products authorised in the European Union (mostly from 2001 onwards), whereas PfDB contains both serious and non-serious reports reported either by healthcare professionals or patients involving products for which Pfizer holds a licence. We could not study the specific influence of different types of reports (non-serious, medically unconfirmed). Masking does not appear to be consequential for a large and diverse set of events that have been assessed as important to public health but affected some of the designated medical events included in the study. The main differences observed between the two databases were mostly due to structural differences because of the products represented in the database rather than other factors (seriousness of the reports). We did not observe any important differences concerning the masking affecting for most of the events used in our study, including the events not usually considered to be medically serious (rash maculo-papular or confusional state)[14] suggesting a marginal role of the seriousness of the reports on the magnitude of the effect. Our finding of a greater degree of masking in pharmaceutical company database is also in keeping with theoretical expectations of Gould.[1]

We have observed that the removal of the masking marginally affects the original ranking given by the quantitative method and results in a lowering of the original threshold chosen for the disproportionality analysis with predictable consequences. Our choice of threshold to define consequential masking was directed by the distribution of the MR and the rate of true and false positives potentially likely to be revealed by the removal of

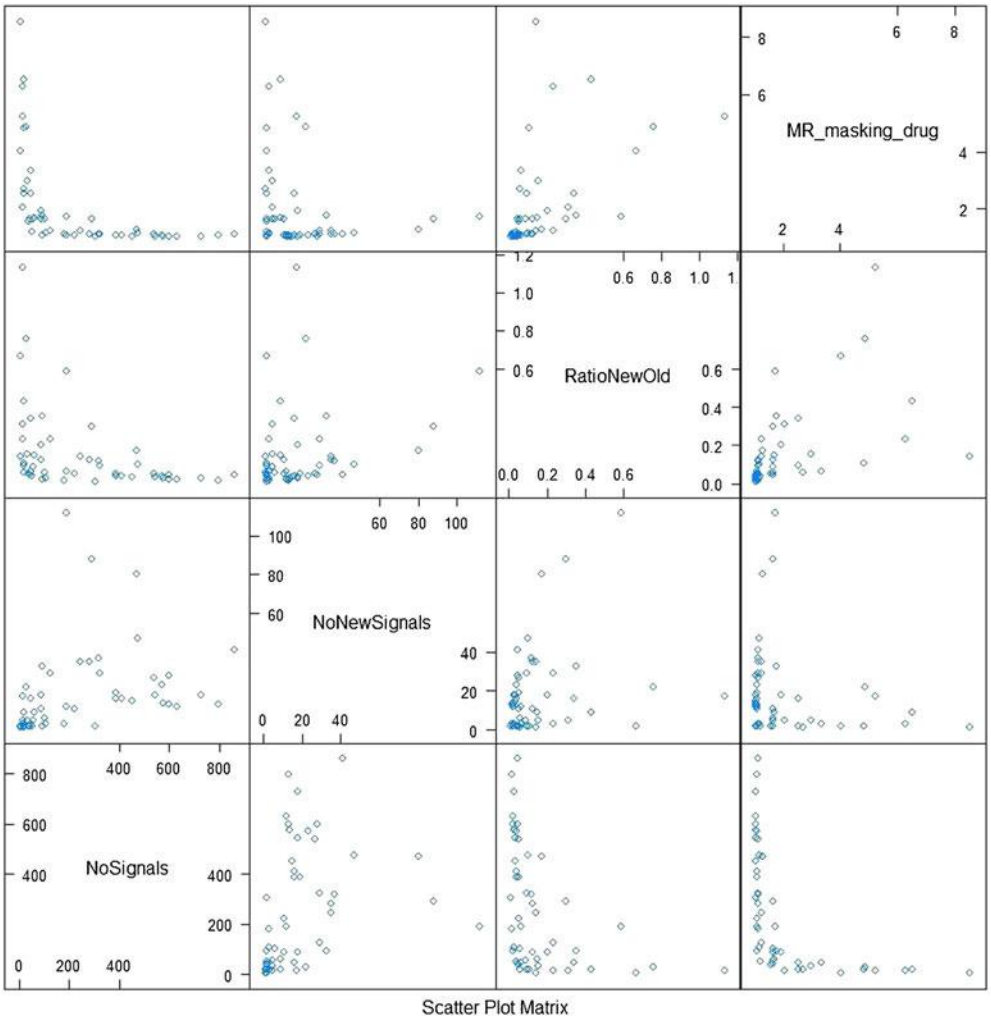


Figure 3. Scatter plot matrix of the number of signals of disproportionate reporting (SDRs) originally present in the dataset (NoSignals) from Eudravigilance, the number of new SDRs revealed by the removal of the highest masking drug (NoNewSignals), the ratio between these two variables and the value of the exact masking ratio of the highest masking drug (masking ratio masking drug). No clear correlation between these variables can be observed

Table 7. New signals of disproportionate reporting revealed by the removal of the highest masking product for the Medical Dictionary for Regulatory Activities preferred terms 'Anaphylactic shock'

Reaction PT	INN	PRR (before)	New PRR	Difference
Anaphylactic shock	Bupivacaine	1.92	2.001168	0.081168
Anaphylactic shock	Cyanocobalamin	1.93	2.011406	0.081406
Anaphylactic shock	Fenspiride hydrochloride	1.98	2.063492	0.083492
Anaphylactic shock	Fusafungine	1.94	2.021811	0.081811
Anaphylactic shock	Hexetidine	1.94	2.021805	0.081805
Anaphylactic shock	Lactobacillus acidophilus	1.98	2.063492	0.083492
Anaphylactic shock	Lactose monohydrate	1.92	2.000984	0.080984
Anaphylactic shock	Miconazole	1.99	2.073965	0.083965
Anaphylactic shock	Ofloxacin hydrochloride	1.95	2.032261	0.082261
Anaphylactic shock	Pegaspargase	1.98	2.063503	0.083503
Anaphylactic shock	Penicillin nos	1.96	2.042666	0.082666
Anaphylactic shock	Triamcinolone	1.93	2.011478	0.081478

PT = preferred term; INN = International Nonproprietary Name; PRR = proportional reporting ratio.

DOI: 10.1002/pds

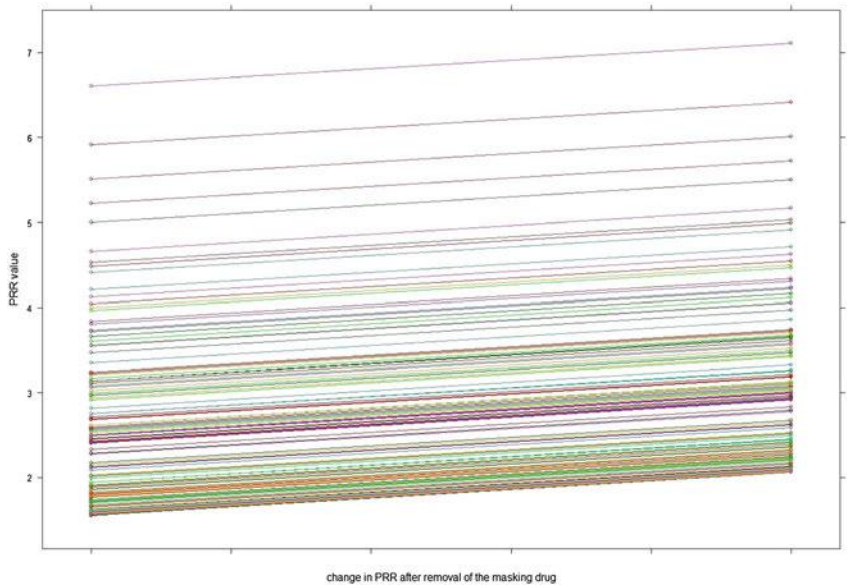


Figure 4. This graph displays the value of the proportional reporting ratio (y-axis) before (left hand side of the x-axis) and after (right hand side of the x-axis) the removal of the highest masking drug (paracetamol) for the event of interest (Medical Dictionary for Regulatory Activities preferred term acute hepatic failure) (EudraVigilance). The exact masking ratio has been used for this computation. The removal of the masking drug increases the number of signals detected by the quantitative method (the value of the proportional reporting ratio will be above the [arbitrary] chosen threshold). However, this graph shows that the ranking of the drug-event combinations is marginally affected by the removal of the drug inducing the highest masking effect (the selected Medical Dictionary for Regulatory Activities preferred term acute hepatic failure is an event commonly reported in EudraVigilance)

masking (unpublished results). Because of their important prevalence, MRs below 1.1 would indistinctly unravel an important number of DECs that are extremely close to threshold chosen for the disproportionality analysis hence increasing the number of false positive.

The two databases used in our study are very large. In addition, our analysis involves a small subset of the MedDRA terminology. The results do not probably apply to smaller databases. We used a PRR to define an SDR as the masking mechanisms associated with

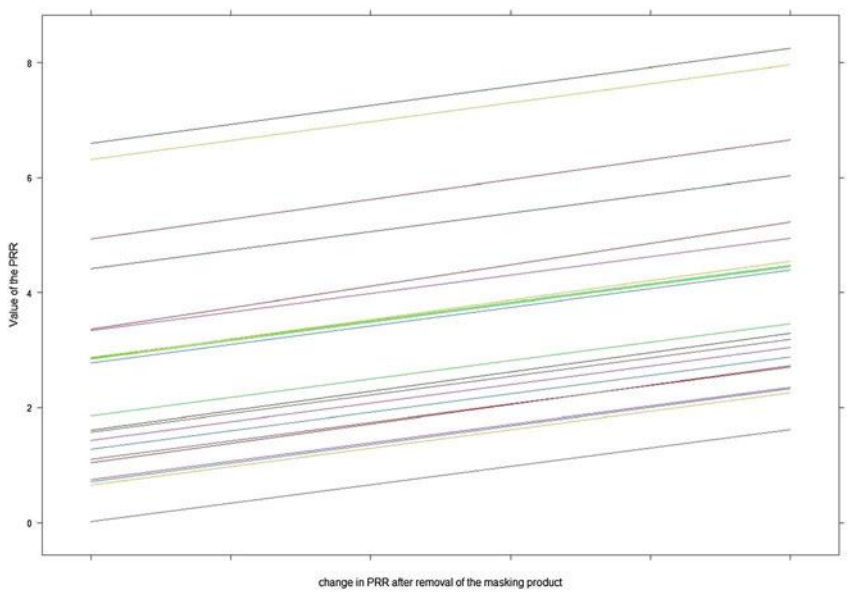


Figure 5. Effect of the removal of the highest masking product for an event rarely reported in EudraVigilance. The graph displays the value of the proportional reporting ratio before (left hand side of x-axis) and after (right hand side of x-axis) the removal of the drug inducing the highest masking effect (medicinal products containing rh-growth hormone) for the event of interest (Medical Dictionary for Regulatory Activities preferred term epiphysiolysis). The removal of the masking drug increases the number of signals detected by the quantitative method (the value of the proportional reporting ratio will be above the [arbitrary] chosen threshold). However, this graph clearly shows that the ranking of the drug-event combinations is marginally affected by the removal of the masking drug

DOI: 10.1002/pds

Table 8. Breakdown of the values of the (approximate) masking ratio $\left(1+\frac{n_{21}}{n_{11}+n_{31}}\right)$ in Pfizer database and in EudraVigilance

PfAST			EV		
Maximum Approximate MR	# of DECs	% total	Maximum Approximate MR	# of DECs	% total
>1.1	204	1.07%	>1.1	89	0.29%
>2	25	0.13%	>2	20	0.07%
>4	8	0.04%	>4	7	0.02%
>5	6	0.03%	>5	5	0.02%
>6	6	0.03%	>6	3	0.01%
>7	6	0.03%	>7	2	0.01%
>8	6	0.03%	>8	2	0.01%
>9	3	0.02%	>9	2	0.01%

MR = masking ratio; DEC = drug-event combination.

other methods (reporting odds ratio and relative reporting ratio) are very similar. Furthermore some data mining protocols involve additional metrics that we did not study ($\chi 2$, statistical unexpectedness). Bayesian algorithms involve shrinkage effects, which could complicate the development of an algorithm. However, our study confirms results obtained with the IC algorithm on the WHO database (VigiBase).[15]

We did not find any clear relation between the value of the MR and the number of SDRs revealed by the removal of the masking. To truly understand the health impacts of this phenomenon, we must distinguish between masking, and what Hochberg *et al*. termed 'consequential masking', namely masking that results in suppression of SDRs involving associations that represent credible 'signals of suspected causality'.[2] Hochberg *et al*. reported that unmasking exercises did not seem to uncover novel associations with significant external evidentiary support, on a very small subgroup with this information available. For a small subset of events, Wang found that unmasked associations tended to be already known had limited evidentiary support. Recent studies[6,13] and our results show a potential public health benefit of removing a consequential masking effect. The benefit gained from the removal of the masking effect needs to be properly quantified. We did not report any characterisation of our SDRs revealed by the masking and acknowledge this limitation of our study. The characterisation of new SDRs poses methodological challenges. There is little agreement between assessors concerning the adjudication of new SDRs together with other issues (sample size and difficulty to conduct an assessment that is not influenced by a hindsight bias). Secondly, the medical adjudication of SDRs has been shown to be conservative, an attitude that may decrease the efficiency of quantitative methods when used prospectively. Gould identified 15 adverse events for which

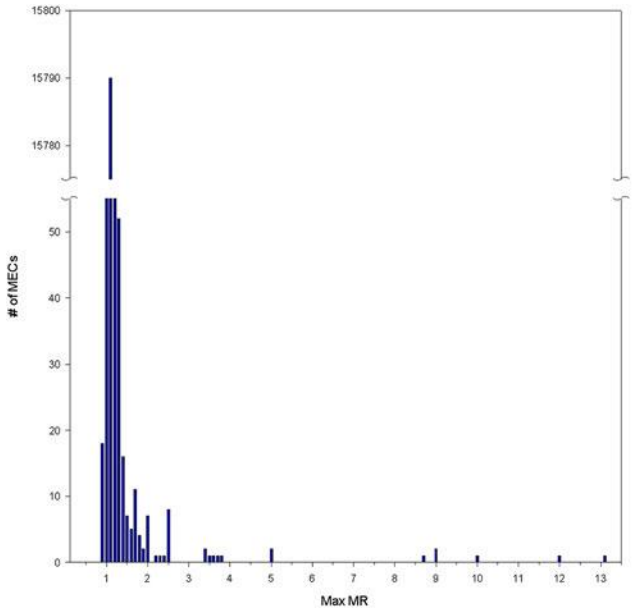


Figure 6. Distribution of the masking ratio on a company database. Compared to the results obtained in EudraVigilance, the distribution is skewed towards higher values

removal of masking drug resulted in the appearance of an SDR. Considering that the unravelling of known effects is an indirect surrogate to assess the real efficiency of signal detection activities, the evaluation of the value of removing the masking effect should be performed (in a blinded way) in a real life situation by assessing SDRs corresponding to true new signals. Therefore a natural extension of the current work would be an analysis that tabulates and prospectively adjudicates unmasked SDRs.

Various approaches to identify and remove the masking have been used in the past. Hochberg and Hauben used the maximum number of reports as a criterion in one exercise to explore masking for the DEC exenatide-pancreatitis in the US Food and Drug Administration's (FDA) AERS database.[16] Wang *et al*. used a disproportionality metric in combination with a measure of statistical unexpectedness but because the latter is influenced by sample size, the authors suggested the top-ranked report count as another option.[3]

Our analysis focused on masking by individual drugs. This may be especially pertinent to a large health authority database in the sense that masking drug groups may be the scaled-up analogue of the phenomenon of single-drug masking in a pharmaceutical company database. The construction of the contingency table and our mathematical formulation can be easily extended to multiple drugs that may have a masking effect together as a group. Wang *et al*. studied the masking effect of groups of drugs, using various

Table 9. Differences of masking observed between Pfizer database and EudraVigilance. The difference of approximate masking ratio is displayed on the right hand column, this difference is positive when the magnitude of the masking was higher in Pfizer database than in EudraVigilance, negative when the contrary was observed. Cells highlighted in pink show differences greater than 1. N/A denotes the absence of reports associated with the Medical Dictionary for Regulatory Activities preferred term in the database. ApproxMR denotes $MR_{PRRApprox1}$ (Table 2)

Reaction PT	Approx. MR EV	Approx. MR Pfizer db	Difference
Acute hepatic failure	1.653	1.255	-0.398
Acute myocardial infarction	1.743	1.244	-0.500
Amnesia	1.130	1.190	0.061
Anaphylactic shock	1.042	1.072	0.030
Anterograde amnesia	1.655	6.781	5.126
Anti-erythropoietin antibody positive	6.518	N/A	
Aplastic anaemia	1.046	1.122	0.075
Bone debridement	5.237	1.500	-3.737
Bone marrow reticulin fibrosis	8.559	N/A	
Bronchiolitis	1.766	1.416	-0.350
Cardiac valve disease	1.224	12.052	10.827
Confusional state	1.021	1.100	0.080
Convulsion	1.036	1.143	0.107
Craniopharyngioma	6.264	6.500	0.236
Depression	1.126	1.226	0.100
Dermatitis bullous	1.048	1.118	0.070
Drug specific antibody present	1.687	2.936	1.248
Dupuytren's contracture	1.193	1.159	-0.034
Electrocardiogram QT prolonged	1.176	1.308	0.131
Epiphysiolysis	4.819	9.978	5.158
Extrapyramidal disorder	1.256	1.462	0.205
Factor IX inhibition	2.625	N/A	
Factor VIII inhibition	2.063	N/A	
Fanconi syndrome	1.630	1.481	-0.149
Fanconi syndrome acquired	1.630	N/A	
Gambling	2.693	3.328	0.636
Haemolytic anaemia	1.081	1.065	-0.016
Intussusception	2.523	1.600	-0.923
Jaw operation	4.882	2.125	-2.757
Mania	1.089	1.290	0.201
Mitochondrial toxicity	1.575	1.600	0.025
Nephrogenic systemic fibrosis	2.521	N/A	
Neuropathy peripheral	1.105	1.255	0.151
Neutropenia	1.067	1.184	0.117
Ovarian hyperstimulation syndrome	1.573	1.500	-0.073
Pancreatitis acute	1.064	1.200	0.136
Pancytopenia	1.102	1.501	0.399
Pathological gambling	2.978	1.737	-1.241
Polyomavirus-associated nephropathy	3.326	2.600	-0.726
Pregnancy with contraceptive patch	4.036	N/A	
Progressive external ophthalmoplegia	11.991	N/A	
Progressive multifocal leukoencephalopathy	1.647	1.633	-0.014
Rapid correction of hyponatraemia	21.999	N/A	
Rash maculo-papular	1.050	1.072	0.022
Renal failure acute	1.034	1.101	0.066
Retrograde amnesia	1.067	1.771	0.703
Rhabdomyolysis	1.289	2.049	0.760
Rosai-Dorfman syndrome	1.167	1.500	0.333
Stevens-Johnson syndrome	1.111	1.366	0.255
Sudden onset of sleep	1.908	2.200	0.292
Suicidal behaviour	1.143	1.590	0.447
Suicide attempt	1.061	1.203	0.142
Thrombocytopenia	1.046	1.085	0.040
Toxic epidermal necrolysis	1.086	1.072	-0.013
Upper gastrointestinal haemorrhage	1.104	1.271	0.167
Venous thrombosis	1.066	1.112	0.046

criteria including pharmacological relatedness, in the US FDA AERS database.[3] The latter was included because it most closely matched the original formulation of Gould in which the masking and masked drug were members of the same drug class.

We did not study the impact of the hypergranular MedDRA hierarchy on the results. Specifically, whether the SDRs gained as a result of unmasking were truly novel would depend on whether there were medically related PT with SDRs prior to the unmasking exercise. This factor may have substantial impacts and has been demonstrated to significantly impact gained and lost SDRs due to stratification by basic covariates (Hauben M. unpublished data).

One important danger of the identification and tackling of the masking is to artificially inflate the type I error rate. This has already been identified as a pitfall in previous disproportionality analyses[10], and great caution should be exercised to avoid post hoc analytical retrofitting. The removal of the masking should not change the current practice of disproportionality analysis. The confirmation of the new SDRs should be performed in accordance with the agreed scientific consensus on signal detection.[17,18]

Furthermore, it is important to point out that there is an inherent bias in discussions of masking as they fail to accommodate unusual data distributions that may have an effect that is opposite to masking, that is, to increase the statistical representation of certain events. In other words, there may be opposing effects that balance each other, but discussions seem to focus on identifying only one of these (the masking) effect. The pitfall is that selectively removing masking drugs without searching and removing unusual effects in the opposite direction may 'unbalance' the data and lead to spurious associations.

CONCLUSION

We acknowledge that our exact algorithm is resource demanding and might require some simplification. In addition, our algorithm might usefully be extended to other methods (confidence intervals, Bayesian) and other types of databases (small, databases for vaccines, longitudinal), which are used in routine signal detection activities.

Although we estimated the prevalence of masking in two databases, we did not study the impact that this would have on real-world signal detection. Specifically, without knowing the distribution of baseline PRRs around the SDR-defining threshold, we do not know what percentage of masking would actually be 'consequential masking' in the sense of newly emergent SDRs. We also limited our examination to removal of individual drugs.

In future studies, it will be useful to quantify the real public health value of removing the masking and identify those situations in which the removal is beneficial. However, our study provides an important insight and a rational approach to the identification, quantification of the masking effect in SRS databases.

CONFLICT OF INTEREST

The following authors, F. M. and J. M. D., have no conflicts of interest with the pharmaceutical industry (declaration of interest available from EMA). M.H. and E.H. are working in World Worldwide Safety and Regulatory, Pfizer Inc. M.H. is on faculty in the Department of Medicine, New York University School of Medicine, New York City, the department of Community and Family Medicine, New York Medical College, Valhalla, NY, and the School of Information Systems, Computing and Mathematics, Brunel University, London, England. L. V. H. is working for GlaxoSmithKline Biologicals. None of the authors have any conflict of interests with any statistical software provider.

The views expressed in this article are the personal views of the author(s) and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medical Agency or one of its committees or working parties or Pfizer Inc. or GlaxoSmithKline Vaccines.

KEY POINTS

- Our estimate of prevalence of significant masking showed that the phenomenon may be rare.
- An important masking effect was consistently associated to products for which the reaction is known or has been extensively or over reported.
- Masking mainly (but not only) affected events rarely reported in our large spontaneous systems databases.
- Differences affecting important medical events were observed between EV and PfDB.
- The original ranking provided by the quantitative methods included in our study was marginally affected by the removal of the masking product.

ETHICS STATEMENT

No individual data on patients were used in this study that did not require any prior approval from an ethics committee.

ACKNOWLEDGEMENTS

Valuable comments on this work were received from Martin Posch and Jim Slattery. The authors thank Jim Slattery for his continuous support when conducting the study. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) for the Innovative Medicine Initiative (www.imi.europa.eu) under Grant Agreement no 115004. The research leading to these results was conducted as part of the PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium, www.imi-protect.eu), which is a public-private partnership coordinated by the European Medicines Agency.

REFERENCES

1. Gould AL. Practical pharmacovigilance analysis strategies. *Pharmacoepidemiol Drug Saf* 2003; **12**(7): 559 574.
2. Hauben M, Aronson JK. Defining 'Signal' and its subtypes in pharmacovigilance based on a systematic review of previous definitions. *Drug Saf* 2009; **32**(2): 99 110.
3. Wang HW, Hochberg AM, Pearson RK, Hauben M. An experimental investigation of masking in the US FDA adverse event reporting system database. *Drug Saf* 2010; **33**(12): 1117 1133.
4. Pariente A, Didailler M, Avillach P, *et al*. A potential competition bias in the detection of safety signals from spontaneous reporting databases. *Pharmacoepidemiol Drug Saf* 2010; **19**(11): 1166 1171.
5. Zeinoun Z, Seifert H, Verstraeten T. Quantitative signal detection for vaccines: effects of stratification, background and masking on GlaxoSmithkline's spontaneous reports database. *Hum Vaccin* 2009; **5**(9): 599 607.
6. Pariente A, Avillach P, Salvo F, *et al*. Effect of competition bias in safety signal generation: analysis of a research database of spontaneous reports in France. *Drug Saf* 2012; **35**(10): 855 864.
7. Maignen F, Hauben M, Hung E, Van HolleL, Dogne JM. A conceptual approach to the masking effect of measures of disproportionality. Submitted for publication to *Pharmacoepidemiol Drug Saf*.
8. Alvarez Y, Hidalgo A, Maignen F, Slattery J. Validation of statistical signal detection procedures in EudraVigilance post-authorisation data. A retrospective evaluation of the potential for earlier signalling. *Drug Saf* 2010; **33**(6): 475 487.
9. European Medicines Agency. EudraVigilance: pharmacovigilance in the European economic area. http://eudravigilance.ema.europa.eu/human/index.asp [26 March 2013].
10. Guideline on the use of statistical signal detection tools in the EudraVigilance data analysis system (Doc. Ref. EMEA/106464/2006 rev. 1). www.ema.europa.eu/pdfs/human/phvwp/10646406enfin.pdf [17 May 2011].
11. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 2001; **10**(6): 483 486.
12. Trifirò G, Pariente A, Coloma PM, *et al*. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? *Pharmacoepidemiol Drug Saf* 2009; **18**(12): 1176 1184.
13. Ooba N, Kubota K. Selected events and reporting odds ratios in signal detection methodology. *Pharmacoepidemiol Drug Saf* 2010; **19**: 1159 1165.
14. Current challenges in pharmacovigilance: pragmatic approaches. Report of CIOMS Working Group V. CIOMS, Geneva 2001.
15. Juhlin K, Ye X, Star K, Norén GN. Outlier removal to uncover patterns in adverse drug reaction surveillance a simple unmasking strategy. *Pharmacoepidemiol Drug Saf* 2013; **22**: 1119 1129.
16. Hauben M, Hochberg A. The importance of reporting negative findings in data mining: the example of Exenatide and pancreatitis. *Pharm Med* 2008; **22**(4): 215 219.
17. Almenoff J, Tonning JM, Gould AL, *et al*. Perspectives on the use of data mining in pharmacovigilance. *Drug Saf* 2005; **28**(11): 981 1007.
18. Report of CIOMS working group VIII. Practical Aspects of Signal Detection in Pharmacovigilance, Geneva 2010.

From: "Ryan, Qin" <[redacted]>
To: "Szarfman, Ana" <[redacted]>, "Niu, Manette" <[redacted]>, "Zinderman, Craig E" <[redacted]>, "Baer, Bethany" <[redacted]>, "Menschik, David" <[redacted]>
Cc: "Stockbridge, Norman L" <[redacted]>
Subject: RE: Our conversation about VAERS of this afternoon.
Date: Sat, 27 Mar 2021 01:00:44 +0000
Importance: Normal
Inline-Images: image001.png

Sure Ana, It will be my pleasure to entertain any questions regarding my experience with Empirica Signal and Empirica Study.

Have a great weekend!

Qin

From: Szarfman, Ana <[redacted]>
Sent: Friday, March 26, 2021 3:49 PM
To: Niu, Manette <[redacted]>; Zinderman, Craig E <[redacted]>; Baer, Bethany <[redacted]>; Menschik, David <[redacted]>
Cc: Stockbridge, Norman L <[redacted]>; Ryan, Qin <[redacted]>
Subject: Our conversation about VAERS of this afternoon.

Hi Manette, Beth, and Craig,

Please refer to the attached files that I displayed this afternoon.

As we talked, the attached excel comparisons between RGPS and MGPS were generated by Bill DuMouchel using the VAERS public domain data incorporated into Empirica Signal.

RGPS is included with the public domain version of Empirica Signal.

Bill and I extensively studied the increased value of RGPS over MGPS for reducing false positives and negative signals.

Oligonucleotides (regulated by CDER) and mRNA vaccines (regulated by CBER) share some common important characteristics, including severe thrombocytopenia; and we are interested in using several resources to understand them better.
Qin Ryan, in the cc is the principal investigator of a project studying this effect with oligonucleotides, having me as a collaborator.

VAERs offers a unique opportunity to study the value of RGPS in improving the detection of early signals in a different, important environment during a pandemic situation whereas the early detection of novel signals is tremendously important for all.

The new methodology being proposed by Bill to study across multiple applications offers the opportunity to benefit from automation, immediate access to a cross comparison of safety signals across multiple treatment arms within multiple applications, and the identification of unbalanced risk factors at baseline. Qin Ryan worked with an earlier prototype of the system, and will answer questions that you may have.

Let me know if you need any additional feedback.

Warmest regards and thanks,

--Ana

Ana Szarfman, MD, PhD, FAMIA,
Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), and Fellow of the American Medical Informatics Association (2020)
Medical Officer, Safety Data Mining Developer and Medical Informatics Analyst,
Celebrating nearly a quarter of a century of successful implementation of safety data mining, interactive patient profiles, and other automated analytical tools.
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted] (office)
[redacted] (personal cell phone and WhatsApp)

From: Bill Dumouchel <[redacted]>
Sent: Wednesday, March 24, 2021 9:38 AM
To: Szarfman, Ana <[redacted]>
Subject: [EXTERNAL] Fw: WVAERS 2021W09 data loaded to slc06lhx

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you recognize the sender and know the content is safe.

From: Bill Dumouchel <[redacted]>
Sent: Tuesday, March 23, 2021 4:27 PM
To: Steve Bright <[redacted]>; Rave Harpaz <[redacted]>; Szarfman, Ana <[redacted]>
Cc: Mohammad Al-Ansari <[redacted]>; Alexander Nip <[redacted]>
Subject: Re: WVAERS 2021W09 data loaded to slc06lhx

I created runID#307 which is the same as #304 but with the new data.
I'm attaching an excel file with 49 examples of extreme masking--that is, RGPS shows a signal where MGPS doesn't, and the confidence intervals don't overlap.
The Covid custom term is just a label for any covid vaccine, no matter the manufacturer. Most of the significant masking involves that, because it gets a larger sample size and thus shorter confidence intervals, with less chance for overlap.

My main worry about these seemingly significant adverse events it that the age grouping is quite course, agegroup6 lumps everyone over 65 together.
So our adjustment for age may not be good.

Appendicitis doesn't show up with the extreme requirement that I imposed on the above search, but, relaxing it slightly, there are fairly extreme estimates for Pfizer & Appendicitis, as shown in sheet two of the attached excel file.

Finally, I've attached a zip file that contains all of the covid-AEs in the results of the run. (50,515 rows)

Enjoy!
Bill

From: Ruixia Song <[redacted]>
Sent: Monday, March 22, 2021 11:26 AM
To: Bill Dumouchel <[redacted]>; Steve Bright <[redacted]>; Rave Harpaz <[redacted]>
Cc: Mohammad Al-Ansari <[redacted]>; Alexander Nip <[redacted]>
Subject: WVAERS 2021W09 data loaded to slc06lhx

Hi All,
WVAERS 2021W09 data has been loaded to slc06lhx.
Ruixia

From: "Baer, Bethany" <[redacted]>
To: "Menschik, David" <[redacted]>
Subject: FW: Data mining question
Date: Fri, 15 Mar 2024 17:29:50 +0000
Importance: Normal
Attachments: Harpaz_datamining_Covid_2022.pdf
Inline-Images: image001.png; image002.jpg; image003.jpg; image004.jpg; image005.jpg; image006.jpg

---

Hi David,
I am just responding to you so you can decide if you want to use this article as an example or not. It goes back to the discussions about Ana's involvement in VAERS data mining and her interest in updating data mining methods.
Bethany

From: Zinderman, Craig <[redacted]>
Sent: Friday, March 15, 2024 1:22 PM
To: Nair, Narayan <[redacted]>; Menschik, David <[redacted]>; Baer, Bethany <[redacted]>
Subject: RE: Data mining question

I'm not aware of literature articles (although I can't say I've looked for it either). I recall Anna talking about masking in the few interactions we had with her, but I don't remember there being references.

Thanks,
Craig

Craig Zinderman, MD, MPH

Associate Director for Medical Policy
Office of Biostatistics and Pharmacovigilance
FDA/Center for Biologics Evaluation and Research
[redacted]

From: Nair, Narayan <[redacted]>
Sent: Friday, March 15, 2024 1:04 PM
To: Menschik, David <[redacted]>; Zinderman, Craig <[redacted]>; Baer, Bethany <[redacted]>
Subject: Data mining question

Good afternoon,
I know in the past we have discussed one of the possible limitations of data mining currently is the vast number of VAERS reports from the COVID vaccines may limit our ability to detect statistical alerts because disproportionality scores may be driven towards the null. Do you know if there is a public reference that discusses this limitation? I have found some references that discuss general limitations for data mining but not sure if there is one that talks about how a large volume of reports from a single class of products could masks results.

Narayan Nair, MD (he/him/his)
Division Director

Division of Pharmacovigilance
Office of Biostatistics and Pharmacovigilance
Center for Biologics Evaluation and Research
U.S. Food and Drug Administration

[redacted]

Drug Safety (2022) 45:765–780
https://doi.org/10.1007/s40264-022-01186-z

ORIGINAL RESEARCH ARTICLE

Signaling COVID-19 Vaccine Adverse Events

Rave Harpaz[1] · William DuMouchel[1] · Robbert Van Manen[1] · Alexander Nip[1] · Steve Bright[1] · Ana Szarfman[2] · Joseph Tonning[3] · Magnus Lerch[1,4]

Accepted: 8 May 2022 / Published online: 23 June 2022

Abstract

Introduction Statistical signal detection is a crucial tool for rapidly identifying potential risks associated with pharmaceutical products. The unprecedented environment created by the coronavirus disease 2019 (COVID-19) pandemic for vaccine surveillance predisposes commonly applied signal detection methodologies to a statistical issue called the masking effect, in which signals for a vaccine of interest are hidden by the presence of other reported vaccines. This masking effect may in turn limit or delay our understanding of the risks associated with new and established vaccines.

Objective The aim is to investigate the problem of masking in the context of COVID-19 vaccine signal detection, assessing its impact, extent, and root causes.

Methods Based on data underlying the Vaccine Adverse Event Reporting System, three commonly applied statistical signal detection methodologies, and a more advanced regression-based methodology, we investigate the temporal evolution of signals corresponding to five largely recognized adverse events and two potentially new adverse events.

Results The results demonstrate that signals of adverse events related to COVID-19 vaccines may be undetected or delayed due to masking when generated by methodologies currently utilized by pharmacovigilance organizations, and that a class of advanced methodologies can partially alleviate the problem. The results indicate that while masking is rare relative to all possible statistical associations, it is much more likely to occur in COVID-19 vaccine signaling, and that its extent, direction, impact, and roots are not static, but rather changing in accordance with the changing nature of data.

Conclusions Masking is an addressable problem that merits careful consideration, especially in situations such as COVID-19 vaccine safety surveillance and other emergency use authorization products.

1 Introduction

As the world contends with ending the coronavirus disease 2019 (COVID-19) pandemic, understanding the risks associated with COVID-19 vaccines is critically urgent. The Vaccine Adverse Event Reporting System (VAERS), co-administered by the US Food and Drug Administration (FDA) and the Centers for Disease Control and Prevention (CDC), is one of several systems used to monitor adverse events (AEs) that occur after vaccination, including the COVID-19 vaccines. Like other safety surveillance systems, VAERS offers the opportunity to rapidly identify potential risks associated with vaccines—a process usually known as signal detection.

According to the World Health Organization (WHO), a safety signal is defined as reported information on a possible causal relationship between an AE and a product, of which the relationship is unknown or incompletely documented [1]. At a very high level, signal detection is the active pursuit of safety signals. The process of signal detection is multifaceted and interdisciplinary and can take many forms, be performed at different levels of evidence and data, and be accomplished in different ways. The specific application considered in this study has previously been termed *data mining*, *screening*, *disproportionality analysis*, and *quantitative signal detection*. It involves the use of statistical techniques that cast a wide net to rapidly explore large databases of reported AEs for statistical patterns or anomalies that may be indicative of new risks that warrant further attention. This approach to signal detection has been routinely applied to safety surveillance systems for over 20 years and has become a de facto standard [2]. To distinguish this approach from other

✉ Rave Harpaz
rave.harpaz@oracle.com

1 Oracle Health Sciences, Burlington, MA, USA

2 U.S. FDA, Silver Spring, MD, USA

3 U.S. Public Health Service/U.S. FDA retired, Silver Spring, MD, USA

4 Lenolution GmbH, Berlin, Germany

Key Points

The masking effect is a statistical issue associated with commonly applied signal detection methodologies in which signals for a product of interest are hidden by the presence of other reported products.

Due to vaccine novelty, and an unprecedented dynamic of reporting, statistical signals of adverse events related to coronavirus disease 2019 (COVID-19) vaccines are more prone to masking and, therefore, to being undetected or delayed.

A more advanced class of signal detection methodologies, based on regression, can address masking and expose strong statistical associations that would otherwise be deemed uninteresting.

The extent, direction, impact, and root causes of masking change in accordance with the changing nature of data.

approaches and activities related to signal detection, we will simply refer to it as *statistical signal detection*, highlighting its statistical foundation. With that, it is important to emphasize that since statistical signal detection is ultimately based on reporting patterns that are influenced by reporting dynamics, it is characterized as hypothesis generating. The presence of a strong statistical signal does not automatically imply a causal relationship and must always be evaluated by other methods, including the clinical review of case-level reports, scientific literature, and relevant studies [2–4]. Likewise, the absence of a strong statistical signal does not automatically rule out the existence of a safety issue. It is also worth mentioning that statistical signal detection can be repurposed to inform suspicions originating from other sources, but that is not the focus of our investigation.

Methodologies for statistical signal detection are based on computing surrogate measures of statistical association between specific pharmaceutical products and AEs that are reported into safety surveillance systems [5]. The measures are typically interpreted as signal scores, with larger values representing stronger statistical associations, which may be more likely to represent true causal associations. In practice, a signal score threshold is often used to screen associations that warrant further attention.

Methodologies for statistical signal detection currently deployed by safety surveillance organizations are largely based on disproportionality statistics. These methodologies use frequency analysis of 2×2 contingency tables to quantify the degree to which a product–AE combination co-occurs disproportionately as compared with that expected if there were no statistical association. To illustrate, we use the relative reporting ratio (RRR), which is a disproportionality statistic underlying several methodologies. The RRR is defined as the ratio of the number of reports mentioning a specific (target) product–event combination to an expected number of reports for the same combination under the assumption that the product and AE occur independently. Based on the values displayed in Table 1, the RRR is formally given by,

$$\mathrm{RRR} = \frac{(a+b+c+d)\cdot a}{(a+b)\cdot(a+c)} \quad (1)$$

and a number of enhancements, such as Bayesian smoothing and stratification, lead to several signal detection methodologies currently utilized by safety surveillance organizations [5].

Given its impact on public health, signal detection is still an active area of research, and since its inception, multiple guidance documents [3, 6–8] have been published with practice recommendations as well as admonitions concerning data and methodological limitations.

Undetected or delayed signals and false alerts are the two primary concerns with signal detection and two objective measures with which the reliability of signal detection can be evaluated. Undetected or delayed signals are especially disconcerting given their direct impact on public health. This study is concerned with those signals undetected by statistical signal detection, which we will refer to as *statistical signals*. Fortunately, multiple other surveillance and signaling efforts are deployed to reduce the chance of undetected signals.

Undetected statistical signals can stem from several sources. Incomplete data and the voluntary nature of reporting to surveillance systems are the primary sources of undetected signals. However, undetected statistical signals can also stem from methodological limitations and, in particular, a widely acknowledged problem called ‘masking’ [3, 9, 10].

Masking is an artifact of commonly applied disproportionality statistics that rely on the analysis of 2×2 contingency tables in which signals of disproportionate reporting may be hidden (hence, masked) by the presence of other non-target products frequently reported with the target AE. As described above, disproportionality statistics based on 2×2 contingency tables are defined as the ratio of the target AE rate for the target product to the background rate for target AE. However, defining the background rate can be problematic. We are prone to think of the background as being scattered randomly across all the non-target products, but this may not be the case. What if one non-target product has half of the target AEs appearing with all non-target products? In that case, under certain conditions eliminating that particular non-target product from the reports database would roughly double our target disproportionality. It would

Table 1 2 × 2 contingency table used to compute disproportionality statistics for signal detection

	Reports with target AE	Reports without target AE	
Reports with target product	*a*	*b*	*a* + *b*
Reports without target product	*c*	*d*	
	a + *c*		*a* + *b* + *c* + *d*

AE adverse event

seem reasonable to do so, because otherwise, the non-target product would be masking the target's true product disproportionality by cutting its value in half. Therefore, a possible solution to address masking is to first identify the 'offending' products and then remove reports containing those products from the calculation of disproportionality statistics. This solution may work in a limited set of scenarios, but is practically infeasible in the general case as it may require examining a combinatorically prohibitive set of product–AE pairs. A more direct and computationally feasible approach to address masking necessitates the use of a more advanced class of methodologies, such as regression, which go beyond the analysis of 2 × 2 contingency tables and can compute statistical associations adjusted for the presence of other products. This investigation makes use of one such methodology called *Regression-Adjusted Gamma Poisson Shrinker* (RGPS) [11].

To illustrate masking with a simple numerical example, consider the values displayed in Tables 2 and 3, which build on the example provided in Table 1 and Eq. (1). Tables 2 and 3 display values used for disproportionality analysis of 2 × 2 contingency tables capturing a hypothetical target AE and a hypothetical target product labeled 'A.' Table 2 introduces a product labeled 'B,' which serves as the 'offending' product that masks the true relationship between the target product 'A' and the target AE. To simplify our example, we assume that products 'A' and 'B' are not co-reported with other products and stress that what is being counted are the number of reports mentioning products/AEs and not co-occurrences. Table 2 shows that most of the reports (80/93) mentioning the target AE are associated with product 'B,' which leads to masking. Applying the RRR (Eq. 1) yields a masked RRR = $(393 \times 3)/(93 \times 13) = 0.98$, indicating that there is no statistical association. However, removing the reports that mention product 'B' yields the counts displayed in Table 3, and an unmasked RRR = $(233 \times 3)/(13 \times 13) = 4.14$ that indicates a strong statistical association between the target AE and target product 'A.'

Conditions that make signal detection especially vulnerable to masking effects include smaller safety databases such as VAERS that may lack diversity, relationships involving rare events, and relationships involving newer products. As such, the novelty of COVID-19 vaccines, coupled with ongoing vaccination programs, and the relatively early stages of COVID-19 vaccine surveillance make signal detection especially susceptible to masking.

The aim of this study is to investigate the problem of masking in relation to signal detection of COVID-19 vaccines and to assess its impact, extent, and root causes. To this end, we evaluate the evolution of signals corresponding to seven distinct AEs with various degrees of evidence linking them to the vaccines, and which demonstrate relatively strong masking effects. Five of these seven AEs are part of a list of AEs deemed to be of special interest for COVID-19 vaccine surveillance by the CDC and the FDA [12, 13]. The remaining two AEs, herpes zoster and tinnitus, are yet to be fully recognized but have accumulated thousands of reports in VAERS and are supported by published studies and case reports. We supplement this temporal investigation of seven AEs with a wider evaluation of masking at the database level. In addition, we center the evaluation on the messenger RNA (mRNA) vaccines from Pfizer-BioNTech (BNT162b2) and Moderna (mRNA-1273), which account for the vast majority of COVID-19 vaccine reports in VAERS.

Table 2 Contingency table used to compute disproportionality statistics with the inclusion of reports containing product 'B' that masks the association of product 'A' with the target AE

	Reports with target AE	Reports without target AE	
Reports with target product A	3	10	13
Reports with product B	80	80	160
Reports without product A or B	10	210	220
	93	300	393

AE adverse event

Table 3 Contingency table used to compute disproportionality statistics with the exclusion of reports containing product 'B' that would mask the association of product 'A' with the target AE

	Reports with target AE	Reports without target AE	
Reports with target product A	3	10	13
Reports with product B (excluded)			
Reports without product A or B	10	210	220
	13	220	233

AE adverse event

2 Materials and Methods

2.1 Data

The investigation was performed using all VAERS reports available at the time of writing this article (1990 to October 1, 2021). These data represent a total of 1,599,958 reports, including 39 weeks of COVID-19 vaccine reports, which are publicly released on a semi-monthly (every 2 weeks) cadence from January 1, 2021 to October 1, 2021. Of those, 778,681 reports include the COVID-19 vaccine from three manufacturers: Pfizer-BioNTech (53%), Moderna (39%), and Janssen (8%). The investigation was based on AEs in VAERS coded at the MedDRA Preferred Term (PT) level and products at the 'manufacturer' level, e.g., 'COVID19_PFIZER/BIONTECH.'

2.2 Adverse Events of Interest

The seven AEs investigated in this study and their associated MedDRA PTs are listed below. The MedDRA PTs associated with each of the seven AEs were used to identify VAERS reports mentioning a given AE.

1. *Bell's palsy* (PT = 'Facial paralysis' or 'Bell's palsy')
2. *Myocarditis* (PT = 'Myocarditis')
3. *Pericarditis* (PT = 'Pericarditis')
4. *Appendicitis* (PT = 'Appendicitis' or 'Appendicitis perforated' or 'Complicated appendicitis')
5. *Pulmonary embolism* (PT = 'Pulmonary embolism')
6. *Herpes zoster* (PT = 'Herpes zoster')
7. *Tinnitus* (PT = 'Tinnitus')

These AEs were selected for our investigation because they demonstrated strong masking effects and are supported by other sources. They were identified using an approach to screen and rank masked associations, which is described in Sect. 2.5 below. As noted in the Introduction, five of these AEs are partially recognized and are part of a list of AEs deemed to be of special interest for COVID-19 vaccine surveillance by the CDC and the FDA [12, 13]. The last two AEs (herpes zoster and tinnitus) were discovered through this investigation but are yet to be fully characterized like the other five AEs. Nonetheless, they are accompanied by strong statistical as well as published support, which is why they are included. Although we discovered other associations that exhibit masking effects, they did not appear strong or serious enough for inclusion in our evaluation, such as injection site pain.

2.3 Signal Detection Methodologies

We evaluated disproportionality statistics produced by four signal detection methodologies. A summary and short description of these methodologies as well as the statistics they compute is provided in Table 4, and further described in the following. Three of these methodologies—Multi-item Gamma Poisson Shrinker (MGPS) [14], Bayesian Confidence Propagation Neural Network (BCPNN) [15], and proportional reporting ratio (PRR) [16]—are well-established and are currently deployed by various organizations worldwide for routine safety surveillance. However, because these three methodologies are based on 2×2 disproportionality analysis, they are unable to, and were not designed to, control masking and certain confounding effects. We use these three methodologies as our baseline to investigate and verify masking effects. The fourth methodology, RGPS [11], is a signal detection methodology based on logistic regression that is designed to produce disproportionality statistics with adjusted background rates that can control masking and more extensive confounding effects. It operates by fitting separate Bayesian logistic regression models to each target AE and by automatically selecting predictors to be included in each regression model. The automatically selected predictors are products (vaccines in this case) that are statistically associated (based on unadjusted disproportionality statistics) with the target event and are represented as indicator variables. In addition, stratification categories are grouped by target AE rates and are represented as multiple regression intercepts. To address masking, RGPS adjusts a given target disproportionality statistic by adjusting its value for the presence of other products that also have large unadjusted disproportionalities (the regression predictors). This adjustment of the target disproportionality can be either positive or negative. When a non-target product with a large disproportionality never shows up in the same report as the target product, then the adjusted background AE rate will be lower and the target AE rate will be higher, in which case the association has been unmasked. Conversely, if that high-disproportionality non-target product is often co-prescribed with the target product, then the AE rate of the two products will be confounded and the adjusted targeted event rate for the two products will each be shrunk to express the uncertainty of which is the true causal factor when all three items, the two products and the target event, occur in the same report.

Additional details on the RGPS methodology are provided in the Supporting Information (SI1) (see the electronic supplementary material), and complete details of the RGPS methodology in Ref. [11].

The stratification categories used for RGPS, MGPS, and BCPNN were age and gender. Stratification by 'report year' was not applied because the vast majority of COVID-19 VAERS reports represent a single year of reporting (2021).

Table 4 Signal detection methodologies and disproportionality statistics used to investigate signals of coronavirus disease 2019 (COVID-19) vaccine adverse events

	Method name	Description	Signal score computed
2 × 2 Disproportionality analysis	Multi-item Gamma Poisson Shrinker (**MGPS**)	Bayesian approach designed to guard against false positives due to multiple comparisons. Computes an adjusted value of the observed-to-expected reporting ratio corrected for temporal trends and confounding by age and sex. Bayesian prior parameters are estimated using Empirical Bayes	**EBGM** (Empirical Bayes Geometric Mean): a centrality measure of the posterior distribution of the true observed-to-expected in the population **EB05/EB95**: lower/upper 5th percentile of the posterior EBGM observed-to-expected distribution
	Proportional reporting ratio (**PRR**)	Method to compute a measure akin to relative risk to quantify the strength of association between a product and event. In its canonical version it does not correct for temporal trends and confounding by age and sex	**PRR**: point estimate (mean) of the relative risk reporting ratio distribution
	Bayesian Confidence Propagation Neural Network (**BCPNN**)	Originally inspired by neural networks, is a Bayesian approach for computing the observed-to-expected reporting ratio corrected for temporal trends and confounding by age and sex. Uses pre-specified Bayesian prior parameters. In practice, produces signal statistics close to those of MGPS	**IC** (Information Component): posterior mean of the log observed-to-expected ratio **IC025/IC975**: lower/upper bounds of the IC 95% confidence interval
Regression-based	Regression-Adjusted Gamma Poisson Shrinker (**RGPS**)	Use of Bayesian logistic regression to guard against masking effects and false signals due to confounding by concomitant products. Computes an adjusted value of the observed-to-expected reporting ratio that is corrected for temporal trends and confounding by age and sex	**ERAM** (Empirical-Bayes Regression-Adjusted Arithmetic Mean): posterior mean of the observed-to-expected distribution **ER05/ER95**: lower/upper 5th percentile of the posterior ERAM observed-to-expected distribution

We applied the canonical version of PRR, which does not require stratification. For RGPS and MGPS, we generated both the point estimates, labeled Empirical-Bayes Regression-adjusted Arithmetic Mean (ERAM) and Empirical Bayes Geometric Mean (EBGM), respectively, and their associated credible intervals labeled ER05–ER95 and EB05–EB95, respectively. Unless specified otherwise, signal scores are represented by the point estimates. The generation of signal scores for the four methodologies considered in this study and analysis thereof was done using Oracle Empirica Signal 9.1 [17].

2.4 Capturing the Evolution of Signals

The evolution of signal scores for each AE was captured by a time series of signal statistics. The time series runs from a period at which initial reports for an AE were available to the latest batch of reports available at the time of writing this article. Each time point corresponds to a semi-monthly public release of VAERS reports, starting from week 3 (W3) January 22, 2021 and ending in week 39 (W39) October 1, 2021, for a total of 19 time points. The signal statistics computed for each time point include the signal score point estimate and its credible interval, e.g., *ER05-ERAM-ER95* for RGPS and *EB05-EBGM-EB95* for MGPS. These were computed based on all data available in VAERS and not only the COVID 19 reports or data within the range of dates underlying the time series.

2.5 Analysis and Evaluation

The comparison of signal detection methodologies for the time series centers on the RGPS and MGPS methodologies. These were chosen as representatives of the two classes of methodologies described in the 'Introduction' and above. That is, MGPS as a representative of the class of methodologies based on 2×2 disproportionality analysis that are unable to address masking, and RGPS as a representative of the more advanced class of methodologies based on regression that can address masking. The information component (IC) statistic [15] computed by the BCPNN methodology produces signal scores that are almost identical to those produced by MGPS and therefore redundant in many parts of our evaluation. The PRR signal statistic in its canonical application does not include smoothing or signal score adjustments for small counts as do the other methodologies and, therefore, does not protect against false alarms as well as the other methodologies. For this reason, a direct comparison against PRR (in its canonical form) would not have allowed us to isolate and explain sources of undetected signals. Nonetheless, both PRR and the IC statistic are used to confirm masking effects using the approach discussed in the following and presented in the 'Results' section.

Table 5 defines several concepts and conditions that we use to evaluate signals and to describe our findings in the 'Results' section. These include the concept of a signaling threshold, criteria to decide if a signal is detected or not (signal present/absent), a condition we use to decide if the difference between signal scores produced by different methodologies is statistically significant, a condition we use to screen candidate associations for masking, and the calculation we use to quantify the size of a masking effect.

Having generated the time series of signal scores for each AE of interest, we investigate and attempt to validate masking sources based on the following:

(1) We select two time periods: an earlier point in the evolution of signals when masking starts to take effect, and the end period (W39). Doing so allows us to examine the origin of the masking sources and whether the sources change over time. The earlier time point corresponds to the earliest point in the time series (for both the Pfizer-BioNTech and Moderna vaccines) for which the RGPS and MGPS signals scores were significantly different, and RGPS's signal score exceeded the signaling threshold as defined above.
(2) For each time point, we evaluate the predictors that are automatically selected by RGPS to be included in the regression model for the target AE. Based on the regression coefficients, we then identify the strongest predictors (vaccines) as potential sources of masking.
(3) As mentioned in the 'Introduction,' once masking sources have been identified, the conventional approach to control masking is to remove all reports containing the maskers, and re-compute signal scores. We use this conventional approach to confirm our findings. That is, we remove reports containing the potential maskers (vaccines) identified by RGPS and re-compute signal scores for the signaling methodologies based on 2×2 disproportionality analysis (MGPS, PRR, BCPNN). Substantial increases in these signal scores as well as their convergence toward the original RGPS signal score is a strong indication that the sources of masking have been correctly identified and a likely explanation for undetected or delayed statistical signals.

3 Results

Figures 1 and 2 and Table 6 depict our findings for each of the seven AEs investigated in this study. The figures display the evolution of signal scores for each AE captured

Table 5 Concepts and conditions used to evaluate signals

Concept	Definition
Signaling threshold	A cutoff value for a given signal score (association statistic) that is used to decide if a signal is present or absent. This investigation uses the value 1.0 (or 0.0 on the log scale), which for association statistics derived from ratios corresponds to the boundary of no statistical association
Signal present	For a given AE and signal score, a signal is present (i.e., detected) if a positive statistical association for the AE is identified. This occurs when the signal score for the AE (or its lower interval limit) exceeds the *signaling threshold.* This investigation requires that the lower limit of the signal score's credible interval exceeds the *signaling threshold*, e.g., ER05 > 1.0 for RGPS and EB05 > 1.0 for MGPS
Signal absent	For a given AE and signal core, a signal is absent (not detected) if the signal score's credible interval contains or falls below the signaling threshold, e.g., ER05 < 1.0 for RGPS and EB05 < 1.0 for MGPS
Statistically significant signal score difference	For a given association, the difference between two signal scores computed by two different methodologies is *statistically significant* if their credible intervals do not overlap. Likewise, we say that there is no difference in signal scores if their credible intervals overlap, e.g., ER05 < EB95 < ER95
Candidate association for masking	Candidate associations for masking are identified as those whose signal statistics satisfy the following condition: ER05 > EB95 and ER05 > 1 and EB05 ≤ 1 That is, an association where RGPS and MGPS disagree by producing signal scores that are statistically significant (non-overlapping credible intervals, ER05 > EB95) with RGPS's interval above the signaling threshold (ER05 > 1) and that of MGPS below or including the threshold (EB05 ≤ 1)
Masking effect size	The masking effect size is defined by the ratio of RGPS's and MGPS's signal scores, i.e., $\frac{\text{ERAM}}{\text{EBGM}} - 1$ In this investigation, the masking effect size will be averaged across the time series to produce a summary statistic and represented as a percentage

AE adverse event, *EBGM* Empirical Bayes Geometric Mean, *ERAM* Empirical-Bayes Regression-Adjusted Arithmetic Mean, *MGPS* Multi-item Gamma Poisson Shrinker, *RGPS* Regression-Adjusted Gamma Poisson Shrinker

as a time series of signal scores, whereas Table 6 provides signal scores for each AE averaged across the time series. As described in the 'Materials and Methods' (Sect. 2.4), the time series ranges from W3 to W39 of COVID-19 reports, for a total of 19 time points in 2-week intervals corresponding to the semi-monthly public release of VAERS reports.

Rows in the figures correspond to AEs, and columns to vaccines (Pfizer/BioNTech vs Moderna). Figure 1 covers the AEs Bell's palsy, myocarditis, pericarditis, and appendicitis, whereas Fig. 2 covers the AEs pulmonary embolism, herpes zoster, and tinnitus. Each figure displays a time series of signal scores for the RGPS and MGPS methodologies. Each point corresponds to the signal score point estimate and its credible interval (shaded region), i.e., *ER05-ERAM-ER95* for RGPS and *EB05-EBGM-EB95* for MGPS. Table 6 summarizes and supplements the figures by providing average signal scores for RGPS and MGPS (ERAM and EBGM, respectively) across each time series, as well as the average masking effect size defined in Sect. 2.5/Table 5. Finally, supporting information (SI2) (see the electronic supplementary material) provides signal statistics for all combinations of AE/vaccine/signaling methodology, including signal statistics for the PRR and BCPNN methodologies.

The figures clearly show several trends:

(1) The time series curves of signal scores produced by RGPS are always above those of MGPS, i.e., the RGPS signal scores are always larger than those of MGPS. This is not an expected pattern and is indicative of masking effects for the AEs of interest. This also suggests that the RGPS methodology would have been able to detect signals missed by MGPS or identify signals at an earlier time point than MGPS. According to Table 6, the average masking effect size ranges from around 40% for Bell's palsy to around 230% for herpes zoster, and the average signal score corrected for masking (RGPS) exceeds the signaling threshold.

(2) For most AEs, RGPS and MGPS initially agree on their signal scores (statistically insignificant differences) and then diverge in their signal scores. The divergence is likely due to the influence of masking effects, the evolution of VAERS data, and possibly changes in reporting practices.

(3) For several AEs, the time series exhibits an acute increase in signal score values at certain time points. These acute increases are likely explained or coincide with external events, such as the availability of a vaccine to certain age groups and the influence of publications.

(4) For certain AEs at certain time points, the signal scores fall below the signaling threshold. This indicates that at those time points statistical signals would have been undetected and that statistical signaling may be time sensitive.

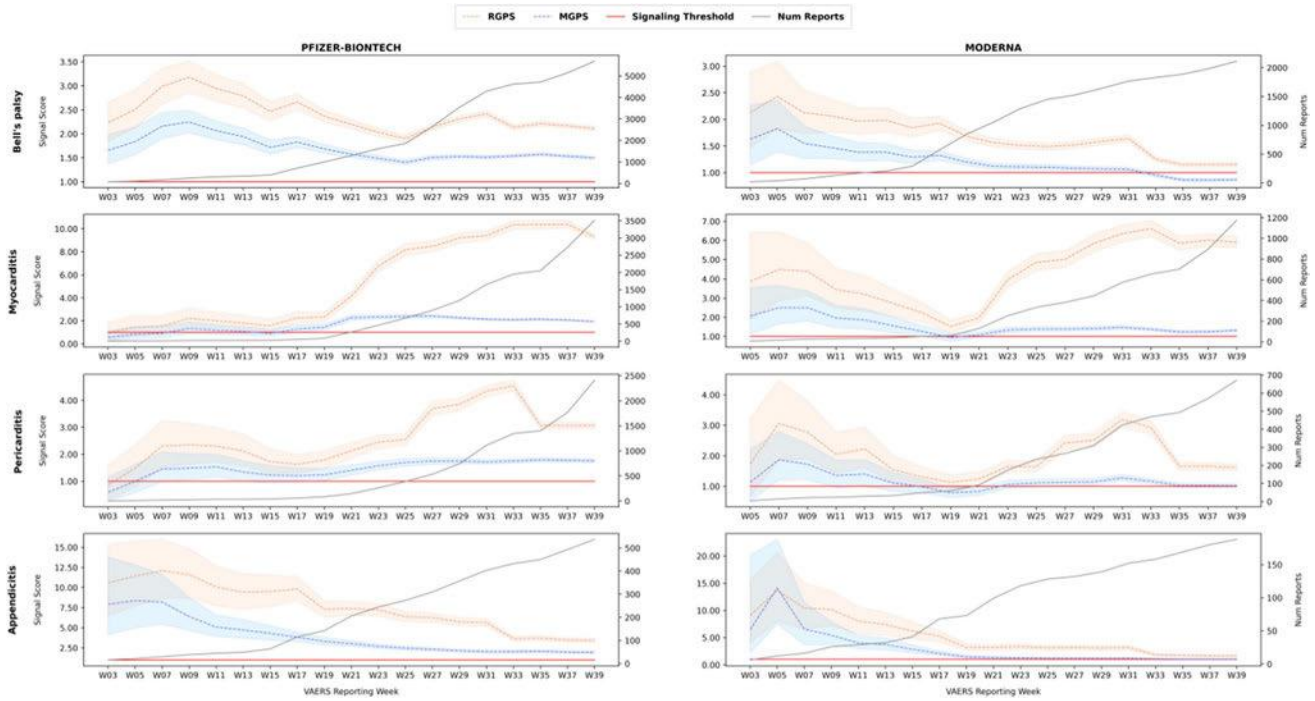


Fig. 1 The evolution of signal scores for Bell's palsy, myocarditis, pericarditis, and appendicitis. *MGPS* Multi-item Gamma Poisson Shrinker, *RGPS* Regression-Adjusted Gamma Poisson Shrinker, *W* week

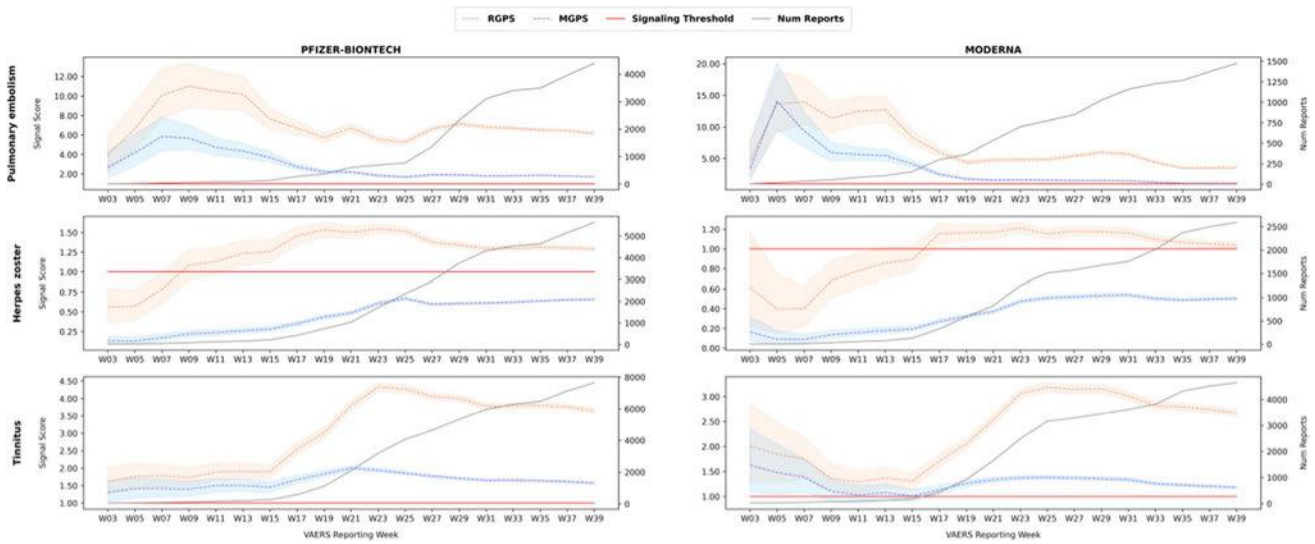


Fig. 2 The evolution of signal scores for pulmonary embolism, herpes zoster, and tinnitus. *MGPS* Multi-item Gamma Poisson Shrinker, *RGPS* Regression-Adjusted Gamma Poisson Shrinker, *W* week

(5) As more data accumulates, signal scores expectedly stabilize. Larger fluctuations are seen for RGPS, indicating that it is sensitive to masking and confounding effects and that the data may still be evolving.

The following describes our findings for each AE of interest.

3.1 Bell's Palsy

Bell's palsy is a form of acute facial paralysis with a weakening and a drooping appearance of the facial muscles usually on just one side of the face. In most cases, the paralysis resolves spontaneously within several weeks. Bell's palsy is due to swelling of the facial nerve, and type I interferons

Table 6 Average signal score and average masking effect for Bell's palsy, myocarditis, pericarditis, appendicitis, pulmonary embolism, herpes zoster, and tinnitus

Adverse event	Pfizer-BioNTech			Moderna		
	RGPS	MGPS	Masking effect size	RGPS	MGPS	Masking effect size
Bell's palsy	2.41	1.70	42%	1.69	1.22	39%
Myocarditis	5.40	1.66	190%	4.35	1.55	196%
Pericarditis	2.60	1.47	72%	2.02	1.17	71%
Appendicitis	7.61	3.94	110%	5.22	3.05	107%
Pulmonary embolism	7.18	2.88	178%	7.05	3.48	167%
Herpes zoster	1.23	0.44	229%	0.96	0.34	232%
Tinnitus	3.02	1.63	82%	2.31	1.27	80%

The average signal score for an AE is based on the individual signal scores underlying its time series displayed in Figs. 1 and 2. Average masking effect size is defined in Sect. 2.5 (not to be confused by the ratio of average signal scores for RGPS and MGPS in the table)

AE adverse event, *MGPS* Multi-item Gamma Poisson Shrinker, *RGPS* Regression-Adjusted Gamma Poisson Shrinker

have been proposed as the potential mechanism [18]. Incidents of Bell's palsy were reported in clinical trials for both the Pfizer-BioNTech and Moderna vaccines, and it has also been documented with the influenza vaccine [19, 20]. The FDA currently recommends its surveillance with larger populations globally. In addition, there have been multiple case reports of Bell's palsy associated with the mRNA vaccines [19, 21–23], and several studies that investigated the association [24–26].

As of W39, there are 7795 reports of Bell's palsy for the mRNA vaccines (5684 Pfizer-BioNTech, 2111 Moderna). The time series in Figure 1 shows that the signal scores produced by each methodology differ by a small amount, with RGPS and MGPS diverging (non-overlapping credible intervals) around W7–9. The figure also shows that a mild masking effect is present (40% averaged across the time series). Regardless of masking, all methods agree early on that the reported co-occurrence of the mRNA vaccines with Bell's palsy is unlikely due to chance (signal scores exceeding the signaling threshold). However, towards the end period (W33) the MGPS signal scores fall below the signaling threshold for the Moderna vaccine.

3.2 Myocarditis and Pericarditis

Myocarditis and pericarditis refer to inflammation of the heart muscle and outermost layer of the heart, respectively. Myocarditis and pericarditis are both thought to be caused by viral infections, and symptoms include chest pain, shortness of breath, and irregular heartbeat appearing within several days after the second dose of the mRNA vaccines. Several case reports of myocarditis and pericarditis developing rapidly after the first and second doses of the mRNA vaccines have been published [27–31], as well as several retrospective studies [13, 32–35] identifying it as a rare complication of the vaccines. One study in mice suggests that inadvertent intravenous injection of COVID-19 mRNA vaccines may induce myopericarditis [36].

The risk of myocarditis following vaccination has been observed to be highest among young males. The CDC has recognized the association with the COVID-19 mRNA vaccines [2], and both myocarditis and pericarditis now appear on the product labels (warning section) of the vaccines [37, 38].

As of W39, there are 4690 reports of myocarditis for the mRNA vaccines (3515 Pfizer-BioNTech, 1175 Moderna) and 3079 reports of pericarditis for the mRNA vaccines (2408 Pfizer-BioNTech, 671 Moderna) in the VAERS system. Relative to the total number of cases for these AEs, 87% of myocarditis cases and 83% of pericarditis cases are associated with the mRNA COVID-19 vaccines.

The changing age distribution of COVID-19 vaccine recipients can be observed in the progression of the time series. Figure 1 shows that both the RGPS and MGPS signal scores for myocarditis were initially not indicative of a safety signal, but around W19–21 (week ending May 30, 2021), as the COVID-19 vaccines were made available in the US to people under 65 years, a substantial increase in both signal scores can be observed. At this point RGPS and MGPS start diverging, with MGPS remaining on point and RGPS showing a gradual increase from a signal score of 2.3 to above 9.0 (Pfizer-BioNTech) and 1.5 to above 5.0 (Moderna). Similar trends of signal score progression are observed for pericarditis, with a slight decrease in RGPS signal scores around W31–33 onwards.

The size of the masking effect for myocarditis is ranked second for the AEs of interest, with an average value around 190%. For pericarditis, the effect size is 70%. The sources of masking for myocarditis were evaluated based on the process described in Sect. 2.5. The two time periods examined were W19 and W39. RGPS automatically selected 20 (W19) and 39 (W39) vaccine predictors for the myocarditis regression model. The strongest predictors for both time points were a set of three smallpox vaccines (at the manufacturer level), which is consistent with published reports recognizing myocarditis as a rare AE of the smallpox vaccine [39–41].

Upon removal of all reports containing the smallpox vaccines on W19, the PRR, EBGM, and IC signal scores indeed reverted to larger signal scores close in magnitude to RGPS's original signal score. The PRR signal score for the Pfizer-BioNTech vaccine increased from 1.44 to 2.48 (72%), and for the Moderna vaccine, from 0.8 to 1.34 (67%). Similarly, the EBGM signal score for the Pfizer-BioNTech vaccine increased from 1.44 to 2.17 (51%), and from 0.94 to 1.42 (51%) for the Moderna vaccine. As more data accumulated in VAERS, the Pfizer-BioNTech and Moderna COVID-19 vaccines were also identified by RGPS as potential maskers. In this case, they masked each other for the myocarditis AE. On W39, the Pfizer-BioNTech vaccine was identified by RGPS as the strongest masker. Removing all reports containing the Pfizer-BioNTech vaccine led to a substantial increase in signal scores for the Moderna–myocarditis association. The PRR signal score increased from 1.2 to 4.98 (315%), and the EBGM score increased from 1.32 to 2.13 (61%). This demonstrates how the Pfizer-BioNTech vaccine is masking the Moderna vaccine, and how masking sources may evolve over time. In addition to the COVID-19 vaccines, the smallpox vaccines were still identified by RGPS as strong sources of masking on W39. Removing both smallpox and Pfizer-BioNTech vaccines led to the following additional increases for the Moderna association: PRR increased from 4.98 to 8.14 (63%) and EBGM increased from 2.13 to 2.4 (13%). Similarly, removing the smallpox and Moderna vaccines led to the following increases for the Pfizer-BioNTech-myocarditis association: PRR increased from 5.42 to 10.96 to 17.94 (230%) and EBGM increased from 1.94 to 2.02 to 2.12 (9%).

3.3 Appendicitis

Appendicitis is an inflammation of the appendix usually caused by an obstruction of the appendiceal lumen; however, the exact etiology of acute appendicitis is often unknown. Appendicitis is the most common cause of acute abdominal pain requiring surgery. If left untreated, acute appendicitis can result in serious complications, such as peritonitis or abscess formation [42, 43]. According to the Pfizer-BioNTech COVID-19 Vaccine Fact Sheet for Healthcare Providers, appendicitis was reported as a serious AE in a clinical trial for eight vaccine participants and four placebo participants (Pfizer-BioNTech COVID-19 vaccine = 10,841; placebo = 10,851), but not during post-authorization experience [37]. The Moderna COVID-19 Vaccine Fact Sheet for Healthcare Providers does not mention appendicitis as an AE in clinical trials or in post-authorization experience [38]. However, both the Pfizer-BioNTech and Moderna Fact Sheets for Healthcare Providers mention lymphadenopathy as a reported AE during clinical trials. Barda et al. demonstrated an elevated risk ratio for appendicitis (risk ratio 1.40; 95% confidence interval [CI] 1.02–2.01) with the Pfizer-BioNTech COVID-19 vaccine in a mass nationwide vaccination setting [44].

As of W39, there are 725 reports of appendicitis for the mRNA vaccines (537 Pfizer-BioNTech, 188 Moderna) in the VAERS system. As shown in Fig. 1, both MGPS and RGPS showed extremely large signal scores early on that attenuated over time but remained high for RGPS, with values above 3.7 for Pfizer-BioNTech and above 1.7 for Moderna. This early signaling by W3 appeared even when the number of reports was small (15 Pfizer-BioNTech, 6 Moderna). RGPS and MGPS started diverging around W11, likely due to masking. The figure shows a relatively large masking effect. Averaged across the time series, the size of the masking effect was high and around the value of 100% for both vaccines.

3.4 Pulmonary Embolism

Pulmonary embolism is a sudden blockage in a lung artery. It usually happens when a blood clot breaks loose and travels through the bloodstream to the lungs. Pulmonary embolism is a serious condition that can cause permanent damage to the lungs, low oxygen levels in the blood, and damage to other organs in the body from not getting enough oxygen. Pulmonary embolism can be life-threatening, especially if a clot is large, or if there are many clots [45].

Systematic reviews and meta-analyses showed high incidences of pulmonary embolism in COVID-19 patients [46, 47]. Barda et al. reported an elevated risk ratio for pulmonary embolism (risk ratio 12.14; 95% CI 6.89–29.20) for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-infected compared to uninfected persons [44].

Besides COVID-19 itself, it appears that COVID-19 vaccines increase the risk for pulmonary embolism; several authors reported the occurrence of pulmonary embolism, often in combination with vaccine-induced thrombotic thrombocytopenia (VITT), following COVID-19 vaccination, mainly for adenovirus-based COVID-19 vaccines [48–54]. Although no increased risk for pulmonary embolism was found by Klein et al. for mRNA vaccines [12] and by Barda et al. for Pfizer-BioNTech [44], some case reports

described the occurrence of pulmonary embolism following vaccination with Pfizer-BioNTech [55–57]. As of this writing, pulmonary embolism is not mentioned in the vaccine labels of the Pfizer-BioNTech and the Moderna COVID-19 vaccines.

As of W39, there are 5869 reports of pulmonary embolism for the mRNA vaccines (4394 Pfizer-BioNTech, 1475 Moderna) in the VAERS system. Figure 2 shows that both MGPS and RGPS exceed the signal threshold for pulmonary embolism already in W3 for both vaccines. In the following weeks, starting on W9, RGPS departs from MGPS and stays on a value level about threefold that of MGPS. Averaged across the time series, the size of the masking effect was high and around the value of 170% for both vaccines. The MGPS time series for Moderna decreases to below the signaling threshold in W39, whereas RGPS remains well above the threshold. For Pfizer-BioNTech, MGPS and RGPS remain above the signaling threshold, with RGPS at about three times the value of MGPS.

3.5 Herpes Zoster

Herpes zoster (shingles) is a painful rash that develops on one side of the face or body. The rash consists of blisters that typically clear within 2–4 weeks [58]. Multiple reports of patients who developed herpes zoster shortly after COVID-19 vaccination have been recently published [59–64], as well as observational studies and systematic reviews [44, 65–67], which suggest a potential link with the mRNA COVID-19 vaccines. Possible mechanisms that explain the pathogenic link are related to the stimulation of innate immunity through toll-like receptors 3, 7 by mRNA-based vaccines [65].

As of W39, there are 8228 reports of herpes zoster for the mRNA vaccines (5637 Pfizer-BioNTech, 2591 Moderna). Figure 2 shows a substantial difference between RGPS and MGPS, with MGPS indicating that there is no statistical association between herpes zoster and the vaccines (signal scores below the signaling threshold), versus RGPS indicating the contrary (signal scores exceeding the signaling threshold) from W13 (Pfizer-BioNTech) and W17 (Moderna) through the remaining time periods. Although the value of the RGPS signal score is not large relative to the other AEs, it indicates that the association is unlikely to be due to chance.

Interestingly, the size of the masking effect for herpes zoster was the largest among the AEs of interest. Averaged across the time series, the size of the masking effect was 230% for both mRNA vaccines. The sources of masking were evaluated and validated based on the process described in Sect. 2.5. The two time periods examined were W17 and W39. RGPS automatically selected 67 (W17) and 44 (W39) vaccine predictors for the herpes zoster regression model. The strongest predictors were the varicella (chickenpox) and the VARZOS (a combination varicella and zoster) vaccines, for a total of six vaccine predictors at the manufacturer level. Although the risk is low, there are documented cases and studies of herpes zoster following varicella and VARZOS vaccination [68–70]. Upon removal of all reports containing the varicella and VARZOS vaccines, we found that the PRR, EBGM, and IC signal scores indeed reverted to larger signal scores close in magnitude to RGPS's original signal score. For example, the PRR signal score for the Pfizer-BioNTech vaccine increased from 0.37 to 1.47 (297%) on W17 and from 0.76 to 2.3 (202%) on W39. Similarly, the EBGM signal score increased from 0.35 to 1.47 (320%) on W17 and from 0.66 to 1.48 (124%) on W39. In addition, we found that these masking sources (i.e., the varicella and VARZOS vaccines) did not change over time and remained consistent at both time periods that were evaluated.

3.6 Tinnitus

Tinnitus is described as the sensation of hearing ringing, hissing, or other noises in one or both ears that is not caused by an external sound. Tinnitus can be intermittent or continuous and can vary in pitch and intensity. Prolonged exposure to loud sounds and a variety of other conditions can lead to tinnitus; however, the mechanism responsible for tinnitus is unclear.

Tinnitus has been linked to other vaccines such as hepatitis, rabies, measles, and H1N1 vaccines [71]. In COVID-19 vaccine trials prior to the release of the Pfizer-BioNTech and Moderna vaccines, no mention was made of the onset of tinnitus or worsening tinnitus for either vaccine. As early as March 2021, in a report from the United Kingdom Medicines and Healthcare products Regulatory Agency (MHRA), 196 tinnitus cases among 33,207 vaccinated persons were recorded for the Pfizer-BioNTech vaccine [72], and since then, several case reports linking tinnitus to the mRNA vaccines as well as to the Janssen and AstraZeneca vaccines have been published [72–75]. In addition, due to an apparently increased number of individuals experiencing tinnitus during the pandemic period, the connection between the vaccines and tinnitus received special attention in various media outlets and professional associations dedicated to tinnitus [76, 77]. As of this writing, tinnitus is not mentioned in the vaccine labels. As mentioned in the 'Introduction,' tinnitus is not contained in the set of AEs of interest recognized by various health organizations. As of W39, there are 12,296 reports of tinnitus for the mRNA vaccines (7649 Pfizer-BioNTech, 4647 Moderna) in the VAERS system. Interestingly, the number of reports for tinnitus is larger by a substantial amount than for any of the other AEs covered in this article. Figure 2 shows that both MGPS and RGPS exceed the signal threshold early on for both vaccines and remain

above the signaling threshold through the remaining time periods (excluding a brief crossing for MGPS and Moderna on W9–15). RGPS and MGPS start diverging on W15–17, with RGPS rapidly increasing to signal score values twice as large in a short amount of time. This appears correlated with the increase in the number of reports available throughout the period and likely the dynamics of masking effects.

Averaged across the time series, the size of the masking effect was high and around the value of 80% for both vaccines. Based on the process described in Sect. 2.5, we evaluated the sources of masking for tinnitus. The two time periods examined were W17 and W39. RGPS automatically selected 21 (W17) and 25 (W39) vaccine predictors for the tinnitus regression model. For W17, the strongest predictors and potential maskers identified by RGPS were the HPV4 (papilloma virus) vaccine and the Janssen and Pfizer-BioNTech COVID-19 vaccines. Hence, on W17, two COVID-19 vaccines were already masking other associations; Janssen masking the Pfizer-BioNTech and Moderna COVID-19 vaccines, and the Janssen and Pfizer-BioNTech vaccines masking the Moderna vaccine. Removing all reports containing these three vaccines (HPV4, Janssen, and Pfizer-BioNTech) resulted in expected signal score increases for the Moderna-tinnitus association, with PRR increasing from 1.79 to 2.5 (40%) and EBGM increasing from 1.13 to 1.43 (27%). Expectedly, on W39, as more data accumulated in VAERS, the Pfizer-BioNTech and Moderna vaccines were identified by RGPS as the strongest maskers (masking each other) in addition to the Janssen vaccine. On W39, the HPV4 vaccine was no longer identified as a strong masker. Removing reports containing the Janssen and Pfizer-BioNTech vaccines led to the following signal score changes for the Moderna-tinnitus association: PRR increasing from 1.8 to 5.5 (205%) and EBGM increasing from 1.18 to 1.69 (43%). Similarly, removing reports containing the Janssen and Moderna vaccines led to the following signal score changes for the Pfizer-BioNTech-tinnitus association: PRR increasing from 2.75 to 6.67 (143%) and EBGM modestly increasing from 1.57 to 1.71 (9%). This demonstrates that the Pfizer-BioNTech and Moderna vaccines may mask each other to varying degrees, in this case, Pfizer-BioNTech having a larger effect on Moderna than vice versa.

3.7 Masking Statistics at the Database Level

Table 7 displays counts for the number of potentially masked associations in VAERS categorized by vaccine type. The conditions that define a potentially masked association are provided in the 'Materials and Methods' (Sect. 2.5; Table 5, candidate association for masking). The table shows that the likelihood of a masked association for the COVID-19 vaccines is 2.3%, which is roughly eight times larger than for non-COVID-19 vaccines (0.3%). This result clearly demonstrates the increased potential and susceptibility of VAERS COVID-19 vaccine surveillance to the problem of masking effects.

4 Discussion

The unprecedented dynamic and extent of reporting into VAERS for the novel class of COVID-19 vaccines may have created conditions that predispose commonly applied signal detection methodologies to the statistical issue known as masking. This in turn may limit our understanding of the risks associated with COVID-19 vaccines, as well as other vaccines and delay their identification.

Signal detection can be approached and accomplished in many ways. In this article, we consider a specific approach and application that is routinely applied by pharmacovigilance organizations, and whose purpose is to computationally explore large databases of reported AEs for statistical patterns that are indicative of new safety issues that warrant further attention. We term this application statistical signal detection and further distinguish two classes of methodologies, one based on 2×2 disproportionality analysis that is prone to masking, and a more advanced class of methods that can cope with masking. Methodologies currently deployed by pharmacovigilance organizations are to a large extent based on the former class of methods and, thus, prone to masking, a motivating reason for this investigation. To abbreviate our discussion, we will refer to this class of methods as the 'standard' methods.

To demonstrate such masking effects, trace their origins, and assess their impact, we center our investigation on seven AEs with various degrees of reported and statistical evidence that link them to the Pfizer-BioNTech and Moderna vaccines. Five of the AEs are largely recognized by various health authorities. The investigation enabled us to discover two potentially new AEs (herpes zoster and tinnitus), which are yet to be recognized by health authorities, but which have overwhelming statistical support in VAERS and are supported by published case reports and studies. These seven AEs were identified and selected for this investigation

Table 7 VAERS counts of masked associations

	Number associations	Number masked associations	% masked associations
All vaccines	265,987	1330	0.50%
Non-COVID-19 vaccines	241,016	753	0.31%
COVID-19 vaccines	24,971	577	2.31%
Pfizer-BioNTech/Moderna	18,588	458	2.46%

COVID-19 coronavirus disease 2019

based on criteria to screen and rank masked associations described in the 'Methods.' We do not extrapolate and claim that masking is often prevalent because it was identified for these seven AEs; neither do we suggest that masking is limited to just these AEs. Rather, we argue that masking is an issue that is important and addressable, and an issue that can be impactful in situations such as COVID-19 vaccine safety surveillance and other emergency use authorization products.

In the investigation, we traced the evolution of signals related to the seven AEs during the course of the initial year of COVID-19 vaccination and the accompanying availability of COVID-19 vaccine AE reports made public in VAERS. This temporal evaluation led to several findings. We surmise that these findings are important not only for the COVID-19 vaccines currently approved and investigated in this article, but are also important for any new COVID-19 vaccines that might be approved in the future and, likewise, should also apply to any new vaccine (or drug) approved for use in the future.

The results show that statistical signals for AEs related to COVID-19, and possibly other vaccines, may go undetected or be delayed due to masking when generated by standard methodologies. The results also suggest that properly identifying and addressing the masking effect exposes strong statistical associations that would otherwise be deemed uninteresting. For example, the tinnitus and herpes zoster signals may have been overlooked partly due to the low signal scores produced for them by standard methodologies. Similarly, signals for the other five AEs may have been delayed by the same standard methodologies. As mentioned, safety surveillance and signal detection are not limited to statistical approaches, and fortunately, these other five AEs had already been well characterized by the FDA, CDC, and other sources.

We found that although the masking effect is rare relative to the entire set of possible associations between vaccines and AEs (representing 0.5% of the total number of unique associations), it is roughly eight times more likely to occur with COVID-19 vaccines than with other vaccines. As mentioned, this may be explained by the unique dynamic and extent of reporting into VAERS for the class of COVID-19 vaccines. Furthermore, the volume of reporting for COVID-19 vaccines is likely to influence future statistical associations with other new vaccines. This suggests that masking may become more frequent and should be carefully considered.

The results also demonstrate that masking is not a static effect but rather a dynamically changing and evolving effect in terms of its origins, direction, and strength. Naturally, this is due to the evolving nature of data. For example, we found that in earlier time periods, non-COVID-19 vaccines could mask signals associated with COVID-19 vaccines, whereas in later time periods, as more COVID-19 reports accumulate, the Pfizer-BioNTech and Moderna vaccines can mask each other and likely other vaccines. This suggests that the assessment of masking should be done on a continuum rather than be a point-in-time exercise and, more generally, that statistical signal detection is time sensitive. Relatedly, it appears that the VAERS data for COVID-19 vaccine surveillance are still evolving and susceptible to external influences, such as vaccination policies, publication influences, reporting practices, and updates to the MedDRA terminology. This in turn could contribute to signal score fluctuations, resulting in time-dependent signaling uncertainty.

Masking effects have been traditionally addressed by removing cases containing the 'offending' product, by using stratification, or by employing regression techniques. However, each of these approaches requires to some extent identifying masking sources prior to signaling, which may limit the utility of signal detection in scenarios where masking is present and where the goal is unconstrained hypothesis generation. This investigation was made possible by using a methodology that automatically identifies and adjusts masking effects. Its ability to correctly identify maskers was verified for three of the seven AEs we investigated (e.g., the smallpox vaccines masking COVID-19 for myocarditis) by using the traditional approach to address masking. That is, by re-applying standard signaling methodologies on data that excludes the maskers.

At a higher level, the results suggest that different signaling approaches may lead to drastically different results—a conclusion that is especially disconcerting in the context of COVID-19 surveillance. Unfortunately, in the absence of an ultimate benchmark, the question of which methodology to rely on is still in debate. Nonetheless, the findings highlight the utility of a more advanced class of signal detection methodologies based on regression. Given present-day computational power and recognized analytic approaches such as regression, there are few reasons to avoid the utilization of these approaches, at the very least to address acknowledged problems such as masking.

The mRNA Pfizer-BioNTech and Moderna vaccines have been demonstrated to be highly effective in preventing infection and severe illness from COVID-19. They also appear to have acceptable safety profiles, suggesting that the benefits of COVID-19 vaccination outweigh the potential risk of AEs. Consequently, AEs such as those highlighted in this article, which are also rare as far as we know, cannot be used to argue against vaccination. Moreover, statistical signal detection is inherently an exploratory hypothesis-generating process. Therefore, associations flagged by signaling approaches do not imply causal relationships and always warrant further scrutiny, including those named in this article. Notwithstanding, the strength of statistical signal detection (as an unconstrained hypothesis-generating process) lies

in being fast and performed in near 'real time.' Analyses can be easily 'tailored' to a specific age group or gender, time frame, and product type. The method also has the advantage of casting a much wider net for AE reporting from millions or hundreds of millions of people and may identify rare AEs not seen in clinical trials. These advantages are critical in the 'real time' and the 'real world' environment of COVID-19 vaccine surveillance.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40264-022-01186-z.

Declarations

Funding No funding was received for the conduct of this research or the preparation of this article.

Disclaimer The findings and conclusions expressed in this report are those of the authors and do not necessarily represent the views of the US FDA or the federal government.

Conflict of interest The authors declare no conflict of interest. Rave Harpaz, William DuMouchel, Robbert Van Manen, Alexander Nip, Steve Bright, and Magnus Lerch are employed by Oracle Health Sciences, which provides the signal detection and management software used to conduct the research for this article.

Ethics approval Ethics approval was not needed for this study.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material The data used for this article are based on the public release of the Vaccine Adverse Event Reporting System (VAERS). The electronic supplementary material (ESM2) provides time-indexed disproportionality statistics for all the methodologies and AEs investigated in this study.

Code availability Not applicable.

Author contributions All authors contributed to the writing, editing, and review of the manuscript and approved the final version for submission. All authors contributed to the analysis and interpretation of data. Conceptualization: RH and WD. Original draft: RH. Software and data resources: RM and AN.

References

1. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet. 2000;356(9237):1255–9.
2. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. Drug Saf. 2002;25(6):381–92.
3. Almenoff J, et al. Perspectives on the use of data mining in pharmacovigilance. Drug Saf. 2005;28(11):981–1007.
4. Pharmacovigilance Signal-Management. https://www.ema.europa.eu/en/human-regulatory/post-authorisation/pharmacovigilance/signal-management.
5. Harpaz R, et al. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther. 2012;91(6):1010–21.
6. Wisniewski AFZ, et al. Good signal detection practices: evidence from IMI PROTECT. Drug Saf. 2016;39(6):469–90.
7. CIOMS Working Group VIII. Practical aspects of signal detection in pharmacovigilance. CIOMS; 2010.
8. Guidance for industry: good pharmacovigilance practices and pharmacoepidemiologic assessment U.S. Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER) March 2005 https://www.fda.gov/media/71546/download.
9. Maignen F, et al. Assessing the extent and impact of the masking effect of disproportionality analyses on two spontaneous reporting systems databases. Pharmacoepidemiol Drug Saf. 2014;23(2):195–207.
10. Juba KM, van Manen RP, Fellows SE. A review of the food and drug administration adverse event reporting system for tramadol-related hypoglycemia. Ann Pharmacother. 2020;54(3):247–53.
11. DuMouchel W, Harpaz R. Regression-adjusted GPS algorithm (RGPS), in Oracle White Paper. https://docs.oracle.com/health-sciences/empirica-signal-811/ESIUG/Regression-Adjusted_GPS_Algorithm.pdf. 2012.
12. Klein NP, Lewis N, Goddard K, et al. Surveillance for adverse events after COVID-19 mRNA vaccination. JAMA. 2021;326(14):1390–99. https://doi.org/10.1001/jama.2021.15072
13. Li X, et al. Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study. BMJ. 2021;373:n1435.
14. Dumouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am Stat. 1999;53(3):177–90.
15. Bate A, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol. 1998;54(4):315–21.
16. Evans SJW, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf. 2001;10(6):483–6.
17. Oracle Empirica Signal. https://docs.oracle.com/en/industries/health-sciences/empirica-signal/9.1/index.html. 2021.
18. Iftikhar H, et al. Bell's Palsy after 24 hours of mRNA-1273 SARS-CoV-2 vaccine. Cureus. 2021;13(6):e15935–e15935.
19. Burrows A, et al. Sequential contralateral facial nerve palsies following COVID-19 vaccination first and second doses. BMJ Case Rep. 2021;14(7): e243829.
20. Ozonoff A, Nanishi E, Levy O. Bell's palsy and SARS-CoV-2 vaccines. Lancet Infect Dis. 2021;21(4):450–2.
21. Cirillo N, Doan R. The association between COVID-19 vaccination and Bell's palsy. Lancet Infect Dis. 2021. https://doi.org/10.1016/S1473-3099(21)00467-9.

22. Wan EYF et al. Bell's palsy following vaccination with mRNA (BNT162b2) and inactivated (CoronaVac) SARS-CoV-2 vaccines: a case series and nested case-control study. Lancet Infect Dis. 2022;22(1):64–72.
23. Repajic M, et al. Bell's Palsy after second dose of Pfizer COVID-19 vaccination in a patient with history of recurrent Bell's palsy. Brain Behav Immunity Health. 2021;13:100217–100217.
24. Tamaki A, et al. Incidence of bell Palsy in patients with COVID-19. JAMA Otolaryngol Head Neck Surg. 2021;147(8):767–8.
25. Renoud L, et al. Association of facial paralysis with mRNA COVID-19 vaccines: a disproportionality analysis using the World Health Organization Pharmacovigilance Database. JAMA Intern Med. 2021;181(9):1243–5.
26. Shemer A, et al. Association of COVID-19 vaccination and facial nerve palsy: a case–control study. JAMA Otolaryngol Head Neck Surg. 2021;147(8):739–43.
27. Kaul R, et al. Myocarditis following COVID-19 vaccination. IJC Heart Vasc. 2021;36: 100872.
28. Koizumi T, et al. Myocarditis after COVID-19 mRNA vaccines. QJM Int J Med. 2021;114:741–3.
29. Alania-Torres E, et al. Case report: probable myocarditis after Covid-19 mRNA vaccine in a patient with arrhythmogenic left ventricular cardiomyopathy. Front Cardiovasc Med. 2021.
30. Ashaari S, Sohaib HA, Bolger K. A case report: symptomatic pericarditis post-COVID-19 vaccination. Eur Heart J Case Rep. 2021;5(10):ytab375.
31. Myocarditis and pericarditis after receipt of mRNA COVID-19 vaccines. https://content.govdelivery.com/attachments/MDMBP/2021/05/28/file_attachments/1839782/Clinician%20Letter_Myocarditis_5.28.21.pdf. 2021.
32. Diaz GA, et al. Myocarditis and pericarditis after vaccination for COVID-19. JAMA. 2021;326(12):1210–2.
33. Aye YN, et al. Acute myocardial infarction and myocarditis following COVID-19 vaccination. QJM Int J Med. 2021.
34. Witberg G, et al. Myocarditis after Covid-19 vaccination in a Large Health Care Organization. New Engl J Med. 2021;385:2132–9.
35. Chelala L, et al. Cardiac MRI findings of myocarditis after COVID-19 mRNA vaccination in adolescents. Am J Roentgenol. 2021;218:651–7.
36. Li C, et al. Intravenous injection of coronavirus disease 2019 (COVID-19) mRNA vaccine can induce acute myopericarditis in mouse model. Clin Infect Dis. 2021.
37. Pfizer-BioNTech Product Label. https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=908ecbe7-2f1b-42dd-94bf-f917ec3c5af8.
38. Moderna Product Label. https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=e0651c7a-2fe2-459d-a766-0d59e919f058.
39. Halsell JS, et al. Myopericarditis following smallpox vaccination among Vaccinia-Naive US Military Personnel. JAMA. 2003;289(24):3283–9.
40. Keinath K, et al. Myocarditis secondary to smallpox vaccination. BMJ Case Rep. 2018. https://doi.org/10.1136/bcr-2017-223523.
41. Cardiac adverse events following smallpox vaccination—United States. 2003. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5212a2.htm.
42. Jameson JL, et al. Harrison's principles of internal medicine, 20th edn. McGraw Hill Medical; 2018.
43. Medlineplus.gov. Accessed 27 Sept 2021.
44. Barda N, et al. Safety of the BNT162b2 mRNA Covid-19 vaccine in a Nationwide setting. N Engl J Med. 2021;385(12):1078–90.
45. MedlinePlus. https://medlineplus.gov/pulmonaryembolism.html.
46. Roncon L, et al. Incidence of acute pulmonary embolism in COVID-19 patients: systematic review and meta-analysis. Eur J Intern Med. 2020;82:29–37.
47. Suh YJ, et al. Pulmonary embolism and deep vein thrombosis in COVID-19: a systematic review and meta-analysis. Radiology. 2021;298(2):E70–80.
48. Islam A, et al. An update on COVID-19 vaccine induced thrombotic thrombocytopenia syndrome and some management recommendations. Molecules. 2021;26(16):5004.
49. Asmat H, et al. A rare case of COVID-19 vaccine-induced thrombotic thrombocytopaenia (VITT) involving the veno-splanchnic and pulmonary arterial circulation, from a UK district general hospital. BMJ Case Rep. 2021;14(9): e244223.
50. Muster V, et al. Pulmonary embolism and thrombocytopenia following ChAdOx1 vaccination. Lancet. 2021;397(10287):1842.
51. Greinacher A, et al. Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. N Engl J Med. 2021;384(22):2092–101.
52. Bersinger S, et al. Using nonheparin anticoagulant to treat a near-fatal case with multiple venous thrombotic lesions during ChAdOx1 nCoV-19 vaccination-related vaccine-induced immune thrombotic thrombocytopenia. Crit Care Med. 2021;49(9):e870–3.
53. Malik B, et al. Pulmonary embolism, transient ischaemic attack and thrombocytopenia after the Johnson & Johnson COVID-19 vaccine. BMJ Case Rep. 2021;14(7): e243975.
54. Clark RT, et al. Early outcomes of bivalirudin therapy for thrombotic thrombocytopenia and cerebral venous sinus thrombosis after Ad26.COV2.S vaccination. Ann Emerg Med. 2021;78(4):511–4.
55. Al-Maqbali JS, et al. A 59-year-old woman with extensive deep vein thrombosis and pulmonary thromboembolism 7 days following a first dose of the Pfizer-BioNTech BNT162b2 mRNA COVID-19 vaccine. Am J Case Rep. 2021;22: e932946.
56. Esba LCA, Al Jeraisy M. Reported adverse effects following COVID-19 vaccination at a tertiary care hospital, focus on cerebral venous sinus thrombosis (CVST). Expert Rev Vaccines. 2021;20(8):1037–42.
57. Nune A, et al. Multisystem inflammatory syndrome in an adult following the SARS-CoV-2 vaccine (MIS-V). BMJ Case Rep. 2021;14(7): e243888.
58. Herpes Zoster. https://www.cdc.gov/shingles/about/symptoms.html.
59. Mohta A, et al. Recurrent herpes zoster after COVID-19 vaccination in patients with chronic urticaria being treated with cyclosporine—a report of 3 cases. J Cosmet Dermatol.
60. Chiu H-H, Wei K-C, Chen A, Wang W-H. Herpes zoster following COVID-19 vaccine: a report of three cases. QJM Int J Med. 2021;114(7):531–32.
61. Eid E, et al. Herpes zoster emergence following mRNA COVID-19 vaccine. J Med Virol. 2021;93(9):5231–2.
62. Arora P, et al. Herpes zoster after inactivated COVID-19 vaccine: a cutaneous adverse effect of the vaccine. J Cosmet Dermatol. n/a(n/a).
63. Thimmanagari K, et al. Ipsilateral zoster ophthalmicus post COVID-19 vaccine in healthy young adults. Cureus. 2021;13(7): e16725.
64. Dasgupta N. Countering hesitancy and misinformation on side effects to complete the course of COVID vaccination. J Patient Exp. 2021;8:23743735211067310.
65. Furer V, et al. Herpes zoster following BNT162b2 mRNA Covid-19 vaccination in patients with autoimmune inflammatory rheumatic diseases: a case series. Rheumatology. 2021;60:SI90–5.
66. Wan EYF, et al. Herpes zoster related hospitalization after inactivated (CoronaVac) and mRNA (BNT162b2) SARS-CoV-2 vaccination: a self-controlled case series and nested case–control study. Lancet Region Health West Pac. 2022;21:100393.
67. Katsikas Triantafyllidis K, et al. Varicella zoster virus reactivation following COVID-19 vaccination: a systematic review of case reports. Vaccines. 2021;9(9):1013.

68. Moodley A, et al. Severe herpes zoster following varicella vaccination in immunocompetent young children. J Child Neurol. 2019;34(4):184–8.
69. Rueda JM, et al. Disseminated herpes zoster after varicella vaccination in a healthy boy. Actas Dermo-Sifiliográficas (Engl Edn). 2017;108(6):587–8.
70. Liang MG, et al. Herpes zoster after varicella immunization. J Am Acad Dermatol. 1998;38(5):761–3.
71. Okhovat S, et al. Sudden onset unilateral sensorineural hearing loss after rabies vaccination. BMJ Case Rep. 2015;2015:bcr2015211977.
72. Parrino D, Frosolini A, Gallo C, De Siati RD, Spinato G, de Filippis C. Tinnitus following COVID-19 vaccination: report of three cases. Int J Audiol. 2022;61(6):526–29.
73. Tseng P-T, et al. The reversible tinnitus and cochleopathy followed first-dose AstraZeneca COVID-19 vaccination. QJM Int J Med. 2021;114:663–4.
74. Buntz B. Is JJS COVID-19 vaccine linked to tinnitus?. https://www.drugdiscoverytrends.com/is-jjs-covid-19-vaccine-linked-to-tinnitus/. 2021.
75. Wichova H, Miller ME, Derebery MJ. Otologic manifestations after COVID-19 vaccination: the House Ear Clinic experience. Otol Neurotol. 2021;42(9):e1213–8.
76. American Tinnitus Association. https://www.ata.org/tinnitus-and-coronavirus.
77. British Tinnitus Association. https://www.tinnitus.org.uk/coronavirus-vaccines-and-tinnitus.

From: "Alimchandani, Meghna" <[redacted]>
To: "Menschik, David" <[redacted]>, "Zinderman, Craig E" <[redacted]>
Subject: RE: Coverage through 8/18
Date: Fri, 5 Aug 2022 14:22:56 +0000
Importance: Normal

Yay sounds good! (I was also hoping we would stop doing that at some point!)

From: Menschik, David <[redacted]>
Sent: Friday, August 5, 2022 10:22 AM
To: Alimchandani, Meghna <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Coverage through 8/18

Thanks!

No – we are no longer routinely sending COVID data mining to CDC (per Narayan after he discussed with Tom S.)

Thanks again,

David

From: Alimchandani, Meghna <[redacted]>
Sent: Friday, August 05, 2022 10:14 AM
To: Menschik, David <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Coverage through 8/18

Thanks, David! Hope you and Craig have good vacations next week.
Quick question…we are still sending data mining for COVID vaccines also on a weekly basis to CDC…is that right?

Thanks to Melvyn for covering urgent BO needs!

Best,
Meghna

From: Menschik, David <[redacted]>
Sent: Friday, August 5, 2022 8:23 AM
To: Alimchandani, Meghna <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: Coverage through 8/18

Hi Meghna and Craig,

Thanks very much for covering me while I'm out from 10:30 am today through 8/18 (MA: covering through Friday 8/12; CZ covering through 8/18).

1. Meetings:
    a. AE Weekly status meeting on 8/12 (MA)
    b. GDIT Composite Report WG 8/16 (CZ)
2. COVID vaccine dose data - Post (drag and drop) spreadsheets from CDC email to Team folder
3. ?Jynneos dose data – John/Tom indicated plan to have this available similar to COVID vaccine dose data and indicated they would similarly share with us (pending)
4. Jynneos data mining – I shared results from Empirica summary table (signals tab) once per attached email. I check this weekly and would plan to share with Tom/John if new PT(s) appear in the table.
5. Melvyn has a list of tasks for which he should be independent with exception of TARS queries/reports and will be working with Chris Jason on this. Melvyn knows that #1 priority is 'customer service' including timely responses to requests for help with BO queries. Please don't hesitate to ask him for help with any BO query.

Thanks again,

David

From: "Zinderman, Craig E" <[redacted]>
To: "Menschik, David" <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts
Date: Mon, 6 Sep 2021 18:03:16 +0000
Importance: Normal
Inline-Images: image001.png; image002.png

David:
Thanks much for laying out that there are disadvantages to not stratifying by year; that's very helpful. Would it also make the case to Steve that the current methodology has been examined and is the best approach? Happy to discuss further Wednesday.

Thanks,
Craig

From: Menschik, David <[redacted]>
Sent: Monday, September 06, 2021 9:03 AM
To: Zinderman, Craig E <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

Thanks and agree. I've given this a lot of thought. Please see my attached draft proposed response to Ana. Would welcome any suggested edits and advice on who to include on 'to' and 'cc' lines though would like to discuss first by phone with you (feel free to call my cell) before proceeding farther...

Thanks,
David

From: Zinderman, Craig E <[redacted]>
Sent: Sunday, September 05, 2021 3:16 PM
To: Menschik, David <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

Thanks David. Happy to discuss options; I would lean towards having some scientific rationale/data to support an approach.

Thanks,
Craig.

From: Menschik, David <[redacted]>
Sent: Saturday, September 04, 2021 7:18 AM
To: Zinderman, Craig E <[redacted]>
Subject: FW: CBER VAERS Signal Management Liaisons/Contacts

FYI and before potential response, let's discuss any thoughts you or I may have by phone when we're back next week.

From: Szarfman, Ana <[redacted]>
Sent: Friday, September 03, 2021 5:50 PM
To: Hendrix, Brian * <[redacted]>; Sydnor, James * <[redacted]>; Menschik, David <[redacted]>
Cc: Lebow, William * <[redacted]>; Baer, Bethany <[redacted]>; Siegel, Jeffrey <[redacted]>; Stockbridge, Norman L <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

Hi Brian,

Thanks so much for the wonderful job you are all doing.

Hi David,

I noticed that you are Board Certified in Clinical Informatics. Congratulations!

Regarding the question I posted to Brian:

Why I am concerned about stratifying the VAERS data by year?

Most of the VAERS reports for 2021 are for the COVID-19 vaccines.
By stratifying by year you are only using one year of data.
For a sound data mining analysis, more than half of the reports need to be for other vaccines.
Usually the control group would have 5 or 10 as many cases as the products of interest.
If you only want to compare the 3 different COVID-19 vaccines with each other, this would OK, but the 3 vaccines could be doing the same bad thing, and you would not know it.
By stratifying by year, the background would be composed by the covid-19 vaccines.
Astra Zeneca in their demo at the Accelerator meeting, presented data not stratified by year, for this same reason.

Using the RGPS data mining algorithm vs MGPS

RGPS is much, much better at unmasking signals than MGPS.
It automatically identifies and corrects for confounders.
This is an important function to have, given the pandemic situation.

I hope we continue helping each other.

Let me know if you need further information.

--Ana

Ana Szarfman, MD, PhD, FAMIA,
[redacted] (office)
(personal cell phone and WhatsApp)

FDA U.S. FOOD & DRUG ADMINISTRATION

From: Hendrix, Brian * <[redacted]>
Sent: Friday, September 3, 2021 3:24 PM
To: Szarfman, Ana <[redacted]>; Sydnor, James * <[redacted]>
Cc: Menschik, David <[redacted]>; Lebow, William * <[redacted]>; Baer, Bethany <B[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

Hi Ana,
Thank you for bringing this up.

Currently all of the VAERS DM runs are being stratified by year.

Given the large proportion of covid-19 events, we will need to look at this going forward.

I've copied David and Bethany here to make them aware as well.

-Brian

From: Szarfman, Ana <[redacted]>
Sent: Friday, September 3, 2021 2:16 PM
To: Hendrix, Brian * <[redacted]>; Sydnor, James * <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

Therefore the background will only be for covid-19 vaccines, instead of for other vaccines. Therefore, masking covid-19 vaccine signals that are common with these vaccines, but not common across other types of vaccines.

From: Szarfman, Ana
Sent: Friday, September 3, 2021 2:07 PM
To: Hendrix, Brian * <[redacted]>; Sydnor, James * <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

For VAERS. Over 95% of the reports in 2021 are for COVID-19 vaccines.

From: Hendrix, Brian * <[redacted]>
Sent: Friday, September 3, 2021 2:06 PM
To: Szarfman, Ana [redacted]>; Sydnor, James * [redacted]
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

For VAERS or across Signal in general?

From: Szarfman, Ana <[redacted]>
Sent: Friday, September 3, 2021 2:06 PM
To: Hendrix, Brian * <[redacted]>; Sydnor, James * [redacted]
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

Thanks Brian and Casey,

Are any of the DM runs being generated NOT BEING stratified by year?

From: Hendrix, Brian * <[redacted]>
Sent: Friday, September 3, 2021 2:02 PM
To: Sydnor, James * <[redacted]>

Cc: Szarfman, Ana <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

Hi Ana,

Can you please let me know which runs you have concerns about? I can provide details of the run structures as needed.

Thank you,
Brian

From: Sydnor, James * <[redacted]>
Sent: Friday, September 3, 2021 1:58 PM
To: Hendrix, Brian * <[redacted]>
Cc: Szarfman, Ana <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

Brian,

Ana has a concern regarding the new CBER VAERS Data Mining and Signal Management runs regarding the possibility that they may be stratifying by Year. I know that there were a number of discussions about the criterial for the runs, so I'm fairly certain that we do not stratify by Year because of the issues with the background that would occur for the most recent months. Please confirm briefly if you can so that Ana can approach David and Bethany with a little bit of background. Thank you!

Best regards,

Casey Sydnor (contractor)
Commonwealth Informatics, Inc.
Empirica Signal Support Team
Ph: [redacted]
-
-
U.S. FOOD & DRUG ADMINISTRATION

From: Sydnor, James *
Sent: Friday, September 3, 2021 1:54 PM
To: Szarfman, Ana <[redacted]>
Cc: Hendrix, Brian * <[redacted]>
Subject: CBER VAERS Signal Management Liaisons/Contacts

Ana,

As we discussed on the phone, you will need to reach out to David Menschik and Bethany Baer (contact info below) in order to discuss your interest in the new CBER VAERS Signal Management runs. Please let Brian and me know if/how we can help after you have discussed with David and Bethany. You can copy us on the correspondence with them if you like, so that we can remain in the loop to know how the conversation is resolved. Best of luck and we wish you a wonderful long weekend!

David Menschik, MD, MPH
Associate Director for Surveillance Informatics
Division of Epidemiology/Office of Biostatistics and Epidemiology

Center for Biologics Evaluation and Research/FDA
[redacted]

Bethany Baer
Physician
Division of Epidemiology/Office of Biostatistics and Epidemiology
Center for Biologics Evaluation and Research/FDA
[redacted]

Best regards,

Casey Sydnor (contractor)
Commonwealth Informatics, Inc.
Empirica Signal Support Team

Office Of Translational Sciences
FDA/CDER/OTS
Ph: [redacted]

From: "Zinderman, Craig E" <[redacted]>
To: "Menschik, David" <[redacted]>
Subject: FW: Write-up
Date: Tue, 14 Sep 2021 19:12:13 +0000
Importance: Normal

David:
I put together a few bullets below, but I'm not wed to them if you have better thoughts; just drafted these to hopefully be of help. You probably have some thoughts to add in any case. Also, I'm not sure if the last sentence is accurate.

Thanks,
Craig.

-CBER appreciates Dr. Szarfman's interest in the data mining procedures for COVID vaccines, as well as her ongoing and past contributions to data mining currently in use at CDER and CBER.
-Over the past several months, Dr. Szarfman has reached out to CBER staff members about data mining using VAERS data for COVID vaccines, and has also raised this topic in various inter-Center data mining related meetings. Dr. Szarfman has been interested in implementing new data mining methods for COVID vaccines, and also in revising the current data mining procedures, currently in place at both CDER and CBER, by removing Year stratifications.
- While there is good reason to remove this stratification (the volume of COVID reports substantially outweighs the volume of non-COVID vaccine reports, potentially masking some of the results), there are also drawbacks, from an epidemiological standpoint, to removing this stratification.
-Importantly, varying data mining methods, such that sensitivity is lower and the resulting number of alerts requiring investigation is higher, doesn't necessarily yield important safety findings. Rather, in CBER's experience, the robust monitoring process already in place for vaccine, and especially for COVID vaccines, is more likely to uncover new safety findings. Increasing the volume of data mining alerts often results in more work for reviewers with little new findings.
-To this end, in May, after discussion with Dr. Anderson, CBER OBE staff asked Dr. Szarfman to stop conducting data mining analyses for COVID vaccines, and to solely use data and products from her own Center for her data mining methods exploration and other work.
-Dr. Szarfman recently reached out to the data mining contractor, and again to CBER staff, to again suggest changes to the procedures in place at CBER. CBER staff explained that, due to the drawbacks of removing the Year stratifications, we prefer to not remove this stratification at this time. Dr. Szarfman is suggesting support for a different data mining calculation, not in place in CDER or CBER, and plans to publish a paper using this methodology on CBER's COVID vaccine data.

From: Nair, Narayan <[redacted]>
Sent: Tuesday, September 14, 2021 1:00 PM
To: Zinderman, Craig E <[redacted]>; Menschik, David <[redacted]>
Subject: FW: Write-up

Dear Craig and David,
Can you provide a few sentences on this issue for me to use in my response to Steve? Happy to discuss by phone if that is better.

Narayan

From: Anderson, Steven <[redacted]>
Sent: Tuesday, September 14, 2021 12:34 PM
To: Nair, Narayan <[redacted]>
Subject: Write-up

Dear Narayan,

Can you send us a one-paragraph write-up about Ana and the challenges it is posing for DE? I would like to share it with Dr. Marks to potentially share with CDER leadership.

Let me know if you wish to discuss.

Regards,

Steve

Steve Anderson, Ph.D., M.P.P.
Director
Office of Biostatistics and Epidemiology
Center for Biologics Evaluation and Research
U. S. Food & Drug Administration
[redacted]

Phone: [redacted]
email: [redacted]

From: "Su, John (CDC/DDID/NCEZID/DHQP)" <[redacted]>
To: "Menschik, David (FDA/CBER)" <[redacted]>
Subject: [EXTERNAL] RE: data mining limitations
Date: Wed, 22 Sep 2021 16:43:32 +0000
Importance: Normal

Hi David,

Signal detection with VAERS data has always been tricky business. I think we're always looking for any quantitative tools to help make sense of what we're seeing. That said, FDA has always made clear the limitations of data mining. Those of us who work with VAERS data frequently are mindful of those limitations; I just figured I share those limitations with folks who aren't as familiar with VAERS (e.g., CISA).

Appreciate the below language. Thanks!

- John

From: Menschik, David <[redacted]>
Sent: Wednesday, September 22, 2021 12:33 PM
To: Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: data mining limitations

Hi John,

In the mRNA vaccine review article that we're co-authors on, we recently expanded data mining limitations section as per attached work-in-progress draft (Hannah indicated acceptance of the language) and excerpt below for convenience:

> EB data mining has multiple limitations[22] including that an absence of a disproportionality alert does not rule out presence of a safety problem. Additionally, since most reports received during this surveillance period involved COVID-19 vaccines, disproportionately scores (which are adjusted by year to control for time-dependent, potentially confounding, exposure and outcome variables) can be muted by COVID-19 vaccine reports contributing substantially to the comparator group, particularly if both mRNA COVID-19 vaccines are associated with the same adverse event.

Thought it might be helpful to share this manuscript update with you, especially if folks on your end may be placing excess value on data mining alerts (EB05>2) or the absence of specific data mining alerts.

Best,
David

PS: If you'd like to discuss more, happy to do so by phone (better suited than email…)

Message

From: Anderson, Steven [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=D4C0C242FEBA45FA954F4F9B05EB3557-ANDERSONST]
Sent: 9/15/2021 2:40:18 AM
To: Nair, Narayan [redacted]
Subject: RE: Write-up

Dear Narayan,

Thank you for this – it is well done. I may modify this a bit by adding some more context since I am sharing with Dr. Marks who may share it with CDER and possibly agency leadership.

Regards,

Steve

Steve Anderson, Ph.D., M.P.P.
Director
Office of Biostatistics and Epidemiology
Center for Biologics Evaluation and Research
U. S. Food & Drug Administration
[redacted]

Phone: [redacted]
email: [redacted]

From: Nair, Narayan <[redacted]>
Sent: Tuesday, September 14, 2021 3:54 PM
To: Anderson, Steven <[redacted]>
Subject: RE: Write-up

Dear Steve,

Happy to discuss further if needed. Here is a brief paragraph –

We are very appreciative of Ana's extensive knowledge and expertise related to data mining. However, we have concerns about her communicating data mining findings using CBER VAERS data to CBER and non-CBER personnel. While we think these efforts are well intentioned, we would request she refrain from using her FDA email or communicating data mining findings using CBER VAERS data given she is a CDER employee.

Some additional info (not sure if this is helpful): When we began planning our approach to passive surveillance for the COVID-19 vaccines in the summer of 2020, our overarching strategy was to build on existing, established systems whenever possible. With regard to Data Mining, I felt that we should utilize the standardized, established system that has been in use for other vaccines. My concern was a novel approach that had not been validated would add another layer of uncertainty in the context of an EUA when rapid retrieval and interpretation of data would be imperative. We recognize as with all passive surveillance our current data mining process has limitations. In particular, we are well aware that if there is a class-effect (e.g., if both mRNA COVID-19 vaccines are associated with the same adverse event) it may be missed by data mining.

My thinking was to re-evaluate the data mining (as well as our other processes) once the data from active surveillance is available. In the best of circumstances data mining is only hypothesis generating and I thought it would be helpful once active surveillance has confirmed the hypothesis related to a certain safety signal(s) to re-evaluate our approach. This would be a suitable time to determine why it wasn't detected by data mining (if that is the case). I think this retrospective approach is more downstream than what Ana is proposing but to my mind would be preferable to shifting midstream.

From: Anderson, Steven <[redacted]>
Sent: Tuesday, September 14, 2021 12:34 PM
To: Nair, Narayan <[redacted]>
Subject: Write-up

Dear Narayan,

Can you send us a one-paragraph write-up about Ana and the challenges it is posing for DE? I would like to share it with Dr. Marks to potentially share with CDER leadership.

Let me know if you wish to discuss.

Regards,

Steve

Steve Anderson, Ph.D., M.P.P.
Director
Office of Biostatistics and Epidemiology
Center for Biologics Evaluation and Research
U. S. Food & Drug Administration
[redacted]

Phone: [redacted]
email: [redacted]

Appointment

From: Szarfman, Ana [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=856B5EF9696E4C6F973C8F10409D88EE-SZARFMAN]
Sent: 12/22/2020 7:37:36 PM
To: Forshee, Richard [redacted]; Niu, Manette [redacted]; Allende, Maria [redacted]; Ryan, Qin [redacted]; Weichold, Frank [redacted]; Ellenberg, Susan [redacted]; Janet Wittes [redacted]; Kluetz, Paul [redacted]; Concato, John [redacted]; Pennello, Gene [redacted]; Stockbridge, Norman L [redacted]; Temple, Robert [redacted]; bill.dumouchel ([redacted]; Pease-Fye, Meg [redacted]; Zhang, Rongmei [redacted]; Janet Wittes [redacted]; Nair, Narayan [redacted]; Lababidi, Samir [redacted]; Tomita, York [redacted]; Chuk, Meredith [redacted]; Kim, Tamy [redacted]; Walderhaug, Mark O [redacted]; cgrochester[redacted]; Dal Pan, Gerald [redacted]; antonioparedes14[redacted] Pucino, Frank [redacted]
CC: Blum, Michael [redacted]; Huang, Lei [redacted]; Yang, Ye [redacted]; Lin, Lisa [redacted]

Subject: Bill DuMouchel's plan for a software that can assess multiple studies and applications at once
Attachments: DuMouchel - Addressing the need for improving the comprehensiveness and transparency of our decision processes to address the COVID-19 pandemic (002).pptx
Location: Via webex

Start: 12/23/2020 5:00:00 PM
End: 12/23/2020 6:00:00 PM
Show Time As: Tentative

Importance: High

Required Attendees: Forshee, Richard; Niu, Manette; Allende, Maria; Qin Ryan ([redacted]); Weichold, Frank; Ellenberg, Susan; Janet Wittes; Kluetz, Paul; Concato, John; Pennello, Gene; Stockbridge, Norman L; Temple, Robert; bill.dumouchel ([redacted]; [redacted] Pease-Fye, Meg; Zhang, Rongmei ([redacted]; Janet Wittes; Nair, Narayan; Lababidi, Samir; Tomita, York; Chuk, Meredith; Kim, Tamy; Walderhaug, Mark O; cgrochester[redacted] Dal Pan, Gerald; antonioparedes14[redacted] Pucino, Frank
Optional Attendees: Blum, Michael; Huang, Lei; Yang, Ye; Lin, Lisa

Hi All,

Many thanks for all your hard work to keep us safe.

We took the liberty to schedule this meeting so close to Christmas because this topic will help improve our analytical capabilities to fight effectively this pandemic situation.

→Will this date and time work for you? Please RSVP.

We have planned this small meeting to gain support for the attached project developed of Bill DuMouchel to address many of the analytical problems we are facing during this dire pandemic. Bob Temple and Norman Stockbridge are aware of this potential project.

DuMouchel developed the safety data mining methodology being used to safety data mine AERS/FAERS and VAERS.

We have been encouraging Bill DuMouchel to implement a methodology that can assess multiple studies and applications at once (i.e., all vaccines for COVID-19, all anticoagulants) where the goal is to compare different treatments that don't all appear in any one study, which is the "Network Meta-Analysis" paradigm.

As Bill DuMouchel will explain, this approach will compare subgroups based on several covariates and in addition where there are multiple endpoints measured on each patient.

All this would usually require an analysis of the patient level data in order to fit the Bayesian shrinkage model.

An advantage of having one program that can do an analysis of patient level data from many studies and many applications at once is that if we can add or remove some data we can get a quick analysis without having to do a lot of intermediate analyses and redo a meta-analysis based on the new summary statistics.

--Ana

Ana Szarfman, MD, PhD, FAMIA,
Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), and
Fellow of the American Medical Informatics Association (2020)
Medical Officer, Safety Data Mining Developer and Medical Informatics Analyst,
Celebrating nearly a quarter of a century of successful implementation of safety data mining, interactive patient profiles, and other automated analytical tools.
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted]
[redacted](office)
[redacted](personal cell phone and WhatsApp)

U.S. FOOD & DRUG
ADMINISTRATION

Addressing the need for improving the comprehensiveness and transparency of our decision processes to address the COVID-19 pandemic

Introductory words to the presentation by Dr. William DuMouchel

by

Ana Szarfman, MD, PhD, FAMIA
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research
Food and Drug Administration

- The good news is that we now have multiple new potential types of vaccines and medical interventions.
- These interventions are unprecedented for:
  - The compressed time for their development and study
  - The large number of potential new interventions (20 +)
  - The enormous size of the global population impacted (hundreds of millions of people)
- The usual review process for new vaccines and molecular entities at the Agency is designed at assessing a much smaller impacted population

- Hearing the summary options of scientists that helped control in the past much smaller outbreaks would not be sufficient.
- Consistent efficacy and safety signals may remain hidden
  - in disjoined Clinical Trial applications and observational studies or by analyses that do not properly adjust for multiplicity and small counts across all the data being generated.
- We need an automated, interactive analytical process in place that can compare the data of all interventions being continuously collected in CTs and observational studies.
  - The automated reanalysis of the data and identification of consistent signals by specific interventions that can be exhaustively examined by experts will provide transparency to the decision making process.

- We have been encouraging Bill DuMouchel to implement a methodology that can assess multiple studies and applications at once (i.e., all vaccines for COVID-19, all anticoagulants) where the goal is to compare different treatments that don't all appear in any one study, which is the "Network Meta-Analysis" paradigm.
- As Bill DuMouchel explained, this approach will compare subgroups based on several covariates and in addition where there are multiple endpoints measured on each patient.
- All this would usually require an analysis of the patient level data in order to fit the Bayesian shrinkage model.

- An advantage of having one program that can do an analysis of patient level data from many studies and many applications at once is that if we can add or remove some data we can get a quick analysis without having to do a lot of intermediate analyses and redo a meta-analysis based on the new summary statistics.

ORACLE

Challenges Involving Models for Complicated Full-Data Meta-Analyses

Analyses of pooled studies of vaccines, biologics, or drug products from prospective or observational studies and involving multiple treatments, subgroups and endpoints

William DuMouchel, PhD
Chief Statistician
Oracle Health Sciences

Many New Vaccines and Proposed Treatments for Covid-19 Are Being Tested

Comparisons of Efficacy and Safety Are Complicated

- Multiple Efficacy and Multiple Safety Endpoints
- Within-Study Comparisons of Treatment Arms
- Comparing Subsets Depending on Covariate Values (age, sex, concomitant meds, etc.)
- Combining Studies (Including Prospective and Observational Studies)
  - Studies May Differ by Indication, Populations or Medical History
  - Bayesian Approaches May Give Higher Weight to Certain Study Designs

Network Meta-Analysis

https://training.cochrane.org/resource/key-concepts-network-meta-analysis-nma

Several Products Need to Be Compared

- Each Study Typically Only Includes Two or a Few Products
- But a Chain of Within-Study Comparisons May Connect Every Pair of Products
- Pooled Studies Can Create Efficient Indirect Comparisons Among Them All
- Allowance Must Be Made for Population Differences Between Studies

Both Observational and Prospective Studies May Be In the Pool

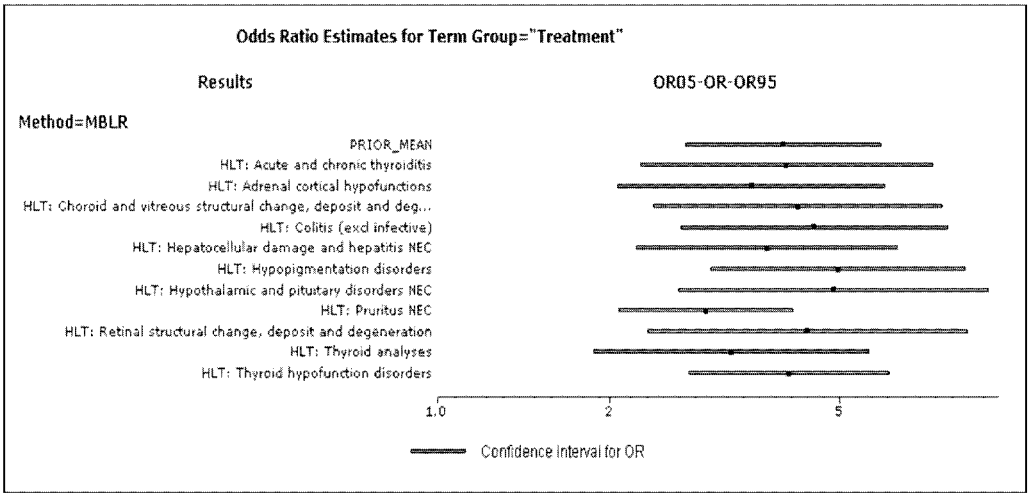
- Bayesian Approaches May Give Higher Weight to Certain Study Designs

Eight Studies of the ICPI Nivolumab versus other Active Comparators

How to compare the various treatment effects?

Study	ARM 1	Arm 2
37		INVESTIGATOR CHOICE
17		DOCETAXEL
25		EVEROLIMUS
57		DOCETAXEL
66	NIVOLUMAB	DACARBAZINE
26		INVESTIGATOR CHOICE
41		INVESTIGATOR CHOICE
27		CHEMO

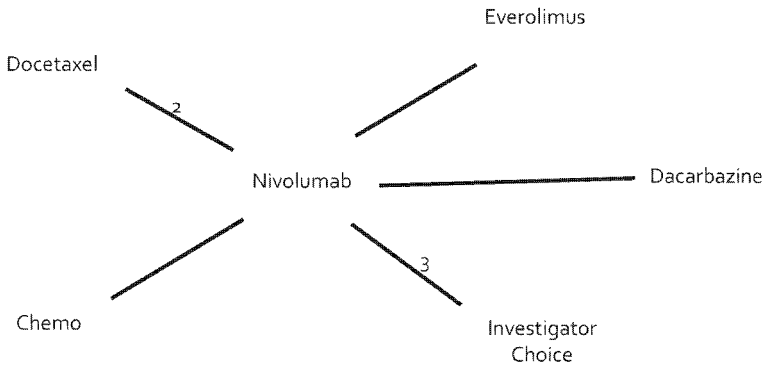
Comparison: Pool All 8 Studies Into 1 Analysis For 11 Safety HLTs



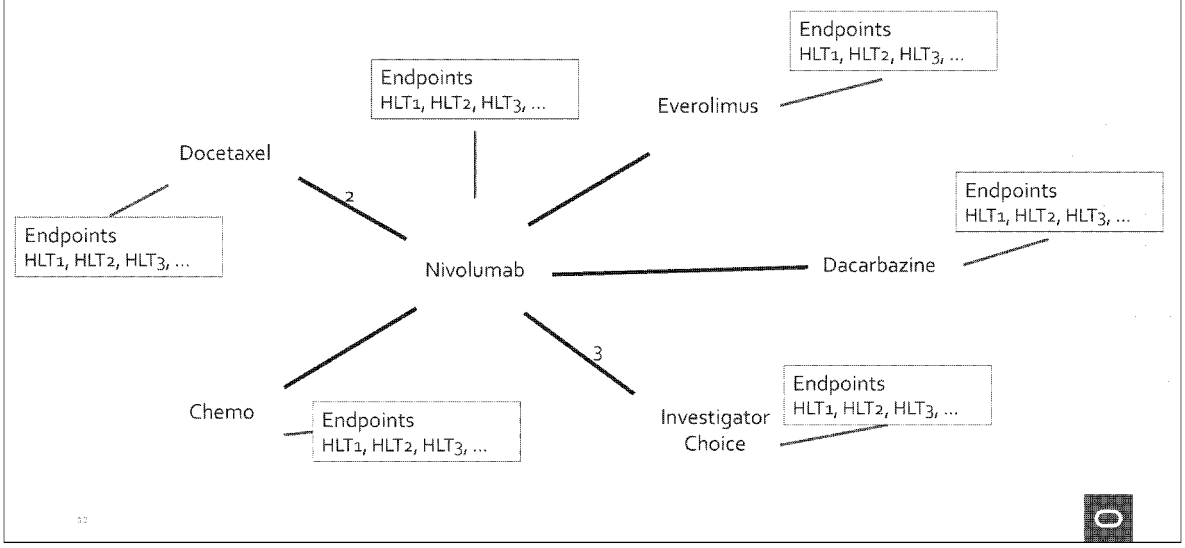
Multivariate Bayesian Logistic Regression (MBLR)
MBLR (also known as BLR) is similar to meta-analysis, but uses the whole data from each study. Adjusts for potentially biasing between-study differences and multiple comparisons. The results of a BLR run may help support any of the following conclusions:
An issue appears related to treatment in a screening analysis, but is not related to treatment when covariates are included in a BLR run, and these covariates show a strong relationship to issue outcome. This may indicate that a randomization error has occurred.

Representing a Pool of Studies as a Network

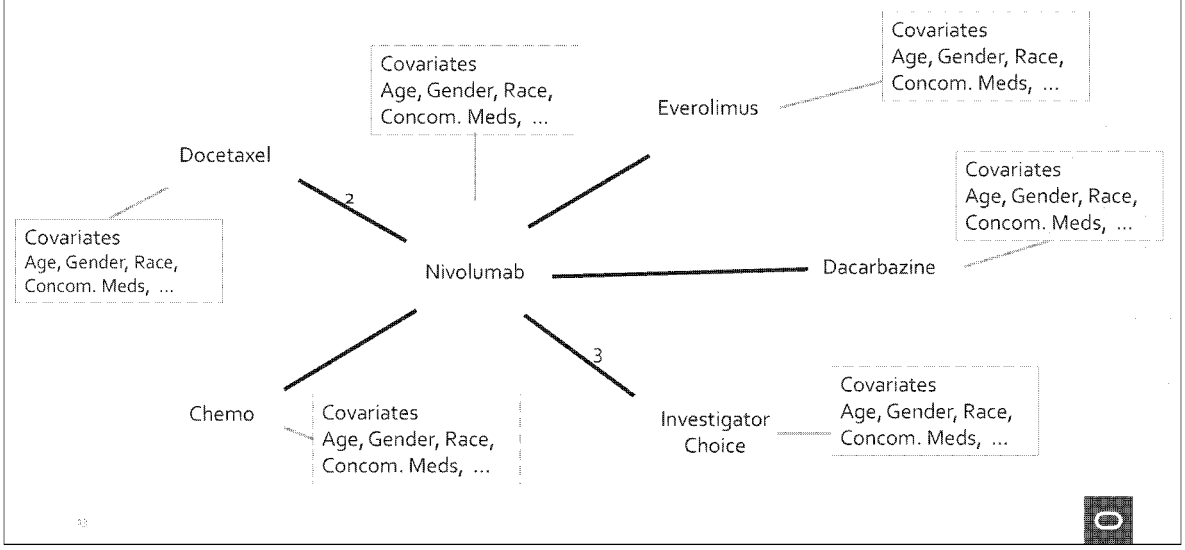
Lines Connecting Two Arms Represent within-Study Comparisons (numbers=study multiplicity)



Each Patient May Have Multiple Medically Similar Endpoint Measurements



Each Patient May Have Multiple Covariates Possibly Influencing Endpoints or Treatment Effects



Hierarchical Bayesian Models and Multiple Comparisons

Selecting a "Best" Treatment, Covariate-Subgroup or Endpoint Post-hoc while Accounting for Random Variation

Standard Approach: Require Small p-Values or Wide Confidence Intervals

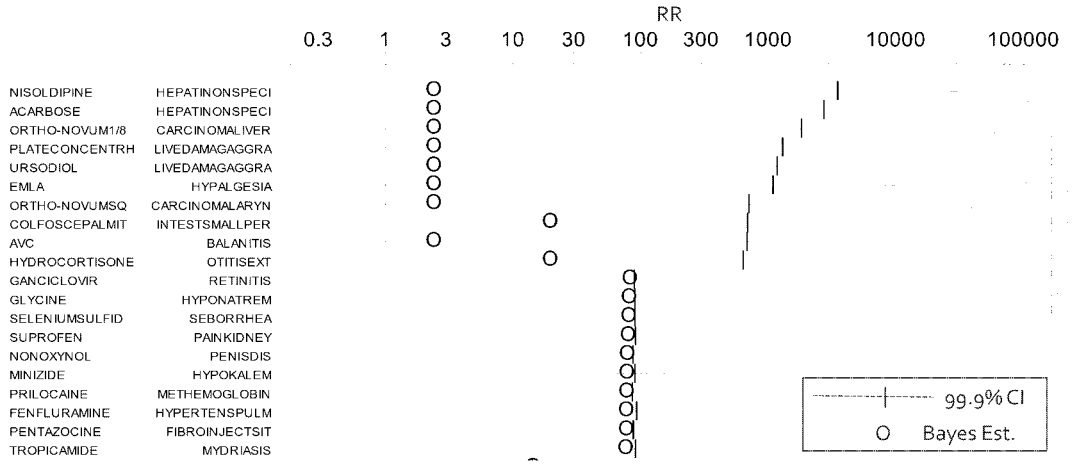
- When there are very many comparisons, these become too conservative
- True Positives Can Remain Hidden
- Very Large Datasets Can Generate too many False Positives

Hierarchical Bayesian Approach

- Fit Models that "Shrink" the Differences among the Estimates Being Compared
- Estimates typically Move toward the Average of all stand-alone Estimates
- But Error Bars typically become shorter, not longer and Effect strengths are more accurately ranked

Example of Bayesian "Shrinkage": Spontaneous Report Disproportionalities

Drug-Event Combinations with large ratios of RR = N/E = Observed/Expected counts



How Bayesian Models Decide How Much to Shrink

Shrinking Estimated Differences Provides Multiple Comparisons Protection

Requires Estimation of Variance Component or Prior Variability

- Compare Data Variation to that Expected by the Null Hypothesis
- Excess Data Variance Allows Estimate of Prior Variability
- Bayesian Calculations Produce the "Shrinkage" Estimates and Error Bars (Individual Values Move Toward Average of All Values)

Next: Two examples of advanced meta-analyses where estimation of excess data variance helped evaluate prediction accuracy

Meta-Regression for Extrapolating Across Biological Systems

DuMouchel W, Harris J (1983) Bayes methods for combining the results of cancer studies in humans and other species, *J Amer Stat Assoc* 78: 293-315 (w Discussion)

Filled cells are those BioSystem – Chemical combinations with data for fitting a dose-response model. Goal is to get better estimates of Human Lung Cancer Risk from Diesel Emissions.

γ = log of dose-response slope

$$y_{kl} = \mu + \alpha_k + \gamma_l + \delta_{kl} + \epsilon_{kl},$$

$$Y = X\beta + \delta + \epsilon, \text{ where } \theta = X\beta + \delta$$

$$\sigma \sim \Pi,$$

$$(\beta \mid \sigma) \sim N(b, V),$$

$$(\theta \mid \beta, \sigma) \sim N(X\beta, \sigma^2 I),$$

$$(Y \mid \theta, \beta, \sigma) \sim N(\theta, C).$$

	ROOF TAR	COKE OVEN	GAS ENGINE	BaP	CIG	DIESEL A	DIESEL B	DIESEL C	DIESEL D	DIESEL E
LUNG CANCER	●	●			●	?	?	?	?	●
SKIN TUMOR INIT	●	●	●	●	●	●	●		●	
VIRAL TRANSFORM	●	●	●	●	●	●	●	●	●	
MUTAGENESIS -MA	●	●	●		●	●	●	●	●	
MUTAGENESIS +MA	●	●	●		●	●	●	●	●	
	1	2	3	4	5	6	7	8	9	10

Biological Effects of Ionizing Radiation [B.E.I.R. Report IV]

from: Health Risks of Radon and other Internally Deposited Alpha-Emitters, 1988, Nat Acadamy Press [Annex 7A, by P Groer and W DuMouchel]

- *Making Better Use of Radium Dial Painters' Data by Combining Studies of Bone Cancer Risk from 4 Isotopes across 4 Biological Systems*

	ISOTOPE			
BIOLOGICAL SYSTEM	Ra-226	Ra-228	Pu-238	Pu-239
Human	●	●	?	?
Beagle Dog (Injection)	●	●		●
Beagle Dog (Inhalation)			●	
Rat			●	●

Plutonium Bone Cancer Risk Estimate:
300 Cancer Deaths per Million Person-Rad
95% Interval = (80, 1100)
5 to 10 times Larger than Risk from Radon

Complicated Problems Require Combining Several Shrinkage Calculations

Combining Multiple Studies Requires Estimating Random Study Effects

Multiple Treatment Effects Require a Separate Shrinkage Calculation

Covariate by Treatment Interactions Require More Shrinkage Parameters

Evaluating Multiple Endpoints: Choose Variables that Are Probably Correlated

- Ex: Safety ADRs—Choose Multiple MedDRA Preferred Terms within the Same Higher Level Grouped Term

Rationale for Use of Covariates

When Studies Are Randomized, Why Adjust for Covariates? Won't They all Balance Out Anyway?

- Depending on sample sizes, will not be perfect balance
- If covariates have strong effects, adjustment for them will reduce residual variance and therefore Treatment effect uncertainty
- Less focus on a single pre-specified model for safety analyses than for efficacy analyses

Main Rationale—Treatment by Covariate Interactions

- Estimating Treatment x Covariate interactions in a safety analysis is equivalent to searching for vulnerable subgroups
- MBLR– cross every Covariate with the Treatment effect

Rationale for EB Model Across End Points

Coping with Fine Granularity of Adverse Event Data or Several Efficacy Measures

- Compare T vs. C on K potential AEs or Efficacy measures that are similar in meaning
- Approach 1—separate analyses of all K measures
  - Small counts lead to non significant comparisons
  - Adjustment for multiple comparisons further reduces sensitivity
- Approach 2—define a single measure as the union or mean of the K measures
  - Significant T vs C difference may be washed out by the pooling
  - Even if significant, little information about the original K measures

Compromise Approach—EB Hierarchical Model

- K individual estimates that "borrow strength" from each other
- Estimate separate vector of coefficients for each response measure
  - But a prior distribution shrinks corresponding coefficients across responses toward each other
  - The amount of shrinkage is controlled by certain prior variances that are also estimated from the data
  - Treatment-Covariate interaction effects, which are *apriori* less likely, are also shrunk toward the null hypothesis value of 0

Bayesian Shrinkage Models

Statistical Validity of Searching for Extreme Differences

- Most significant adverse event or patient subgroup

Classical Approach to Post-Hoc Interval Estimates

- Maintain centers of CI at observed differences
- Expand widths of every CI
- Expansion is greater the more differences you look at
  - If you look at too many, the CI's are too wide to be useful

Bayesian Approach

- Requires a prior distribution for differences
  - Can estimate it from the multiple observed differences available
- Centers of CI's are "shrunk" toward average or null difference
  - High-variance differences shrink the most
- Widths of CI's usually shrink a little too
  - The more you look at, the better you can model the prior dist.

Safety Study Difficulties

Analysis of safety data versus studies of efficacy

1. Prespecification of end points is rarely possible for safety analyses
2. Sample sizes are typically inadequate for many safety issues
3. The optimal granularity of adverse event definitions is often uncertain
4. Subset analyses across subpopulations can rarely be prespecified
5. Pooled analyses of many studies are necessary to compare product safety profiles
6. Combining safety information from clinical trials and observational data may be necessary

All of the above issues can be thought of as variations on problems of multiple comparisons

Hierarchical Bayesian analysis methods analyze commonalities among the diverse effects Shrinkage techniques help the estimates "borrow strength" from each other

W DuMouchel, "Multivariate Bayesian Logistic Regression for Analysis of Clinical Study Safety Issues", *Statistical Science*, 2012, vol. 27, no. 3, 319-349). The cited pages include three invited discussions of the methodology.

Robustness to Post-Hoc Selection

Simulation Study of Bayesian Estimation from Article Cited on last Slide

- Draw "true parameters" from the prior distributions 1000 times
- Estimate main and interaction effects each time
  - **Get both MBLR and Standard "unshrunk" estimates**

Focus on Estimating the "Most Significant" Interaction (Subset Difference)

- 80 Interactions (8 covariates x 10 response events)
- For each simulation, select β_{gk} that has *largest* b_{gk}/se_{gk}
- Compare accuracy of estimates and confidence limits

Note that Bayesian Shrinkage Eliminates Selection Bias!

	SIM.COEF	SD.SIMC	BIAS	RMSE	Z.SCORE	CI.05	CI.95
MBLR	1.7651	0.6094	0.0005	0.2923	-0.0052	0.067	0.056
Stnd	1.7445	0.5981	0.2184	0.4330	0.5794	0.008	0.135

Proposal for Enhancement of the Current MBLR Method

Use Network Meta-Analysis to Compare Multiple Interventions Across Studies

Add Estimation of a Study-Level Variance Component (Shrinkage Parameter)

- Allow Study-Level Variables such as *Prospective* vs *Observational*
- Incorporate Extra Uncertainty for Observational Studies

Allow Borrowing Strength Across Multiple Efficacy Endpoints

- Example: Multiple Values of Duration Since Vaccination

Explore Markov Chain Monte Carlo Computational Approach

Recommended Aspects of Analysis Methodology

Graphical Representation of Comparisons with Confidence Intervals

- Covariate selection and interpretation of subset analyses
- Comparison of results from different input assumptions—Sensitivity Analyses

Collaboration with statisticians from FDA and elsewhere during program development

Example pools of studies with their analysis results for training purposes

Not a Replacement for Study analyses, but a uniform methodology for comparing estimates across Treatments, Studies, Endpoints and Subsets

Message

From: Alimchandani, Meghna [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=963019BC771F43CFB664C1729AAF5A17-MEGHNA.ALIM]
Sent: 12/1/2022 4:16:29 PM
To: Nair, Narayan [redacted]
Subject: RE: [EXTERNAL] RE: Weekly data mining

THANKS!!!

From: Nair, Narayan <[redacted]>
Sent: Thursday, December 1, 2022 11:14 AM
To: Moro, Pedro L (CDC) <[redacted]>
Cc: Alimchandani, Meghna <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: [EXTERNAL] RE: Weekly data mining

Hi Pedro,
I hope you are well. FYI - we have shifted some of our responsibilities in our Division so David will no longer be responsible for fielding questions about data mining. Feel free to contact me if questions come up. With regard to Alison's question, we have not had any disproportionality alerts from Data Mining for the mRNA COVID-19 vaccines for any new safety concerns (including none for Parsonage Turner Syndrome.)
A couple of key points (you probably are already familiar with these) :

- Results from data mining are considered hypothesis generating and do not, by themselves, demonstrate causal associations. They may serve as an indication for further investigation.
- The absence of disproportionality does not confirm the absence of a safety signal nor negate a signal detected by other methods.
- We generally try and avoid referring to disproportionality/data mining alerts as "signals" or "safety signals". From a regulatory perspective the terms signal and/or safety signal have certain connotations and may trigger actions so we try not conflate data mining alerts with signals.

I hope this is helpful. Thanks!

Narayan

From: Moro, Pedro (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Saturday, November 26, 2022 7:55 PM
To: Menschik, David <[redacted]>
Cc: Broder, Karen R (CDC) <[redacted]>; Lale, Allison (CDC) <[redacted]>; Nair, Narayan <[redacted]>
Subject: [EXTERNAL] RE: Weekly data mining

Hi David,

I hope your weekend is going well. I used to get these weekly data mining outputs from you but because of the FOIAS we were getting we may have asked you to stop sending them. Per Alison's comment and interest in the latest output do you think you could send us the latest data mining output you have?

Thanks

Pedro

From: Lale, Allison (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Saturday, November 26, 2022 6:28 PM
To: Moro, Pedro (CDC/DDID/NCEZID/DHQP) <[redacted]>
Cc: Broder, Karen (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: FW: Weekly data mining

Hi Pedro,

I was just wondering if we still get these data-mining alerts from FDA? In the past, we have checked this list for our own verification before presenting a CISA consult.

For example, we have an upcoming case of Parsonage Turner syndrome (PT: Neuralgic Amytrophy) following COVID-19 vaccine. We performed a VAERS search with Elaine's help, but it could be nice to say this event has not preliminarily signaled in VAERS.

Thanks,
Allison

p.s. Hope you had a good holiday!

From: Menschik, David <[redacted]>
Sent: Tuesday, July 5, 2022 7:42 AM
To: Shimabukuro, Tom (CDC/DDID/NCEZID/DHQP) <[redacted]>; Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>; Moro, Pedro (CDC/DDID/NCEZID/DHQP) <[redacted]>
Cc: Zinderman, Craig E (FDA/CBER) <[redacted]>; Nair, Narayan (FDA/CBER) <[redacted]>; Alimchandani, Meghna (FDA/CBER) <[redacted]>; Broder, Karen (CDC/DDID/NCEZID/DHQP) <[redacted]>; McNeil, Michael (CDC/DDID/NCEZID/DHQP) <[redacted]>; Lale, Allison (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: Weekly data mining

Good morning all,

Attached please find a list of all (i.e., unvetted and regardless of notability) PTs with data mining alerts (i.e., EB05 $\geq$2) for all SARS-CoV-2 vaccine VAERS reports from our weekly 'US Signals Summary Table' ('as of date' 7/1/22). Please feel free to share this hypothesis generating output with your team/command chain, though this is not intended to be shared more broadly.

Thanks,
David

From: Menschik, David
To: Nair, Narayan
Subject: FW: [EXTERNAL] RE: Weekly data mining
Date: Monday, November 28, 2022 6:06:06 AM

Good Morning Narayan,

We've previously discussed concerns about sharing our data mining output externally given history including over reliance on data mining output (e.g., 'signaling' as used below would be an inappropriate term for a data mining disproportionality finding, not to mention all the misinterpretations about absence of disproportionality misinterpretations as being reassuring, particularly in context of all the limitations including masking, etc.) and presenting findings out of context that can be misconstrued. Understood we were reverting to our traditional method of primary reviewer for a product evaluating statistical signals of disproportionality along with other methods and traditional working up of/vetting potential signals from all sources up the chain to division level prior to potential sharing of possible signals externally.
If I misunderstood, let's meet briefly to discuss. If I understood correctly, I could respond to Pedro indicating something to the effect of: 'Given all the limitations of our data mining, our standard practice is for assigned reviewers to evaluate any disproportionality findings in the context of other available data with vetting of potential signals up the chain to our division level prior to potential sharing of disproportionality data.'
Please advise.

Thanks,
David

From: Moro, Pedro (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Saturday, November 26, 2022 7:55 PM
To: Menschik, David <[redacted]>
Cc: Broder, Karen R (CDC) <[redacted]>; Lale, Allison (CDC) <[redacted]>; Nair, Narayan <[redacted]>
Subject: [EXTERNAL] RE: Weekly data mining

Hi David,

I hope your weekend is going well. I used to get these weekly data mining outputs from you but because of the FOIAS we were getting we may have asked you to stop sending them. Per Alison's comment and interest in the latest output do you think you could send us the latest data mining output you have?

Thanks

Pedro

From: Lale, Allison (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Saturday, November 26, 2022 6:28 PM
To: Moro, Pedro (CDC/DDID/NCEZID/DHQP) <[redacted]>
Cc: Broder, Karen (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: FW: Weekly data mining

Hi Pedro,

I was just wondering if we still get these data-mining alerts from FDA? In the past, we have checked this list for our own verification before presenting a CISA consult.

For example, we have an upcoming case of Parsonage Turner syndrome (PT: Neuralgic Amytrophy) following COVID-19 vaccine. We performed a VAERS search with Elaine's help, but it could be nice to say this event has not preliminarily signaled in VAERS.

Thanks,
Allison

p.s. Hope you had a good holiday!

From: Menschik, David <[redacted]>
Sent: Tuesday, July 5, 2022 7:42 AM
To: Shimabukuro, Tom (CDC/DDID/NCEZID/DHQP) <[redacted]>; Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>; Moro, Pedro (CDC/DDID/NCEZID/DHQP) <[redacted]>
Cc: Zinderman, Craig E (FDA/CBER) <[redacted]>; Nair, Narayan (FDA/CBER) <[redacted]>; Alimchandani, Meghna (FDA/CBER) <[redacted]>; Broder, Karen (CDC/DDID/NCEZID/DHQP) <[redacted]>; McNeil, Michael (CDC/DDID/NCEZID/DHQP) <[redacted]>; Lale, Allison (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: Weekly data mining

Good morning all,

Attached please find a list of all (i.e., unvetted and regardless of notability) PTs with data mining alerts (i.e., EB05 ≥2) for all SARS-CoV-2 vaccine VAERS reports from our weekly 'US Signals Summary Table' ('as of date' 7/1/22). Please feel free to share this hypothesis generating output with your team/command chain, though this is not intended to be shared more broadly.

Thanks,
David

Message

From: Niu, Manette [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=EE2A4A5155724814A9836019691CABA7-MANETTE.NIU]
Sent: 4/29/2021 10:35:38 AM
To: Zinderman, Craig E [redacted]
CC: Baer, Bethany [redacted]; Menschik, David [redacted]
Subject: FW: a more efficient way to find events of interest
Attachments: Audit trail for the TTP (3D 2010-2021) DM run.pdf; Output for the TTP (3D 2010-2021) DM Run.pdf

fyi

From: Szarfman, Ana <[redacted]>
Sent: Thursday, April 29, 2021 12:49 AM
To: Allende, Maria <[redacted]>; Niu, Manette <[redacted]>
Cc: Stockbridge, Norman L <[redacted]>
Subject: a more efficient way to find events of interest

Hi Maria and Manette,

I am sharing an analysis that was requested at your end.

Please refer to the attached audit trail and to the companion 3D data mining analysis displaying all TTP cases reported for COVID-19 vaccines as of April 23, 2021 .

By grouping PTs and HLTs representing TTP into a custom term, and using a 3D display of "vaccine--PT--custom term" it enables the reviewer to focus on every associated single event of interest with each vaccine. The associated reports can be easily grouped and accessed by drilling down techniques.

See highlighted in yellow potential events that may be associated with brain TTP.

Let me know if you need any additional feedback.

--Ana

Ana Szarfman, MD, PhD, FAMIA,
Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), and Fellow of the American Medical Informatics Association (2020)
Medical Officer, Safety Data Mining Developer and Medical Informatics Analyst,
Celebrating nearly a quarter of a century of successful implementation of safety data mining, interactive patient profiles, and other automated analytical tools.
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted]
[redacted](office)
[redacted](personal cell phone and WhatsApp)

U.S. FOOD & DRUG ADMINISTRATION

Help

Detail for Run "TTP(3D, 2010-21)"

ID:	31881	
Type:	MGPS	
Name:	TTP(3D, 2010-21)	
Description:	Data as 4/23/2021 - Vaccine Name, PT; 3D; Stratified by Sex, AgeGroup11, 2010-2021; Minimum count=2 Added fibrin d dimer increased & Platelet factor 4	
Project:	Clinical Informatics	
Configuration:	VAERS_M_TS	
Configuration description:	Vaers data extracted from VAERS_M account; Data is refreshed weekly.	
As of date:	04/23/2021 00:00:00	
Database restriction:	Received Year equals any of the following values: '2010', '2011', '2012', '2013', '2014', '2015', '2016', '2017', '2018', '2019', '2020', '2021'	
Item variables:	Vaccine Name, Symptom: PT	
Custom terms:	TTP_fibrin D dimer_platelet factor 4 (Custom Term) for Symptom: PT	Selection logic: ((1 union 4) intersect (2 union 3 union 5)) 1) Symptom: PT equals any of the following values: 'Acquired amegakaryocytic thrombocytopenia', 'Amegakaryocytic thrombocytopenia', 'Autoimmune heparin-induced thrombocytopenia', 'Congenital thrombocytopenia', 'Cutaneovisceral angiomatosis with thrombocytopenia', 'Disseminated intravascular coagulation', 'Fibrin D dimer increased', 'Fibrin abnormal', 'Fibrin degradation products increased', 'Fibrin increased', 'Fibrinolysis', 'Fibrinolysis increased', 'Haemangioma-thrombocytopenia syndrome', 'Heparin-induced thrombocytopenia', 'Heparin-induced thrombocytopenia test', 'Heparin-induced thrombocytopenia test positive', 'Immune thrombocytopenia', 'Neonatal alloimmune thrombocytopenia', 'Non-immune heparin associated thrombocytopenia', 'Platelet count decreased', 'Platelet dysfunction', 'Platelet factor 4', 'Platelet factor 4 increased', 'Severe fever with thrombocytopenia syndrome', 'Spontaneous heparin-induced thrombocytopenia syndrome', 'Thrombocytopenia', 'Thrombocytopenia neonatal', 'Thrombocytopenia-absent radius syndrome' 2) Narrative contains 'FIBRIN D DIMER INCREASED' 3) Narrative contains 'factor 4' 4) Symptom: HLT equals 'Thrombocytopenias' 5) Symptom: HLT equals any of the following values: 'Aortic embolism and thrombosis', 'Cerebrovascular embolism and thrombosis', 'Cerebrovascular venous and sinus thrombosis', 'Gastrointestinal embolism and thrombosis', 'Gastrointestinal vascular occlusion and infarction', 'Hepatic and portal embolism and thrombosis', 'Non-site specific embolism and thrombosis', 'Peripheral embolism and thrombosis', 'Pulmonary embolism and thrombosis', 'Pulmonary thrombotic and embolic conditions', 'Renal embolism and thrombosis', 'Retinal embolism and

	embolism and thrombosis', 'Retinal embolism and thrombosis', 'Site specific embolism and thrombosis NEC', 'Vena caval embolism and thrombosis'
Stratification variables:	Sex, AgeGroup11
Event Hierarchy:	MedDRA 23.1
Highest dimension:	3
Minimum count:	2
Calculate PRR:	No
Calculate ROR:	No
Fill in hierarchy values:	No
Exclude single itemtypes:	Yes
Fit separate distributions:	No
Save intermediate files:	No
Created by:	Ana Szarfman
Created on:	04/26/2021 21:22:31 EDT
User:	Ana Szarfman
Source database:	Source Data: VAERS data as of April23, 2021 from VAERS data from VAERS_M as of 04/23/2021 00:00:00 loaded on 2021-04-25 00:00:00.0

Close

Configuration: VAERS_M_TS **Run :** TTP(3D, 2010-21) **Run ID:** 31881
Dimension: 3 **Selection Criteria:** Vaccine Name(COVID19 (COVID19 (JANSSEN)), COVID19 (COVID19 (MODERNA)), COVID19 (COVID19 (PFIZER-BIONTECH))) + Symptom: PT(PT=TTP_fibrin D dimer_platelet factor 4 (Custom Term)) + Any Symptom: PT(PT)
548 rows Sorted by EBGM desc

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	Platelet count	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	10	7.63	12.8	20.5	1.52
COVID19 (COVID19 (JANSSEN))	Cerebral haemorrhage	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	9	6.37	11.0	18.0	1.32
COVID19 (COVID19 (JANSSEN))	Cerebral venous sinus thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	13	6.87	10.9	16.5	0.647
COVID19 (COVID19 (JANSSEN))	Cerebral haematoma	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	4.94	10.8	21.2	0.539
COVID19 (COVID19 (JANSSEN))	Blood fibrinogen	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	4.90	10.7	21.1	0.936
COVID19 (COVID19 (JANSSEN))	Portal vein thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	4.90	10.7	21.0	0.568
COVID19 (COVID19 (JANSSEN))	Mean cell haemoglobin concentration normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	4.86	10.6	20.9	1.07
COVID19 (COVID19 (JANSSEN))	Cerebral mass effect	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	4.96	10.1	18.8	0.706
COVID19 (COVID19 (JANSSEN))	Prothrombin time shortened	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	4.62	10.1	19.8	1.02
COVID19 (COVID19 (JANSSEN))	Fibrin D dimer	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	5.43	10.0	17.3	0.920
COVID19		TTP_fibrin D dimer_platelet					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(COVID19 (JANSSEN))	Platelet factor 4	factor 4 (Custom Term)	5	4.84	9.85	18.3	0.425
COVID19 (COVID19 (JANSSEN))	Heparin-induced thrombocytopenia test positive	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	4.47	9.77	19.2	0.400
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Thrombectomy	6	5.04	9.72	17.3	0.599
COVID19 (COVID19 (JANSSEN))	Platelet count normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	5.01	9.66	17.2	1.10
COVID19 (COVID19 (JANSSEN))	Magnetic resonance imaging head	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	4.98	9.61	17.1	1.10
COVID19 (COVID19 (JANSSEN))	Blood fibrinogen decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	8	5.36	9.53	15.9	1.18
COVID19 (COVID19 (JANSSEN))	Subarachnoid haemorrhage	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	3.91	9.42	19.9	0.860
COVID19 (COVID19 (JANSSEN))	Blood smear test normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	3.21	8.90	20.8	0.670
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Venogram normal	3	3.69	8.88	18.7	0.723
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Transverse sinus thrombosis	5	4.33	8.82	16.4	0.436
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Vasogenic cerebral oedema	2	3.17	8.80	20.5	0.507
COVID19 (COVID19 (JANSSEN))	Superior sagittal sinus thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	4.30	8.76	16.3	0.520
COVID19 (COVID19 (JANSSEN))	Angiogram cerebral abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom	4	4.00	8.74	17.2	0.856

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
		Term)					
COVID19 (COVID19 (JANSSEN))	Computerised tomogram head abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	14	5.62	8.74	13.1	1.23
COVID19 (COVID19 (JANSSEN))	Blood sodium normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	3.60	8.66	18.3	0.790
COVID19 (COVID19 (JANSSEN))	Computerised tomogram neck	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	3.58	8.63	18.2	0.787
COVID19 (COVID19 (JANSSEN))	Pneumatosis intestinalis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	3.08	8.54	19.9	0.373
COVID19 (COVID19 (JANSSEN))	Haematocrit normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	3.90	8.52	16.7	0.856
COVID19 (COVID19 (JANSSEN))	Albumin globulin ratio	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	3.05	8.45	19.7	0.633
COVID19 (COVID19 (JANSSEN))	Blood sodium	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	3.03	8.42	19.6	0.666
COVID19 (COVID19 (JANSSEN))	Peripheral artery thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	3.03	8.40	19.6	0.338
COVID19 (COVID19 (JANSSEN))	Blood smear test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.99	8.30	19.3	0.657
COVID19 (COVID19 (JANSSEN))	Scan with contrast abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	4.06	8.28	15.4	0.893
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Venogram	3	3.43	8.27	17.4	0.581
COVID19		TTP_fibrin D					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(COVID19 (JANSSEN))	Heparin-induced thrombocytopenia test	dimer_platelet factor 4 (Custom Term)	5	4.02	8.20	15.3	0.487
COVID19 (COVID19 (JANSSEN))	Blood chloride normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.95	8.19	19.1	0.649
COVID19 (COVID19 (JANSSEN))	Gastrointestinal necrosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.92	8.11	18.9	0.502
COVID19 (COVID19 (JANSSEN))	Posturing	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.92	8.11	18.9	0.642
COVID19 (COVID19 (JANSSEN))	Cerebral congestion	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.92	8.09	18.9	0.403
COVID19 (COVID19 (JANSSEN))	Haemoglobin	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.91	8.08	18.8	0.639
COVID19 (COVID19 (JANSSEN))	Blood bilirubin	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.90	8.06	18.8	0.638
COVID19 (COVID19 (JANSSEN))	Aspartate aminotransferase	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.89	8.01	18.7	0.634
COVID19 (COVID19 (JANSSEN))	Blood lactic acid	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	3.91	7.97	14.8	0.859
COVID19 (COVID19 (JANSSEN))	Alanine aminotransferase	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.86	7.94	18.5	0.628
COVID19 (COVID19 (JANSSEN))	Haptoglobin increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.84	7.87	18.3	0.623
COVID19 (COVID19	Retching	TTP_fibrin D dimer_platelet factor 4 (Custom	2	2.82	7.84	18.3	0.620

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(JANSSEN))		Term)					
COVID19 (COVID19 (JANSSEN))	Full blood count abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	3.55	7.76	15.2	0.779
COVID19 (COVID19 (JANSSEN))	Computerised tomogram head	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	4.19	7.74	13.3	0.921
COVID19 (COVID19 (JANSSEN))	Blood alkaline phosphatase normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.79	7.74	18.0	0.612
COVID19 (COVID19 (JANSSEN))	Haemorrhagic stroke	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.79	7.73	18.0	0.612
COVID19 (COVID19 (JANSSEN))	Cerebral infarction	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.75	7.63	17.8	0.604
COVID19 (COVID19 (JANSSEN))	Brain herniation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	3.17	7.63	16.1	0.625
COVID19 (COVID19 (JANSSEN))	Hemiparesis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	3.48	7.62	15.0	0.765
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	White blood cell count normal	7	4.12	7.59	13.1	0.904
COVID19 (COVID19 (JANSSEN))	Hepatitis B surface antigen	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.73	7.56	17.6	0.599
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Troponin I normal	2	2.68	7.45	17.4	0.590
COVID19 (COVID19 (JANSSEN))	Monocyte percentage	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.67	7.40	17.3	0.586
COVID19 (COVID19	Red cell distribution	TTP_fibrin D dimer_platelet	3	3.03	7.29	15.4	0.665

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(JANSSEN))	width normal	factor 4 (Custom Term)					
COVID19 (COVID19 (JANSSEN))	Mean platelet volume normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.62	7.27	16.9	0.576
COVID19 (COVID19 (JANSSEN))	Blood alkaline phosphatase	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.62	7.26	16.9	0.574
COVID19 (COVID19 (JANSSEN))	Myocardial strain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.60	7.22	16.8	0.292
COVID19 (COVID19 (JANSSEN))	Anticoagulant therapy	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	3.92	7.22	12.4	0.555
COVID19 (COVID19 (JANSSEN))	Haptoglobin	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.60	7.20	16.8	0.491
COVID19 (COVID19 (JANSSEN))	Computerised tomogram abdomen	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.96	7.14	15.0	0.651
COVID19 (COVID19 (JANSSEN))	Blood potassium normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.96	7.13	15.0	0.650
COVID19 (COVID19 (JANSSEN))	Cholelithiasis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.53	7.03	16.4	0.557
COVID19 (COVID19 (JANSSEN))	Haemoglobin normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	3.20	7.00	13.7	0.703
COVID19 (COVID19 (JANSSEN))	Mean cell volume normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.52	7.00	16.3	0.554
COVID19 (COVID19 (JANSSEN))	Splenic vein thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.52	6.99	16.3	0.311

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	Mean cell haemoglobin normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.52	6.99	16.3	0.554
COVID19 (COVID19 (JANSSEN))	Blood bilirubin normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.49	6.91	16.1	0.547
COVID19 (COVID19 (JANSSEN))	Glomerular filtration rate	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.47	6.86	16.0	0.543
COVID19 (COVID19 (JANSSEN))	Red blood cell count normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.84	6.85	14.4	0.625
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	White blood cell count	2	2.46	6.82	15.9	0.540
COVID19 (COVID19 (JANSSEN))	Angiogram cerebral normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.44	6.78	15.8	0.537
COVID19 (COVID19 (JANSSEN))	Activated partial thromboplastin time	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.42	6.70	15.6	0.531
COVID19 (COVID19 (JANSSEN))	Coma scale abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.39	6.64	15.5	0.526
COVID19 (COVID19 (JANSSEN))	Abdominal X-ray	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.73	6.57	13.8	0.599
COVID19 (COVID19 (JANSSEN))	Blood albumin normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.36	6.53	15.2	0.517
COVID19 (COVID19 (JANSSEN))	Aphasia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.35	6.53	15.2	0.517
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Venogram abnormal	5	3.20	6.52	12.1	0.403

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	Blood thyroid stimulating hormone normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.33	6.47	15.1	0.512
COVID19 (COVID19 (JANSSEN))	Photophobia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.68	6.46	13.6	0.589
COVID19 (COVID19 (JANSSEN))	COVID-19 pneumonia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.32	6.45	15.0	0.511
COVID19 (COVID19 (JANSSEN))	Pulmonary hypertension	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.32	6.43	15.0	0.509
COVID19 (COVID19 (JANSSEN))	Oxygen saturation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.31	6.41	14.9	0.438
COVID19 (COVID19 (JANSSEN))	Pupil fixed	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.30	6.37	14.9	0.505
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound abdomen normal	2	2.29	6.35	14.8	0.503
COVID19 (COVID19 (JANSSEN))	Metabolic function test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	3.29	6.34	11.3	0.722
COVID19 (COVID19 (JANSSEN))	Protein S	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.25	6.24	14.6	0.411
COVID19 (COVID19 (JANSSEN))	Blood calcium normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.25	6.23	14.5	0.493
COVID19 (COVID19 (JANSSEN))	Antiphospholipid antibodies	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.23	6.18	14.4	0.490
COVID19 (COVID19	Immunology test	TTP_fibrin D dimer_platelet factor 4 (Custom	2	2.23	6.18	14.4	0.489

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(JANSSEN))		Term)					
COVID19 (COVID19 (JANSSEN))	COVID-19	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	3.03	6.18	11.5	0.666
COVID19 (COVID19 (JANSSEN))	Full blood count	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	3.18	6.13	10.9	0.699
COVID19 (COVID19 (JANSSEN))	Cardiolipin antibody	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.21	6.13	14.3	0.485
COVID19 (COVID19 (JANSSEN))	Computerised tomogram head normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	3.14	6.04	10.8	0.689
COVID19 (COVID19 (JANSSEN))	Angiogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	3.12	6.02	10.7	0.686
COVID19 (COVID19 (JANSSEN))	Migraine	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.49	6.00	12.6	0.548
COVID19 (COVID19 (JANSSEN))	Lung opacity	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.49	6.00	12.6	0.547
COVID19 (COVID19 (JANSSEN))	Cerebrovascular accident	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.12	5.89	13.7	0.466
COVID19 (COVID19 (JANSSEN))	Jugular vein thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.45	5.89	12.4	0.325
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Urinary system X-ray	2	2.12	5.89	13.7	0.466
COVID19 (COVID19 (JANSSEN))	International normalised ratio normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	2.68	5.86	11.5	0.589
COVID19	TTP_fibrin D						

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(COVID19 (JANSSEN))	dimer_platelet factor 4 (Custom Term)	Thrombocytopenia	17	3.92	5.86	8.49	0.860
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Venous occlusion	3	2.43	5.85	12.3	0.343
COVID19 (COVID19 (JANSSEN))	Aspartate aminotransferase normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.09	5.80	13.5	0.459
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound scan	6	2.97	5.73	10.2	0.653
COVID19 (COVID19 (JANSSEN))	Alanine aminotransferase normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	2.06	5.71	13.3	0.452
COVID19 (COVID19 (JANSSEN))	Mental status changes	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	2.80	5.69	10.6	0.614
COVID19 (COVID19 (JANSSEN))	SARS-CoV-2 test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	2.60	5.68	11.1	0.571
COVID19 (COVID19 (JANSSEN))	Protein total normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.34	5.63	11.9	0.513
COVID19 (COVID19 (JANSSEN))	Platelet transfusion	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	2.54	5.56	10.9	0.558
COVID19 (COVID19 (JANSSEN))	Antinuclear antibody	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.95	5.40	12.6	0.427
COVID19 (COVID19 (JANSSEN))	Blood glucose normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.93	5.36	12.5	0.425
COVID19 (COVID19 (JANSSEN))	Antithrombin III	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.93	5.36	12.5	0.424
COVID19		TTP_fibrin D					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(COVID19 (JANSSEN))	Brain oedema	dimer_platelet factor 4 (Custom Term)	3	2.21	5.32	11.2	0.486
COVID19 (COVID19 (JANSSEN))	Brain injury	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.92	5.32	12.4	0.421
COVID19 (COVID19 (JANSSEN))	Skin discolouration	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.92	5.32	12.4	0.421
COVID19 (COVID19 (JANSSEN))	Protein C	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.92	5.31	12.4	0.421
COVID19 (COVID19 (JANSSEN))	Anxiety	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	2.74	5.28	9.41	0.602
COVID19 (COVID19 (JANSSEN))	Headache	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	23	3.71	5.25	7.26	0.816
COVID19 (COVID19 (JANSSEN))	Metabolic acidosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.89	5.24	12.2	0.415
COVID19 (COVID19 (JANSSEN))	Pupillary reflex impaired	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.87	5.18	12.1	0.410
COVID19 (COVID19 (JANSSEN))	Scan with contrast	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.85	5.14	12.0	0.407
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound Doppler	6	2.66	5.13	9.14	0.584
COVID19 (COVID19 (JANSSEN))	Contusion	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	2.34	5.12	10.0	0.514
COVID19 (COVID19 (JANSSEN))	SARS-CoV-2 test positive	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	2.33	5.09	9.98	0.511

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	Gene mutation identification test negative	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.09	5.04	10.6	0.460
COVID19 (COVID19 (JANSSEN))	Haemorrhage	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.81	5.02	11.7	0.397
COVID19 (COVID19 (JANSSEN))	Magnetic resonance imaging	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	2.44	4.96	9.23	0.535
COVID19 (COVID19 (JANSSEN))	Electrocardiogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	2.43	4.96	9.22	0.535
COVID19 (COVID19 (JANSSEN))	Angiogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.06	4.95	10.4	0.452
COVID19 (COVID19 (JANSSEN))	Magnetic resonance imaging head normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	2.04	4.90	10.3	0.447
COVID19 (COVID19 (PFIZER-BIONTECH))	Haemorrhage intracranial	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.74	4.82	11.2	0.467
COVID19 (COVID19 (JANSSEN))	Blood sodium increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	2.19	4.79	9.41	0.481
COVID19 (COVID19 (JANSSEN))	Computerised tomogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	8	2.64	4.70	7.87	0.581
COVID19 (COVID19 (JANSSEN))	Seizure	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	2.15	4.70	9.23	0.472
COVID19 (COVID19 (JANSSEN))	Shock	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.69	4.68	10.9	0.370
COVID19	Blood smear test	TTP_fibrin D dimer_platelet					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(COVID19 (JANSSEN))	abnormal	factor 4 (Custom Term)	2	1.69	4.68	10.9	0.370
COVID19 (COVID19 (JANSSEN))	Platelet count decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	30	3.44	4.66	6.20	0.754
COVID19 (COVID19 (JANSSEN))	Electroencephalogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.93	4.64	9.79	0.424
COVID19 (COVID19 (JANSSEN))	Neck pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	2.11	4.61	9.05	0.463
COVID19 (COVID19 (JANSSEN))	Feeling abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.65	4.58	10.7	0.362
COVID19 (COVID19 (MODERNA))	Right ventricular dilatation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.65	4.57	10.7	0.266
COVID19 (COVID19 (JANSSEN))	Activated partial thromboplastin time shortened	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.89	4.55	9.58	0.415
COVID19 (COVID19 (JANSSEN))	Death	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	2.35	4.53	8.08	0.517
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Vena cava filter insertion	2	1.63	4.52	10.5	0.229
COVID19 (COVID19 (JANSSEN))	Magnetic resonance imaging head abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	10	2.65	4.45	7.11	0.582
COVID19 (COVID19 (JANSSEN))	Muscle spasms	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.60	4.44	10.3	0.351
COVID19 (COVID19 (JANSSEN))	Cardiac arrest	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.58	4.40	10.2	0.348

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	Lung infiltration	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.57	4.37	10.2	0.346
COVID19 (COVID19 (JANSSEN))	Immune thrombocytopenia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.57	4.36	10.1	0.345
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Unresponsive to stimuli	3	1.80	4.32	9.11	0.395
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Vision blurred	2	1.55	4.31	10.0	0.341
COVID19 (COVID19 (JANSSEN))	Antiphospholipid antibodies negative	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.55	4.29	10.00	0.340
COVID19 (COVID19 (JANSSEN))	Fibrin D dimer increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	40	3.28	4.27	5.48	0.721
COVID19 (COVID19 (JANSSEN))	Computerised tomogram thorax abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	13	2.66	4.21	6.40	0.585
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Troponin	3	1.73	4.17	8.78	0.380
COVID19 (COVID19 (JANSSEN))	Blood pressure increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.50	4.16	9.70	0.330
COVID19 (COVID19 (PFIZER-BIONTECH))	Interleukin-2 receptor assay	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.45	4.02	9.38	0.295
COVID19 (COVID19 (JANSSEN))	Dizziness	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	2.08	4.00	7.14	0.456
COVID19 (COVID19 (JANSSEN))	Prothrombin time prolonged	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	9	2.30	3.98	6.49	0.506

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	Brain natriuretic peptide normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.42	3.93	9.17	0.311
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound abdomen abnormal	3	1.62	3.90	8.23	0.356
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Thrombophlebitis superficial	3	1.62	3.89	8.20	0.226
COVID19 (COVID19 (PFIZER-BIONTECH))	Hypercoagulation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.37	3.81	8.89	0.350
COVID19 (COVID19 (JANSSEN))	Activated partial thromboplastin time prolonged	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	1.87	3.81	7.08	0.410
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound Doppler abnormal	10	2.24	3.77	6.02	0.493
COVID19 (COVID19 (JANSSEN))	Confusional state	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.56	3.76	7.92	0.343
COVID19 (COVID19 (JANSSEN))	Erythema	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.34	3.73	8.68	0.295
COVID19 (COVID19 (JANSSEN))	Blood urea decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.70	3.73	7.31	0.374
COVID19 (COVID19 (PFIZER-BIONTECH))	SARS-CoV-2 test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.70	3.71	7.28	0.266
COVID19 (COVID19 (JANSSEN))	Angiogram pulmonary abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	9	2.14	3.70	6.04	0.471
COVID19 (COVID19 (JANSSEN))	Atelectasis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.52	3.66	7.70	0.333

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	Peripheral swelling	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.66	3.64	7.13	0.365
COVID19 (COVID19 (JANSSEN))	Abdominal pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	1.87	3.61	6.43	0.411
COVID19 (COVID19 (JANSSEN))	Heart rate increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.50	3.60	7.59	0.329
COVID19 (COVID19 (JANSSEN))	Immunoglobulin therapy	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	1.93	3.56	6.14	0.424
COVID19 (COVID19 (JANSSEN))	Oxygen saturation decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.27	3.52	8.21	0.279
COVID19 (COVID19 (PFIZER-BIONTECH))	Ischaemic stroke	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.26	3.51	8.18	0.375
COVID19 (COVID19 (PFIZER-BIONTECH))	Lipase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.26	3.50	8.15	0.633
COVID19 (COVID19 (JANSSEN))	Pulmonary embolism	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	28	2.53	3.47	4.66	0.557
COVID19 (COVID19 (MODERNA))	PCO2 decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.25	3.46	8.06	0.446
COVID19 (COVID19 (PFIZER-BIONTECH))	Cyanosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.24	3.44	8.03	0.624
COVID19 (COVID19 (JANSSEN))	Glomerular filtration rate decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.43	3.44	7.25	0.314
COVID19 (COVID19	TTP_fibrin D dimer_platelet factor 4	Thrombosis	10	2.05	3.44	5.50	0.281

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(JANSSEN))	(Custom Term)						
COVID19 (COVID19 (PFIZER-BIONTECH))	Chest X-ray	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	1.85	3.41	5.88	0.930
COVID19 (COVID19 (PFIZER-BIONTECH))	COVID-19 pneumonia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.22	3.38	7.88	0.272
COVID19 (COVID19 (PFIZER-BIONTECH))	Pulmonary mass	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.40	3.36	7.08	0.701
COVID19 (COVID19 (JANSSEN))	Endotracheal intubation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	8	1.89	3.36	5.62	0.415
COVID19 (COVID19 (JANSSEN))	Blood lactic acid increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.52	3.32	6.52	0.334
COVID19 (COVID19 (PFIZER-BIONTECH))	COVID-19	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	1.63	3.31	6.16	0.248
COVID19 (COVID19 (JANSSEN))	Musculoskeletal stiffness	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.19	3.30	7.68	0.261
COVID19 (COVID19 (PFIZER-BIONTECH))	Pleuritic pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	1.62	3.29	6.13	0.813
COVID19 (COVID19 (MODERNA))	Echocardiogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	1.70	3.27	5.83	0.813
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Troponin increased	3	1.35	3.26	6.86	0.297
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Vomiting	7	1.74	3.22	5.54	0.383
COVID19 (COVID19	SARS-CoV-2 test	TTP_fibrin D dimer_platelet	13	2.02	3.20	4.86	0.833

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(MODERNA))	negative	factor 4 (Custom Term)					
COVID19 (COVID19 (JANSSEN))	Leukocytosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.15	3.19	7.44	0.253
COVID19 (COVID19 (JANSSEN))	Oedema peripheral	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.14	3.18	7.40	0.251
COVID19 (COVID19 (PFIZER-BIONTECH))	International normalised ratio	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.14	3.16	7.36	0.407
COVID19 (COVID19 (JANSSEN))	Echocardiogram normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.14	3.15	7.35	0.250
COVID19 (COVID19 (JANSSEN))	Disseminated intravascular coagulation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.13	3.13	7.29	0.248
COVID19 (COVID19 (JANSSEN))	Computerised tomogram thorax	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.42	3.10	6.09	0.311
COVID19 (COVID19 (JANSSEN))	Pleural effusion	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.29	3.10	6.52	0.282
COVID19 (COVID19 (JANSSEN))	Antinuclear antibody negative	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.28	3.07	6.48	0.280
COVID19 (COVID19 (JANSSEN))	International normalised ratio increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	1.65	3.04	5.23	0.362
COVID19 (COVID19 (PFIZER-BIONTECH))	Echocardiogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	1.48	3.02	5.61	0.702
COVID19 (COVID19 (PFIZER-BIONTECH))	SARS-CoV-2 test negative	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	11	1.83	3.01	4.72	0.506

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (PFIZER-BIONTECH))	Pericardial effusion	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.24	3.00	6.31	0.552
COVID19 (COVID19 (PFIZER-BIONTECH))	Mouth haemorrhage	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.08	2.98	6.96	0.530
COVID19 (COVID19 (JANSSEN))	Intensive care	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	11	1.81	2.97	4.66	0.397
COVID19 (COVID19 (PFIZER-BIONTECH))	Brain natriuretic peptide normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.23	2.96	6.24	0.360
COVID19 (COVID19 (MODERNA))	Liver function test increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.06	2.95	6.88	0.642
COVID19 (COVID19 (JANSSEN))	Pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	9	1.70	2.93	4.79	0.373
COVID19 (COVID19 (PFIZER-BIONTECH))	SARS-CoV-2 test positive	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.34	2.92	5.73	0.202
COVID19 (COVID19 (JANSSEN))	Syncope	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.21	2.92	6.14	0.266
COVID19 (COVID19 (PFIZER-BIONTECH))	Oxygen therapy	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.05	2.91	6.78	0.465
COVID19 (COVID19 (PFIZER-BIONTECH))	Lung opacity	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.20	2.90	6.10	0.291
COVID19 (COVID19 (PFIZER-BIONTECH))	Full blood count normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.19	2.88	6.06	0.600
COVID19 (COVID19	Computerised	TTP_fibrin D dimer_platelet					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(PFIZER-BIONTECH))	tomogram abdomen	factor 4 (Custom Term)	2	1.03	2.87	6.68	0.377
COVID19 (COVID19 (JANSSEN))	Mechanical ventilation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.31	2.87	5.63	0.288
COVID19 (COVID19 (PFIZER-BIONTECH))	Neutrophil count	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.03	2.86	6.67	0.450
COVID19 (COVID19 (PFIZER-BIONTECH))	C-reactive protein normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.03	2.86	6.66	0.518
COVID19 (COVID19 (JANSSEN))	Brain natriuretic peptide increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.03	2.86	6.66	0.226
COVID19 (COVID19 (JANSSEN))	Deep vein thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	13	1.79	2.83	4.30	0.330
COVID19 (COVID19 (JANSSEN))	Mean cell volume increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.02	2.82	6.57	0.223
COVID19 (COVID19 (MODERNA))	Blood test normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.17	2.82	5.94	0.706
COVID19 (COVID19 (PFIZER-BIONTECH))	Hypopnoea	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	1.01	2.80	6.53	0.508
COVID19 (COVID19 (JANSSEN))	Respiratory failure	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.28	2.80	5.50	0.281
COVID19 (COVID19 (JANSSEN))	Pain in extremity	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	1.52	2.80	4.82	0.333
COVID19 (COVID19 (JANSSEN))	Condition aggravated	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.16	2.79	5.89	0.255

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound Doppler normal	2	1.00	2.79	6.50	0.221
COVID19 (COVID19 (MODERNA))	Oxygen saturation decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	1.36	2.76	5.14	0.617
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood culture	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.995	2.76	6.44	0.500
COVID19 (COVID19 (PFIZER-BIONTECH))	Cardiac failure	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.15	2.76	5.82	0.469
COVID19 (COVID19 (MODERNA))	Blood urea normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.994	2.76	6.43	0.600
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Troponin normal	4	1.26	2.75	5.40	0.545
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound scan abnormal	5	1.35	2.75	5.11	0.296
COVID19 (COVID19 (PFIZER-BIONTECH))	Pulmonary infarction	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.14	2.74	5.78	0.328
COVID19 (COVID19 (JANSSEN))	Blood gases abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.25	2.74	5.38	0.275
COVID19 (COVID19 (PFIZER-BIONTECH))	Rales	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.987	2.74	6.38	0.496
COVID19 (COVID19 (JANSSEN))	Rash	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.982	2.73	6.35	0.216
COVID19 (COVID19 (JANSSEN))	Painful respiration	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.980	2.72	6.34	0.215

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	Mean cell haemoglobin increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.977	2.71	6.32	0.215
COVID19 (COVID19 (PFIZER-BIONTECH))	Procalcitonin increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.24	2.71	5.32	0.539
COVID19 (COVID19 (PFIZER-BIONTECH))	Activated partial thromboplastin time shortened	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.23	2.68	5.27	0.617
COVID19 (COVID19 (MODERNA))	Body temperature increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.958	2.66	6.20	0.578
COVID19 (COVID19 (PFIZER-BIONTECH))	Electrocardiogram normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	1.44	2.65	4.57	0.722
COVID19 (COVID19 (JANSSEN))	Nausea	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	8	1.48	2.64	4.42	0.326
COVID19 (COVID19 (JANSSEN))	Chest X-ray	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.947	2.63	6.12	0.208
COVID19 (COVID19 (PFIZER-BIONTECH))	Laboratory test normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.944	2.62	6.10	0.475
COVID19 (COVID19 (PFIZER-BIONTECH))	Cerebrovascular accident	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.942	2.61	6.09	0.385
COVID19 (COVID19 (MODERNA))	Hiatus hernia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.941	2.61	6.08	0.568
COVID19 (COVID19 (JANSSEN))	Blood lactate dehydrogenase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.937	2.60	6.06	0.206
COVID19 (COVID19	Computerised	TTP_fibrin D dimer_platelet					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(PFIZER-BIONTECH))	tomogram thorax	factor 4 (Custom Term)	7	1.40	2.58	4.44	0.504
COVID19 (COVID19 (JANSSEN))	Blood chloride decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.926	2.57	5.99	0.203
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Troponin increased	5	1.26	2.56	4.76	0.471
COVID19 (COVID19 (JANSSEN))	Laboratory test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.921	2.56	5.96	0.202
COVID19 (COVID19 (PFIZER-BIONTECH))	Oxygen saturation decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.16	2.54	4.98	0.497
COVID19 (COVID19 (MODERNA))	Blood lactic acid	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.02	2.46	5.18	0.384
COVID19 (COVID19 (JANSSEN))	Acute respiratory failure	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.886	2.46	5.73	0.195
COVID19 (COVID19 (JANSSEN))	Pneumonia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.02	2.45	5.16	0.224
COVID19 (COVID19 (JANSSEN))	Chest discomfort	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.02	2.45	5.16	0.223
COVID19 (COVID19 (PFIZER-BIONTECH))	Acute myocardial infarction	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.881	2.44	5.70	0.283
COVID19 (COVID19 (MODERNA))	Lung opacity	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	1.01	2.42	5.11	0.328
COVID19 (COVID19 (PFIZER-BIONTECH))	Red blood cell count increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.869	2.41	5.62	0.437

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (PFIZER-BIONTECH))	Mean cell volume decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.863	2.40	5.58	0.434
COVID19 (COVID19 (MODERNA))	Scan with contrast	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.862	2.39	5.58	0.423
COVID19 (COVID19 (JANSSEN))	Computerised tomogram abdomen abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.993	2.39	5.04	0.218
COVID19 (COVID19 (MODERNA))	Computerised tomogram abdomen normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.859	2.38	5.55	0.518
COVID19 (COVID19 (JANSSEN))	Chest X-ray normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.988	2.38	5.02	0.217
COVID19 (COVID19 (JANSSEN))	Computerised tomogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	1.23	2.37	4.23	0.270
COVID19 (COVID19 (JANSSEN))	Electrocardiogram normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.983	2.37	4.99	0.216
COVID19 (COVID19 (PFIZER-BIONTECH))	Chest pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	14	1.52	2.36	3.54	0.762
COVID19 (COVID19 (JANSSEN))	Blood creatine phosphokinase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.971	2.34	4.93	0.213
COVID19 (COVID19 (MODERNA))	Computerised tomogram thorax	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	1.27	2.34	4.02	0.512
COVID19 (COVID19 (JANSSEN))	Cough	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.06	2.32	4.54	0.233
COVID19 (COVID19	TTP_fibrin D dimer_platelet factor 4	Troponin	4	1.06	2.31	4.54	0.542

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(MODERNA))	(Custom Term)						
COVID19 (COVID19 (PFIZER-BIONTECH))	Depressed level of consciousness	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.833	2.31	5.39	0.419
COVID19 (COVID19 (PFIZER-BIONTECH))	Monocyte percentage decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.833	2.31	5.39	0.419
COVID19 (COVID19 (MODERNA))	Presyncope	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.830	2.30	5.36	0.501
COVID19 (COVID19 (PFIZER-BIONTECH))	Mean cell haemoglobin decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.825	2.29	5.34	0.415
COVID19 (COVID19 (MODERNA))	SARS-CoV-2 test positive	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.04	2.28	4.47	0.499
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound Doppler	6	1.18	2.28	4.06	0.594
COVID19 (COVID19 (PFIZER-BIONTECH))	Electrocardiogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.03	2.25	4.42	0.507
COVID19 (COVID19 (PFIZER-BIONTECH))	Scan with contrast abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.808	2.24	5.22	0.281
COVID19 (COVID19 (PFIZER-BIONTECH))	Coagulopathy	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.808	2.24	5.22	0.398
COVID19 (COVID19 (JANSSEN))	Back pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.928	2.24	4.71	0.204
COVID19 (COVID19 (MODERNA))	Pulmonary infarction	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.799	2.22	5.17	0.180

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	White blood cell count increased	9	1.28	2.21	3.60	0.281
COVID19 (COVID19 (JANSSEN))	Acute kidney injury	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.913	2.20	4.64	0.201
COVID19 (COVID19 (PFIZER-BIONTECH))	Palpitations	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	1.00	2.20	4.32	0.505
COVID19 (COVID19 (MODERNA))	Resuscitation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.911	2.20	4.62	0.345
COVID19 (COVID19 (JANSSEN))	Lethargy	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.791	2.20	5.12	0.174
COVID19 (COVID19 (JANSSEN))	Hypotension	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.908	2.19	4.61	0.199
COVID19 (COVID19 (JANSSEN))	Chills	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	1.07	2.18	4.05	0.235
COVID19 (COVID19 (PFIZER-BIONTECH))	Protein total increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.780	2.16	5.04	0.392
COVID19 (COVID19 (PFIZER-BIONTECH))	Dyspnoea exertional	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	1.06	2.16	4.02	0.534
COVID19 (COVID19 (PFIZER-BIONTECH))	Cardiac arrest	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.770	2.14	4.98	0.239
COVID19 (COVID19 (PFIZER-BIONTECH))	Glomerular filtration rate decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	1.10	2.13	3.80	0.555
COVID19 (COVID19	Red blood cell count decreased	TTP_fibrin D dimer_platelet factor 4 (Custom	7	1.15	2.13	3.66	0.253

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(JANSSEN))		Term)					
COVID19 (COVID19 (JANSSEN))	Haemoglobin decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	11	1.29	2.11	3.31	0.283
COVID19 (COVID19 (JANSSEN))	Decreased appetite	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.761	2.11	4.92	0.167
COVID19 (COVID19 (MODERNA))	Anticoagulant therapy	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	1.03	2.09	3.89	0.489
COVID19 (COVID19 (JANSSEN))	Dyspnoea exertional	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.745	2.07	4.82	0.164
COVID19 (COVID19 (PFIZER-BIONTECH))	Hypoxia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	1.07	2.06	3.67	0.444
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Tachypnoea	5	1.01	2.06	3.82	0.609
COVID19 (COVID19 (PFIZER-BIONTECH))	Painful respiration	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.853	2.05	4.33	0.429
COVID19 (COVID19 (MODERNA))	Chest X-ray normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	1.06	2.04	3.64	0.638
COVID19 (COVID19 (MODERNA))	Cardio-respiratory arrest	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.735	2.04	4.75	0.260
COVID19 (COVID19 (MODERNA))	Loss of consciousness	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.732	2.03	4.74	0.442
COVID19 (COVID19 (PFIZER-BIONTECH))	Fibrin D dimer increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	39	1.54	2.00	2.58	0.772
COVID19		TTP_fibrin D					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(COVID19 (JANSSEN))	Blood creatinine increased	dimer_platelet factor 4 (Custom Term)	4	0.916	2.00	3.93	0.201
COVID19 (COVID19 (PFIZER-BIONTECH))	Bilirubin conjugated increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.720	2.00	4.65	0.362
COVID19 (COVID19 (MODERNA))	Acute respiratory distress syndrome	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.719	2.00	4.65	0.434
COVID19 (COVID19 (PFIZER-BIONTECH))	Pulmonary embolism	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	33	1.49	1.99	2.62	0.586
COVID19 (COVID19 (JANSSEN))	Areflexia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.718	1.99	4.64	0.158
COVID19 (COVID19 (JANSSEN))	Blood chloride increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.825	1.99	4.19	0.181
COVID19 (COVID19 (PFIZER-BIONTECH))	Brain oedema	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.712	1.98	4.61	0.358
COVID19 (COVID19 (PFIZER-BIONTECH))	Chest discomfort	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.970	1.98	3.68	0.488
COVID19 (COVID19 (JANSSEN))	Chest pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	1.02	1.98	3.52	0.225
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Wheezing	3	0.820	1.98	4.16	0.495
COVID19 (COVID19 (PFIZER-BIONTECH))	Angiogram pulmonary abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	10	1.16	1.96	3.13	0.454
COVID19 (COVID19 (MODERNA))	Seizure	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.812	1.96	4.12	0.490

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (JANSSEN))	Pyrexia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	9	1.13	1.95	3.19	0.249
COVID19 (COVID19 (JANSSEN))	Muscular weakness	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.703	1.95	4.55	0.154
COVID19 (COVID19 (PFIZER-BIONTECH))	Oedema peripheral	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.806	1.94	4.09	0.405
COVID19 (COVID19 (MODERNA))	Metabolic acidosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.699	1.94	4.52	0.422
COVID19 (COVID19 (PFIZER-BIONTECH))	Pleural effusion	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.946	1.93	3.59	0.476
COVID19 (COVID19 (JANSSEN))	Arthralgia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.798	1.92	4.05	0.175
COVID19 (COVID19 (MODERNA))	Pulmonary embolism	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	35	1.44	1.91	2.49	0.544
COVID19 (COVID19 (MODERNA))	Angiogram pulmonary abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	11	1.16	1.91	2.99	0.600
COVID19 (COVID19 (MODERNA))	Blood test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.987	1.90	3.39	0.595
COVID19 (COVID19 (JANSSEN))	Red blood cells urine positive	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.685	1.90	4.43	0.151
COVID19 (COVID19 (JANSSEN))	Neutrophil count increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.684	1.90	4.43	0.150
COVID19 (COVID19		TTP_fibrin D dimer_platelet					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(PFIZER-BIONTECH))	Pallor	factor 4 (Custom Term)	2	0.684	1.90	4.42	0.344
COVID19 (COVID19 (PFIZER-BIONTECH))	Computerised tomogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.980	1.89	3.37	0.492
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Thrombectomy	3	0.778	1.87	3.95	0.276
COVID19 (COVID19 (MODERNA))	Haemoglobin normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.676	1.87	4.37	0.408
COVID19 (COVID19 (JANSSEN))	Dyspnoea	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	12	1.16	1.87	2.88	0.255
COVID19 (COVID19 (JANSSEN))	Aspartate aminotransferase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	1.01	1.86	3.20	0.221
COVID19 (COVID19 (PFIZER-BIONTECH))	Metabolic function test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.770	1.85	3.91	0.387
COVID19 (COVID19 (MODERNA))	Cardiac arrest	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.664	1.84	4.29	0.255
COVID19 (COVID19 (JANSSEN))	SARS-CoV-2 test negative	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.763	1.84	3.87	0.168
COVID19 (COVID19 (PFIZER-BIONTECH))	Lung infiltration	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.661	1.84	4.28	0.332
COVID19 (COVID19 (JANSSEN))	Paraesthesia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.660	1.83	4.27	0.145
COVID19 (COVID19 (PFIZER-BIONTECH))	Respiratory distress	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.648	1.80	4.19	0.326

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (PFIZER-BIONTECH))	Dyspnoea	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	23	1.26	1.79	2.47	0.636
COVID19 (COVID19 (PFIZER-BIONTECH))	Angiogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.741	1.78	3.76	0.307
COVID19 (COVID19 (PFIZER-BIONTECH))	Cerebral haemorrhage	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.740	1.78	3.75	0.269
COVID19 (COVID19 (PFIZER-BIONTECH))	International normalised ratio normal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.641	1.78	4.15	0.322
COVID19 (COVID19 (PFIZER-BIONTECH))	Mean cell haemoglobin concentration decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.641	1.78	4.14	0.322
COVID19 (COVID19 (MODERNA))	General physical health deterioration	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.640	1.78	4.14	0.386
COVID19 (COVID19 (PFIZER-BIONTECH))	Lymphocyte count decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.807	1.76	3.46	0.406
COVID19 (COVID19 (MODERNA))	COVID-19	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.732	1.76	3.72	0.324
COVID19 (COVID19 (PFIZER-BIONTECH))	Haematocrit increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.635	1.76	4.11	0.319
COVID19 (COVID19 (PFIZER-BIONTECH))	Serum ferritin increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.633	1.76	4.09	0.318
COVID19 (COVID19 (JANSSEN))	Hypertension	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.629	1.75	4.07	0.138
COVID19		TTP_fibrin D dimer_platelet					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(COVID19 (JANSSEN))	Myalgia	factor 4 (Custom Term)	2	0.617	1.71	3.99	0.136
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Walking aid user	2	0.610	1.69	3.94	0.368
COVID19 (COVID19 (PFIZER-BIONTECH))	Anion gap decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.608	1.69	3.93	0.306
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Thrombectomy	2	0.607	1.68	3.93	0.166
COVID19 (COVID19 (PFIZER-BIONTECH))	Condition aggravated	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.767	1.68	3.29	0.385
COVID19 (COVID19 (PFIZER-BIONTECH))	Cough	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.869	1.68	2.99	0.437
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Thrombosis	11	1.02	1.67	2.62	0.512
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Wheezing	2	0.603	1.67	3.90	0.303
COVID19 (COVID19 (MODERNA))	Heart rate increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.694	1.67	3.52	0.419
COVID19 (COVID19 (MODERNA))	Angiogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.599	1.66	3.87	0.361
COVID19 (COVID19 (JANSSEN))	Blood urea increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.759	1.66	3.26	0.167
COVID19 (COVID19 (PFIZER-BIONTECH))	Injection site pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.598	1.66	3.87	0.301

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Thrombosis	12	1.01	1.63	2.51	0.610
COVID19 (COVID19 (MODERNA))	Dyspnoea exertional	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.744	1.63	3.19	0.449
COVID19 (COVID19 (PFIZER-BIONTECH))	Haematuria	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.584	1.62	3.78	0.294
COVID19 (COVID19 (JANSSEN))	Blood albumin decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.740	1.62	3.18	0.163
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Troponin	2	0.582	1.62	3.76	0.222
COVID19 (COVID19 (MODERNA))	Electroencephalogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.579	1.61	3.74	0.349
COVID19 (COVID19 (JANSSEN))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Tachycardia	4	0.732	1.60	3.14	0.161
COVID19 (COVID19 (MODERNA))	Full blood count	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.664	1.60	3.37	0.401
COVID19 (COVID19 (PFIZER-BIONTECH))	Brain natriuretic peptide increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.726	1.59	3.12	0.365
COVID19 (COVID19 (JANSSEN))	Blood potassium decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.656	1.58	3.33	0.144
COVID19 (COVID19 (PFIZER-BIONTECH))	Computerised tomogram thorax abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	10	0.930	1.56	2.50	0.468
COVID19 (COVID19 (JANSSEN))	Blood calcium decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.761	1.55	2.88	0.167

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Troponin increased	3	0.642	1.55	3.26	0.272
COVID19 (COVID19 (PFIZER-BIONTECH))	Influenza virus test negative	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.557	1.54	3.60	0.280
COVID19 (COVID19 (JANSSEN))	Fatigue	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.757	1.54	2.87	0.166
COVID19 (COVID19 (MODERNA))	Angiogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.640	1.54	3.25	0.349
COVID19 (COVID19 (JANSSEN))	Alanine aminotransferase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.751	1.53	2.84	0.165
COVID19 (COVID19 (MODERNA))	Computerised tomogram thorax abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	11	0.931	1.53	2.40	0.562
COVID19 (COVID19 (PFIZER-BIONTECH))	Hyperhidrosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.549	1.52	3.55	0.276
COVID19 (COVID19 (MODERNA))	Chest X-ray	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.628	1.51	3.19	0.379
COVID19 (COVID19 (MODERNA))	Fibrin D dimer increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	32	1.11	1.49	1.97	0.672
COVID19 (COVID19 (PFIZER-BIONTECH))	Computerised tomogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	8	0.839	1.49	2.50	0.422
COVID19 (COVID19 (PFIZER-BIONTECH))	Computerised tomogram head abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.733	1.49	2.78	0.368
COVID19 (COVID19 (PFIZER-	Blood culture negative	TTP_fibrin D dimer_platelet factor 4 (Custom	3	0.619	1.49	3.14	0.311

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
BIONTECH))		Term)					
COVID19 (COVID19 (PFIZER-BIONTECH))	Back pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.679	1.48	2.92	0.341
COVID19 (COVID19 (MODERNA))	Atelectasis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.604	1.45	3.06	0.364
COVID19 (COVID19 (MODERNA))	Dyspnoea	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	20	0.999	1.45	2.04	0.603
COVID19 (COVID19 (JANSSEN))	Chest X-ray abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.661	1.44	2.84	0.145
COVID19 (COVID19 (PFIZER-BIONTECH))	Respiratory failure	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.707	1.44	2.68	0.356
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound scan abnormal	5	0.705	1.44	2.67	0.355
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood glucose increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	10	0.844	1.42	2.26	0.424
COVID19 (COVID19 (PFIZER-BIONTECH))	Death	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.733	1.41	2.52	0.277
COVID19 (COVID19 (PFIZER-BIONTECH))	Deep vein thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	12	0.865	1.39	2.15	0.408
COVID19 (COVID19 (JANSSEN))	Protein total decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.500	1.39	3.24	0.110
COVID19 (COVID19 (JANSSEN))	Haematocrit decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.708	1.36	2.43	0.155

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood lactate dehydrogenase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.489	1.36	3.16	0.246
COVID19 (COVID19 (JANSSEN))	Echocardiogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.489	1.36	3.16	0.107
COVID19 (COVID19 (PFIZER-BIONTECH))	Acute kidney injury	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.665	1.36	2.52	0.334
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound Doppler normal	2	0.486	1.35	3.14	0.244
COVID19 (COVID19 (JANSSEN))	Red cell distribution width increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.482	1.34	3.12	0.106
COVID19 (COVID19 (PFIZER-BIONTECH))	Full blood count	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.482	1.34	3.12	0.242
COVID19 (COVID19 (MODERNA))	Chest pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	8	0.738	1.31	2.20	0.445
COVID19 (COVID19 (PFIZER-BIONTECH))	Carbon dioxide decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.544	1.31	2.76	0.274
COVID19 (COVID19 (PFIZER-BIONTECH))	Red cell distribution width increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.679	1.31	2.33	0.341
COVID19 (COVID19 (JANSSEN))	Blood glucose increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.593	1.30	2.54	0.130
COVID19 (COVID19 (JANSSEN))	Malaise	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.464	1.29	3.00	0.102
COVID19 (COVID19		TTP_fibrin D dimer_platelet					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(PFIZER-BIONTECH))	Laboratory test	factor 4 (Custom Term)	2	0.463	1.29	3.00	0.233
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Transfusion	2	0.463	1.29	3.00	0.233
COVID19 (COVID19 (PFIZER-BIONTECH))	Syncope	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.462	1.28	2.99	0.232
COVID19 (COVID19 (PFIZER-BIONTECH))	Prothrombin time prolonged	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.659	1.27	2.27	0.331
COVID19 (COVID19 (MODERNA))	Hypoxia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.578	1.26	2.48	0.304
COVID19 (COVID19 (MODERNA))	Metabolic function test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.455	1.26	2.94	0.275
COVID19 (COVID19 (MODERNA))	Death	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.655	1.26	2.25	0.241
COVID19 (COVID19 (PFIZER-BIONTECH))	Acute respiratory failure	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.524	1.26	2.66	0.263
COVID19 (COVID19 (PFIZER-BIONTECH))	Confusional state	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.449	1.25	2.90	0.226
COVID19 (COVID19 (PFIZER-BIONTECH))	C-reactive protein increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.570	1.25	2.45	0.286
COVID19 (COVID19 (MODERNA))	Electrocardiogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.444	1.23	2.87	0.260
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Tachycardia	6	0.634	1.22	2.18	0.297

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (PFIZER-BIONTECH))	Chest X-ray abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	9	0.709	1.22	2.00	0.356
COVID19 (COVID19 (PFIZER-BIONTECH))	Red blood cell count decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	9	0.705	1.22	1.99	0.355
COVID19 (COVID19 (PFIZER-BIONTECH))	Chills	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.597	1.22	2.26	0.300
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.504	1.21	2.56	0.253
COVID19 (COVID19 (PFIZER-BIONTECH))	Activated partial thromboplastin time prolonged	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.501	1.21	2.54	0.252
COVID19 (COVID19 (PFIZER-BIONTECH))	Neutrophil count increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.500	1.20	2.54	0.251
COVID19 (COVID19 (MODERNA))	Computerised tomogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	0.644	1.19	2.05	0.389
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound Doppler normal	2	0.426	1.18	2.76	0.257
COVID19 (COVID19 (JANSSEN))	Lymphocyte percentage decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.489	1.18	2.48	0.107
COVID19 (COVID19 (MODERNA))	Syncope	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.424	1.18	2.74	0.256
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood alkaline phosphatase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.422	1.17	2.73	0.212
COVID19 (COVID19 (PFIZER-	TTP_fibrin D dimer_platelet factor 4	Tachypnoea	2	0.419	1.16	2.71	0.211

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
BIONTECH))	(Custom Term)						
COVID19 (COVID19 (MODERNA))	Fall	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.418	1.16	2.70	0.252
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound scan	2	0.417	1.16	2.70	0.252
COVID19 (COVID19 (MODERNA))	Myalgia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.479	1.15	2.43	0.289
COVID19 (COVID19 (MODERNA))	Laboratory test	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.416	1.15	2.69	0.251
COVID19 (COVID19 (MODERNA))	Dizziness	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.475	1.14	2.41	0.287
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound Doppler	3	0.473	1.14	2.40	0.286
COVID19 (COVID19 (PFIZER-BIONTECH))	Malaise	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.519	1.14	2.23	0.261
COVID19 (COVID19 (PFIZER-BIONTECH))	Pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.581	1.12	2.00	0.292
COVID19 (COVID19 (MODERNA))	Sinus tachycardia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.402	1.12	2.60	0.243
COVID19 (COVID19 (PFIZER-BIONTECH))	Pyrexia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	9	0.645	1.11	1.82	0.324
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood lactic acid increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.461	1.11	2.34	0.232
COVID19 (COVID19	Diarrhoea	TTP_fibrin D dimer_platelet	3	0.460	1.11	2.33	0.277

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(MODERNA))		factor 4 (Custom Term)					
COVID19 (COVID19 (PFIZER-BIONTECH))	Lethargy	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.397	1.10	2.57	0.200
COVID19 (COVID19 (PFIZER-BIONTECH))	Echocardiogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.502	1.10	2.16	0.253
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Thrombocytopenia	6	0.565	1.09	1.94	0.284
COVID19 (COVID19 (MODERNA))	Pneumonia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.497	1.09	2.13	0.300
COVID19 (COVID19 (PFIZER-BIONTECH))	Pain in extremity	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.526	1.07	2.00	0.265
COVID19 (COVID19 (JANSSEN))	Asthenia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.374	1.04	2.43	0.082
COVID19 (COVID19 (MODERNA))	Blood urine present	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.372	1.03	2.40	0.224
COVID19 (COVID19 (PFIZER-BIONTECH))	Decreased appetite	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.367	1.02	2.38	0.185
COVID19 (COVID19 (MODERNA))	Mechanical ventilation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.413	0.994	2.09	0.249
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood creatinine increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.487	0.993	1.85	0.245
COVID19 (COVID19 (PFIZER-BIONTECH))	Anaemia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.352	0.978	2.28	0.177

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood potassium decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.446	0.975	1.91	0.224
COVID19 (COVID19 (MODERNA))	Computerised tomogram	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.404	0.974	2.05	0.244
COVID19 (COVID19 (PFIZER-BIONTECH))	Nausea	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.475	0.968	1.80	0.239
COVID19 (COVID19 (MODERNA))	Cerebral venous sinus thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.344	0.955	2.23	0.208
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood urea increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.495	0.954	1.70	0.249
COVID19 (COVID19 (MODERNA))	Hypotension	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.395	0.952	2.01	0.239
COVID19 (COVID19 (PFIZER-BIONTECH))	Myalgia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.342	0.949	2.21	0.172
COVID19 (COVID19 (MODERNA))	Pyrexia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	8	0.525	0.935	1.56	0.317
COVID19 (COVID19 (MODERNA))	Chills	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.427	0.933	1.83	0.257
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound Doppler abnormal	5	0.457	0.930	1.73	0.276
COVID19 (COVID19 (MODERNA))	Intensive care	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	0.496	0.916	1.58	0.299
COVID19 (COVID19	Electrocardiogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom	6	0.474	0.914	1.63	0.286

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(MODERNA))		Term)					
COVID19 (COVID19 (MODERNA))	Sepsis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.328	0.911	2.12	0.198
COVID19 (COVID19 (PFIZER-BIONTECH))	Computerised tomogram abdomen abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.327	0.907	2.12	0.164
COVID19 (COVID19 (MODERNA))	Chest discomfort	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.327	0.906	2.11	0.197
COVID19 (COVID19 (MODERNA))	Condition aggravated	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.326	0.904	2.11	0.197
COVID19 (COVID19 (PFIZER-BIONTECH))	Hypertension	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.326	0.904	2.11	0.162
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood albumin decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.443	0.902	1.68	0.223
COVID19 (COVID19 (PFIZER-BIONTECH))	Haemoglobin decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	10	0.533	0.895	1.43	0.268
COVID19 (COVID19 (PFIZER-BIONTECH))	Endotracheal intubation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.409	0.894	1.75	0.205
COVID19 (COVID19 (MODERNA))	Cough	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.366	0.882	1.86	0.221
COVID19 (COVID19 (JANSSEN))	Electrocardiogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.316	0.879	2.05	0.069
COVID19 (COVID19 (PFIZER-BIONTECH))	Electrocardiogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.430	0.875	1.63	0.216

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Ultrasound Doppler abnormal	4	0.397	0.868	1.70	0.200
COVID19 (COVID19 (MODERNA))	Chest X-ray abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	7	0.468	0.863	1.49	0.282
COVID19 (COVID19 (PFIZER-BIONTECH))	Fatigue	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.420	0.855	1.59	0.211
COVID19 (COVID19 (MODERNA))	Pleural effusion	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.308	0.854	1.99	0.186
COVID19 (COVID19 (PFIZER-BIONTECH))	Magnetic resonance imaging head abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.354	0.852	1.80	0.178
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood sodium decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.353	0.850	1.79	0.177
COVID19 (COVID19 (PFIZER-BIONTECH))	Abdominal pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.305	0.846	1.97	0.153
COVID19 (COVID19 (PFIZER-BIONTECH))	Haematocrit decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	8	0.466	0.829	1.39	0.234
COVID19 (COVID19 (MODERNA))	Acute respiratory failure	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.295	0.818	1.91	0.178
COVID19 (COVID19 (MODERNA))	Blood glucose increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.420	0.810	1.44	0.254
COVID19 (COVID19 (MODERNA))	Computerised tomogram abdomen abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.292	0.810	1.89	0.176
COVID19 (COVID19	TTP_fibrin D dimer_platelet factor 4	Tachycardia	4	0.368	0.804	1.58	0.222

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(MODERNA))	(Custom Term)						
COVID19 (COVID19 (MODERNA))	Hypertension	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.289	0.803	1.87	0.175
COVID19 (COVID19 (MODERNA))	Endotracheal intubation	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.364	0.796	1.56	0.220
COVID19 (COVID19 (MODERNA))	Fatigue	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.391	0.795	1.48	0.236
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood gases abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.285	0.790	1.84	0.143
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Vomiting	3	0.327	0.788	1.66	0.197
COVID19 (COVID19 (MODERNA))	Abdominal pain	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.282	0.782	1.82	0.170
COVID19 (COVID19 (PFIZER-BIONTECH))	Aspartate aminotransferase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.384	0.782	1.46	0.193
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood bilirubin increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.280	0.778	1.81	0.141
COVID19 (COVID19 (PFIZER-BIONTECH))	Platelet count decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	10	0.460	0.773	1.24	0.231
COVID19 (COVID19 (PFIZER-BIONTECH))	Intensive care	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	5	0.372	0.758	1.41	0.187
COVID19 (COVID19 (MODERNA))	Nausea	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.341	0.746	1.46	0.206
COVID19		TTP_fibrin D					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(COVID19 (MODERNA))	Pain	dimer_platelet factor 4 (Custom Term)	4	0.338	0.739	1.45	0.204
COVID19 (COVID19 (PFIZER-BIONTECH))	Lymphocyte percentage decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.336	0.736	1.44	0.169
COVID19 (COVID19 (PFIZER-BIONTECH))	Pneumonia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.265	0.734	1.71	0.133
COVID19 (COVID19 (MODERNA))	Arthralgia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.255	0.707	1.65	0.154
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood chloride increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.253	0.704	1.64	0.127
COVID19 (COVID19 (MODERNA))	Blood potassium decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.291	0.701	1.48	0.176
COVID19 (COVID19 (PFIZER-BIONTECH))	Asthenia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.285	0.688	1.45	0.143
COVID19 (COVID19 (MODERNA))	Deep vein thrombosis	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	6	0.352	0.679	1.21	0.202
COVID19 (COVID19 (MODERNA))	C-reactive protein increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.245	0.679	1.58	0.148
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	White blood cell count increased	5	0.333	0.678	1.26	0.167
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	White blood cell count decreased	2	0.239	0.663	1.54	0.120
COVID19 (COVID19 (PFIZER-	Neutrophil percentage increased	TTP_fibrin D dimer_platelet factor 4 (Custom	3	0.275	0.662	1.40	0.138

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
BIONTECH))		Term)					
COVID19 (COVID19 (MODERNA))	Pain in extremity	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.269	0.647	1.36	0.162
COVID19 (COVID19 (PFIZER-BIONTECH))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Vomiting	2	0.232	0.644	1.50	0.117
COVID19 (COVID19 (PFIZER-BIONTECH))	Headache	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.289	0.633	1.24	0.145
COVID19 (COVID19 (MODERNA))	Platelet count decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	9	0.366	0.632	1.03	0.221
COVID19 (COVID19 (PFIZER-BIONTECH))	Blood calcium decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.285	0.624	1.22	0.143
COVID19 (COVID19 (MODERNA))	Malaise	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.224	0.622	1.45	0.135
COVID19 (COVID19 (MODERNA))	Asthenia	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.257	0.620	1.31	0.155
COVID19 (COVID19 (PFIZER-BIONTECH))	Alanine aminotransferase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.253	0.611	1.29	0.127
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	White blood cell count decreased	2	0.215	0.598	1.39	0.130
COVID19 (COVID19 (MODERNA))	Headache	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.271	0.592	1.16	0.163
COVID19 (COVID19 (MODERNA))	Echocardiogram abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.211	0.586	1.37	0.127
COVID19		TTP_fibrin D					

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
(COVID19 (MODERNA))	Blood creatinine increased	dimer_platelet factor 4 (Custom Term)	3	0.241	0.581	1.22	0.145
COVID19 (COVID19 (MODERNA))	Magnetic resonance imaging head abnormal	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.209	0.580	1.35	0.126
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	Thrombocytopenia	3	0.236	0.569	1.20	0.142
COVID19 (COVID19 (MODERNA))	Alanine aminotransferase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.235	0.566	1.19	0.142
COVID19 (COVID19 (MODERNA))	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	White blood cell count increased	4	0.235	0.515	1.01	0.142
COVID19 (COVID19 (MODERNA))	Red cell distribution width increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.179	0.496	1.16	0.108
COVID19 (COVID19 (PFIZER-BIONTECH))	International normalised ratio increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.174	0.483	1.12	0.087
COVID19 (COVID19 (MODERNA))	Neutrophil percentage increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.161	0.446	1.04	0.097
COVID19 (COVID19 (MODERNA))	International normalised ratio increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.153	0.424	0.988	0.092
COVID19 (COVID19 (MODERNA))	Lymphocyte percentage decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.143	0.397	0.925	0.086
COVID19 (COVID19 (MODERNA))	Haemoglobin decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	4	0.167	0.366	0.719	0.101
COVID19 (COVID19 (MODERNA))	Aspartate aminotransferase increased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.130	0.360	0.840	0.078

Vaccine Name	Symptom: PT1	Symptom: PT2	N	EB05	EBGM	EB95	INTSS
COVID19 (COVID19 (MODERNA))	Haematocrit decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	3	0.138	0.332	0.701	0.083
COVID19 (COVID19 (MODERNA))	Red blood cell count decreased	TTP_fibrin D dimer_platelet factor 4 (Custom Term)	2	0.118	0.328	0.764	0.071

These data do not, by themselves, demonstrate causal associations; they may serve as a signal for further investigation.

Message

From: Menschik, David [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=0407D7354456470CAB9BC2D3F98D6D3C-MENSCHIK]
Sent: 3/18/2021 8:33:35 PM
To: Nguon, Kosal * [redacted]
Subject: RE: special project run

Thanks Kosal!!

From: Nguon, Kosal * <[redacted]>
Sent: Thursday, March 18, 2021 4:30 PM
To: Menschik, David <[redacted]>
Subject: RE: special project run

Hi David,

The updated "SP: SCV" run (ID 30822) has now been completed with an as of date of March 11, 2020. Additionally, I changed the custom name to "COVID mRNA Vaccines All" and edited the description (removed "Unknown" covid vaccine manufacturers as that was not part of the original custom term and query anyway).

I'll get started on the other request regarding excising out all of the COVID reports minus the ones of interest.

Enjoy your time off—thanks!

Best,

Kosal

From: Menschik, David <[redacted]>
Sent: Thursday, March 18, 2021 2:58 PM
To: Nguon, Kosal * [redacted]>
Subject: RE: special project run

Sure – thanks! I'll send an outlook invite…

From: Nguon, Kosal * <[redacted]>
Sent: Thursday, March 18, 2021 2:57 PM
To: Menschik, David <[redacted]>
Subject: RE: special project run

Hi David,

I have a meeting that should end early at 330PM today. Does that work for you? Thanks.

Best,

Kosal

From: Menschik, David <[redacted]>
Sent: Thursday, March 18, 2021 2:54 PM
To: Nguon, Kosal * <[redacted]>
Subject: RE: special project run

Hi Kosal,

I think it would be best if we touched base verbally – what's your availability like?

Thanks,
David

From: Nguon, Kosal * <[redacted]>
Sent: Thursday, March 18, 2021 1:12 PM
To: Menschik, David <[redacted]>
Subject: RE: special project run

Hi David,

I've been thinking deeply about this as I have been working with a few different folks with similar but important differences, and I want to make sure I understand the result/outcome concretely, and we can map out the methodology/custom run. Here's what I have and reasoning:

The concern is that COVID case reports contribute a significant enough portion to the VAERS database, where it may affect the EBGM scores of other vaccines/products-event combinations. In order to accurately calculate an EBGM score for all other product-event combinations, what is the best way to do that (i.e., pre-COVID reports)?

Technically, a report is only counted once; however, a report may have more than one product-event combination (e.g., a person got the COVID vaccine and takes aspirin, and experiences a rash, fever and upset stomach; 2 drugs times 3 events equals 6 drug-event combinations). If you remove COVID, you only remove the COVID-event combinations, but the aspirin-event combinations exist. I believe the VAERS data only has identified vaccines, so the aspirin-related events would not show up. However, what's the likelihood of someone getting a COVID vaccine and another vaccine in a relatively close enough time period and reporting them both?

I think you'll be ok removing COVID vaccines, which will in turn eliminate all the reports assuming there are no other vaccines. This makes it easier than identifying COVID reports by CASE ID and then removing them that way. Let me know if that makes sense as you followed though some of my stream-of-conscious thinking here. Thanks!

Best,

Kosal

From: Menschik, David <[redacted]>
Sent: Thursday, March 18, 2021 11:54 AM
To: Nguon, Kosal * <[redacted]>
Subject: RE: special project run

Thanks and that explanation is helpful. Accordingly, would it make sense to restrict the background so that each actual vaccine report is only counted once? (i.e., remove the custom drug term from the denominator/comparator for this run)

Thanks,
David

From: Nguon, Kosal * <[redacted]>
Sent: Thursday, March 18, 2021 11:46 AM
To: Menschik, David <[redacted]>
Subject: RE: special project run

Hi David,

I'll answer both your questions in this email to keep it all together.

1. I can certainly change the custom drug name to "COVID mRNA Vaccine."

2. Regarding the custom term and comparator groups, I am not exactly sure what you are asking, but I think I do, so I'll try my best to provide an explanation. For the COVID mRNA vaccines like Moderna and Pfizer, if you combine them to make a custom term for "mRNA Covid Vaccines" as we have done. So when it comes to background/denominator, MGPS will use it as:
[custom drug term} + A + B + C + ...+ N, where A, B, C are different specified "vaccines"

I believe the slight changes you see are only with EBGM scores, and these due to the custom terms and how the "definition" of a vaccine or PT has altered.

Please let me know if that second part makes sense, and if not, I am happy to talk through it. Thanks!

Best,

Kosal

From: Menschik, David <[redacted]>
Sent: Thursday, March 18, 2021 10:39 AM
To: Nguon, Kosal * <[redacted]>
Subject: RE: special project run

Hi Kosal,

Also wondering: how is this custom term built in to the denominator/comparator group for MGPS algorithm or is the denominator/comparator group unchanged? (wondering because for the [non-customized] component vaccines, I see slight differences in EBGMs/EB05s relative to using the [non-special project] comparable 'US VAERS vac name run')

Thanks,
David

From: Menschik, David
Sent: Thursday, March 18, 2021 10:34 AM
To: Nguon, Kosal * <[redacted]>
Subject: RE: special project run

Thanks Kosal! As an aside, I noticed in the last special project called "SP: SCV" that the custom term is called "COVID Vaccine All" though it only refers to the mRNA vaccines (Pfizer and Moderna) based on definition in notes section. Accordingly, can you change the custom term from "COVID Vaccine All" to "COVID mRNA Vaccine" ?

Thanks,
David

From: Nguon, Kosal * <[redacted]
Sent: Thursday, March 18, 2021 10:07 AM
To: Menschik, David <[redacted]
Cc: Baer, Bethany <[redacted]
Subject: RE: special project run

Hi David,

Taking Brian off.

Not a problem and looking forward to it.

Best,

Kosal

From: Menschik, David <[redacted]
Sent: Thursday, March 18, 2021 7:08 AM
To: Nguon, Kosal * <[redacted]
Cc: Baer, Bethany <[redacted]; Hendrix, Brian * <[redacted]
Subject: RE: special project run

Thanks very much Kosal. I'll plan to circle back to you on this later in the day.

Thanks,
David

From: Nguon, Kosal * <[redacted]
Sent: Wednesday, March 17, 2021 5:46 PM
To: Menschik, David <[redacted]; Hendrix, Brian * <[redacted]
Cc: Baer, Bethany <[redacted]
Subject: RE: special project run

Hi David,

Not a problem, and I can definitely help with this request and Brian can focus on the VSafe and remaining runs.

Just a few questions and comments:
1. From your statement, " [a] run that excludes from the comparison group all COVID vaccine reports except those involving the COVID vaccine report in the numerator (e.g., Pfizer vaccine in this example)." I think we'd have to create a special data mining run for each COVID vaccine (e.g., analogous run with Pfizer, analogous run with Moderna, analogous run with Jansen/J&J). Is that ok for you?

2. Do you want to create one for the Jansen/J&J now or wait until it reaches a certain threshold? We can create it, but it's not going to be useful as there's not much data for it.
3. How do you want to approach vaccines marked as "Unknown manufacturer"? Create its own group? Exclude altogether? Or combine with another COVID vaccine?

Thanks for thinking about this, and I can get started on it.

Best,

Kosal

From: Menschik, David <[redacted]>
Sent: Wednesday, March 17, 2021 3:41 PM
To: Hendrix, Brian * <[redacted]>; Nguon, Kosal * <[redacted]>
Cc: Baer, Bethany <[redacted]>
Subject: RE: special project run

Apologies for neglecting to include Kosal in my email below – this may be more in his lane (understand he has created 'special project' runs in the past)
Best,
David

From: Menschik, David
Sent: Wednesday, March 17, 2021 2:58 PM
To: Hendrix, Brian * <[redacted]>
Cc: Baer, Bethany <[redacted]>
Subject: special project run

Hi Brian,

I was exploring the time trend graph in the context of theoretical muting effect of so many COVID vaccine reports in the denominator/comparator and found that the first vaccine authorized under EUA (12/10) graph (attached) shows a muting trend for PTs. Consequently, I think it would be valuable to establish as a 'special project' run (visible only to us), an analogous (using "US VAERS Vac Name Monthly Cumulative") run that excludes from the comparison group all COVID vaccine reports except those involving the COVID vaccine report in the numerator (e.g., Pfizer vaccine in this example). Can you please advise?

Thanks,
David

-----Original Appointment-----
From: Menschik, David
Sent: Tuesday, December 01, 2020 10:16 AM
To: Menschik, David; Nguon, Kosal *; Sydnor, James *; Casey Sydnor; Lebow, William *; Hendrix, Brian *; Baer, Bethany
Subject: Bi-weekly touching base on VAERS Empirica updates in support of SC2V surveillance
When: Wednesday, March 17, 2021 2:00 PM-2:30 PM (UTC-05:00) Eastern Time (US & Canada).
Where: WebEx

Pushing back this one meeting 30 minutes to accommodate scheduling conflict. Please advise if this does not work.
Thanks!
Also removing Brendan (as discussed)

This is intended as a placeholder in case there is anything for Commonwealth and FDA to discuss regarding the VAERS Empirica update projects in support of SC2V surveillance. (If not, plan to cancel)

-- Do not delete or change any of the following text. --

When it's time, join your Webex meeting here.

Join meeting

More ways to join:

Join from the meeting link
[redacted]

Join by meeting number
Meeting number (access code): [redacted]
Meeting password: [redacted]

Tap to join from a mobile device (attendees only)
[redacted]

Join by phone
[redacted]

If you are a host, click here to view host information.

Need help? Go to https://help.webex.com

Message

From: Menschik, David [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=0407D7354456470CAB9BC2D3F98D6D3C-MENSCHIK]
Sent: 5/9/2021 9:42:59 AM
To: Zinderman, Craig E [redacted]
Subject: RE: Data Mining feedback

Thanks Craig

From: Zinderman, Craig E <redacted>
Sent: Friday, May 07, 2021 5:56 PM
To: Menschik, David <redacted>; Niu, Manette <redacted>
Subject: FW: Data Mining feedback

Fyi..

From: Szarfman, Ana <redacted>
Sent: Friday, May 07, 2021 5:09 PM
To: Zinderman, Craig E <redacted>
Cc: Nair, Narayan <redacted>; Stockbridge, Norman L <redacted>
Subject: RE: Data Mining feedback

Hi Craig,

Thanks for your email. I understand your tremendous workload and the fantastic work you are all doing ,that I tremendously respect.

I will only deliver analyses when I am specifically requested to do so, and only to the reviewer making such requests.

We are testing a new data mining methodology, and given the circumstances, it will be good for all to understand its performance with such important data. This is a method that also strongly reduces confounding, so it may be helpful in certain future circumstances.

Let me know if I misunderstood anything in your email.

Warmest regards,

--Ana

Ana Szarfman, MD, PhD, FAMIA,
Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), and Fellow of the American Medical Informatics Association (2020)
Medical Officer, Safety Data Mining Developer and Medical Informatics Analyst,
Celebrating nearly a quarter of a century of successful implementation of safety data mining, interactive patient profiles, and other automated analytical tools.
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted]
(office)
[redacted]

U.S. FOOD & DRUG ADMINISTRATION

From: Zinderman, Craig E <[redacted]>
Sent: Friday, May 7, 2021 4:25 PM
To: Szarfman, Ana <[redacted]>
Cc: Nair, Narayan <[redacted]>
Subject: Data Mining feedback

Good Afternoon Ana,

Thank you again for talking with us back in March about your work exploring new data mining approaches and discussing your interest in CBER's COVID-19 vaccine products. We are writing to kindly ask you to please hold off on creating and sending data mining reports and analyses using COVID-19 vaccine AE data.

In CBER OBE, we have been reviewing the various COVID-19 vaccine data mining results that you have been forwarding. While we appreciate your interest in sharing your results, they haven't contributed to the already robust process for reviewing VAERS (and other vaccine safety) data. I will describe a little bit below:
-AESIs: CBER and CDC have established sets of AESIs (sets of PTs representing Adverse Events of Special Interest) for key events. Incoming reports captured by these AESIs are highlighted for FDA reviewer screening, as well as medical record follow-up and chart abstraction by CDC reviewers. Many of the alerts that you have been sending relate to AESIs for which we are already screening and reviewing reports, such as AMI, TTS, Thromboembolic events, and other forms of coagulopathy. Having our staff examine your alerts creates an extra, and somewhat redundant workstream for them, since these AESIs are already under close observation. AESIs for which we are seeing substantial or notable reporting are further evaluated via comparisons to background rates (using known COVID vaccine administration data tracked by CDC and provided weekly to FDA) as well as in population-based data sources both at FDA and CDC (e.g., BEST, CMS, Vaccine Safety Datalink (VSD)).
-Serious report screening: FDA MOs review serious reports coming into VAERS daily until meeting pre-specified milestones (i.e., certain time since authorization and doses administered) and then review aggregated PT counts weekly, by seriousness, AESI, Lot number, most frequent PTs, weekly changes in PT rankings, and other metrics.
-Pre-screening: for a couple of notable issues, the VAERS program contractor flags reports when they hit the door: these pre-screened events will have expedited gathering of medical records and CDC review and abstraction. TTS and anaphylaxis both have fallen into this category.

Of note, data mining alerts, which are designed to generate hypotheses of possible safety issues, are no longer particularly useful for our pharmacovigilance purposes when a signal has already been identified, such as for TTS, or when an issue (e.g., an AESI such as AMI) is being worked up in a more robust (e.g., active surveillance) system.

Further, we have a standard process for data mining screening in place for VAERS data; this screening was in place at the start of, and throughout, the COVID vaccine campaign; the frequency/nature of the calculations, stratifications and other parameters, are known and understood by us and our stakeholders. We understand that exploring new approaches might improve the methodology and is of interest to you. However, from our perspective, the approach employed during a period of intense, high profile surveillance should be standard, predictable, and road-tested. Results from adjusting parameters that raise or lower sensitivity of the alerts as the vaccination campaign is underway could lead to confusion and have unintended consequences (e.g., regarding vaccine confidence).

So, while we appreciate your work and interest at CDER on the COVID vaccine VAERS data, in the above context, we have found no indication for action based on your findings, which have been consuming resources at a time when

resources are stretched across preexisting robust pharmacovigilance activities. So, we are asking that you please hold off on creating and sending data mining results for COVID-19 vaccine AE data. Thanks much for your time and understanding, and sorry for the long email.

Kind regards,

Craig Zinderman, MD, MPH
Associate Director for Medical Policy
Office of Biostatistics and Epidemiology
FDA/Center for Biologics Evaluation and Research
[redacted]

Craig Zinderman, MD, MPH

Associate Director for Medical Policy
Office of Biostatistics and Epidemiology
FDA/Center for Biologics Evaluation and Research
[redacted]

Message

From: Menschik, David [/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=0407D7354456470CAB9BC2D3F98D6D3C-MENSCHIK]
Sent: 8/16/2021 3:18:21 PM
To: Richardson, Judith [redacted]
Subject: RE: Comments from the CDER/CBER/Commonwealth call from today

Will discuss at our 1:1 later today

From: Baer, Bethany <[redacted]>
Sent: Monday, August 16, 2021 11:17 AM
To: Richardson, Judith <[redacted]>
Cc: Menschik, David <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Comments from the CDER/CBER/Commonwealth call from today

Hi Judy,
I know that you caught part of the conversation on the call so I wanted you to be aware of things.
Thanks,
Bethany

From: Richardson, Judith <[redacted]>
Sent: Monday, August 16, 2021 10:48 AM
To: Baer, Bethany <[redacted]>; Menschik, David <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Comments from the CDER/CBER/Commonwealth call from today

Thanks Bethany! (for including me ☺)

From: Baer, Bethany <[redacted]>
Sent: Thursday, August 12, 2021 2:37 PM
To: Menschik, David <[redacted]>; Zinderman, Craig E <[redacted]>
Cc: Richardson, Judith <[redacted]>
Subject: Comments from the CDER/CBER/Commonwealth call from today

Hi David and Craig,
I wanted to pass on to you that during the Commonwealth contractor call today Ana Szarfman specifically brought up concerns directed at CBER's vaccine data mining and the use of the "20-year-old MGPS model which could potentially mask signals." She is especially concerned by the large (97% is what she stated, I believe) proportion of 2021 reports that are for COVID vaccines so that the expected comparison is lost. I let her know that you both were aware of these considerations. She may reach out to you directly again. I let her know that I am not the right person to respond to her concerns.
Thanks,
Bethany

From: "Forshee, Richard" <[redacted]>
To: "Anderson, Steven" <[redacted]>
Subject: FW: Issue #1 -- Death signal --> WVAERS 2021W21 data loaded on slc06lhx
Date: Mon, 12 Jul 2021 20:57:09 +0000
Importance: Normal
Attachments: VaccineHLT.xlsx
Inline-Images: image002.png

Hi Steve,

Here is what Ana just sent me. There is not much on the methods, just an Excel table. I haven't reviewed this yet.

Thanks,
--Rich

From: Szarfman, Ana <[redacted]>
Sent: Monday, July 12, 2021 4:36 PM
To: Forshee, Richard [redacted]>
Cc: Stockbridge, Norman L <[redacted]>; Weichold, Frank <[redacted]>
Subject: Issue #1 -- Death signal --> WVAERS 2021W21 data loaded on slc06lhx

Hi Dear Richard,

Many thanks for all the extremely important work you are all doing!

As we talked over the phone, I became aware last Fri that scientists from Cornell are concerned of an increased mortality signal with the COVID-19 vaccines.

We detected such a signal using the data collected by VAERS during the week ending on May 30, 2021, and made public one or two weeks later.

Please refer to the attached spreadsheet and to the email from Bill DuMouchel that I am forwarding, dated June 20, 2021.

Note that Bill used RGPS, a method that automatically unmask signals that remain hidden by other data mining methodologies, including by MGPS (a method we implemented in 1998).

For the COVID-19 analyses, Bill does not stratify by year, since in 2021 over 95% of the VAERS reports are for COVID-19 vaccines, and we would not have a proper background from all other vaccines to make comparisons.

Let me know if you have any questions.

Many thanks,

--Ana

Ana Szarfman, MD, PhD, FAMIA,
Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), and Fellow of the American Medical Informatics Association (2020)
Medical Officer, Safety Data Mining Developer and Medical Informatics Analyst,

Celebrating nearly a quarter of a century of successful implementation of safety data mining, interactive patient profiles, and other automated analytical tools.
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted] (office)
([redacted])

FDA U.S. FOOD & DRUG ADMINISTRATION

From: Bill DuMouchel <[redacted]>
Date: June 20, 2021 at 22:46:05 EDT
Subject: Re: WVAERS 2021W21 data loaded on slc06lhx

I created two runs based on Week21 VAERS:
ID 412: Vaccine Type vs PT
ID413: Vaccine+Manufacturer vs HLT

I'm attaching an Excel file with results from run 413. Sheet 1 has 24 masked DECs and Sheet 2 has all DECs.

Masking is here defined as ER05 > EB95 and ER05 > 1 and ERAM > 1.5*EBGM

It seems to me that when a strong signal shows up at the HLT level, it should be hard to discount it.

For sheet 1, note signals for the two HLTs D*eath and sudden death* and *Non-site specific embolism and thrombosis* show up for all three COVID19 vaccines.

Are we just supposed to ignore over 4000 of the former and 1500 of the latter HLT reports?
Can anyone propose theories of what potential biases are causing them to have such high disproportionalities? We hoped that use of AgeGroup11 would eliminate the main bias.

-Bill

From: Ruixia Song <[redacted]>
Sent: Thursday, June 17, 2021 9:34 AM
To: Bill DuMouchel [redacted]>; Steve Bright <[redacted]>; Rave Harpaz <[redacted]>
Cc: Mohammad Al-Ansari <[redacted]>; Alexander Nip <[redacted]>
Subject: WVAERS 2021W21 data loaded on slc06lhx

Hi Bill, Steve, Rave,

WVAERS 2021W21 data has been loaded to slc06lhx server.

Ruixia

From: "Miller, Elaine R. (CDC/DDID/NCEZID/DHQP)" <[redacted]>
To: "Miller, Elaine R. (CDC/DDID/NCEZID/DHQP)" <[redacted]>
Subject: FW: COVID19 vaccine safety in people with psoriasis; [CDC-1901314-X2R3D2] CRM:03381288
Date: Sat, 16 Oct 2021 02:11:24 +0000
Importance: Normal
Attachments: Shimabukuro_et_al_VAERS_2015.pdf

From: Shimabukuro, Tom (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Friday, October 15, 2021 10:11:21 PM (UTC-05:00) Eastern Time (US & Canada)
To: Syed, Maha N <[redacted]>; COVID19VaxSafety <[redacted]>
Cc: Gelfand, Joel <[redacted]>
Subject: RE: COVID19 vaccine safety in people with psoriasis; [CDC-1901314-X2R3D2] CRM:03381288

Dear Dr. Sayed,

FDA conducts empirical Bayesian data mining to assess for disproportionality in VAERS. We consider EB data mining the gold standard for disproportionality analysis. You can access the references for EB datamining in the attached paper. I would recommend against using the VAERS public data. CDC provides the public data as a public service and for transparency, but it's not a database that should be used for serious surveillance or research. Furthermore, the COVID-19 vaccination program is so unique and the reporting patterns to VAERS are so different that historical comparisons or comparisons with other vaccines would be uninformative and likely misleading. I think your best bet is to conduct a retrospective observational study in an EHR or claims database or a clinical study involving prospective data collection. Psoriasis is relatively common so I think a study if feasible. VAERS data are unlikely to be useful.

Regards,

Tom

Tom Shimabukuro, MD, MPH, MBA
Captain, U.S. Public Health Service
Deputy Director
Immunization Safety Office
Centers for Disease Control and Prevention (CDC)
1600 Clifton Road, [redacted], Atlanta, GA 30329

Phone: [redacted], Fax: [redacted]
Email: [redacted]

From: Syed, Maha N [redacted]
Sent: Friday, October 15, 2021 3:28 PM
To: Shimabukuro, Tom (CDC/DDID/NCEZID/DHQP) <[redacted]>; COVID19VaxSafety [redacted]
Cc: Gelfand, Joel <[redacted]>
Subject: COVID19 vaccine safety in people with psoriasis; [CDC-1901314-X2R3D2] CRM:03381288

Dear Tom Shimabukuro,

I hope my email finds you well.

We are planning on conducting a study using the VAERS database to assess for any signals of COVID19 vaccines triggering or exacerbating psoriasis, for the purpose of hypothesis generation.

Would you be able to elaborate the comparator vaccinee group you used in the disproportionality analysis from your publication? In our planned study, we were aiming to compare psoriasis patients who received covid-19 vaccinations during the period of December 2020-October 2021 versus psoriasis patients who were administered the flu vaccine during the same time period. However, the challenge is that AE reports for the comparator Flu vaccine group are very limited based on a feasibility analysis I conducted.

I look forward to your response

Kind Regards,

Maha
Maha N. Syed, MBBS
Post-doctoral Research Fellow
University of Pennsylvania
Perelman School of Medicine, Department of Dermatology
Tel: [redacted] | Mobile: [redacted]
Email: [redacted]

Vaccine 33 (2015) 4398–4405

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Review

Safety monitoring in the Vaccine Adverse Event Reporting System (VAERS)

Tom T. Shimabukuro[a,*], Michael Nguyen[b], David Martin[b], Frank DeStefano[a]

[a] Immunization Safety Office, Division of Health care Quality Promotion, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States

[b] Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD, United States

ARTICLE INFO

Article history:
Received 26 December 2014
Received in revised form 9 July 2015
Accepted 11 July 2015
Available online 22 July 2015

Keywords:
Vaccination
Vaccine adverse event
Adverse event following immunization
Adverse reaction
Adverse effect
Spontaneous reporting
Passive surveillance
Vaccine safety
Vaccine Adverse Event Reporting System (VAERS)

ABSTRACT

The Centers for Disease Control and Prevention (CDC) and the U.S. Food and Drug Administration (FDA) conduct post-licensure vaccine safety monitoring using the Vaccine Adverse Event Reporting System (VAERS), a spontaneous (or passive) reporting system. This means that after a vaccine is approved, CDC and FDA continue to monitor safety while it is distributed in the marketplace for use by collecting and analyzing spontaneous reports of adverse events that occur in persons following vaccination. Various methods and statistical techniques are used to analyze VAERS data, which CDC and FDA use to guide further safety evaluations and inform decisions around vaccine recommendations and regulatory action. VAERS data must be interpreted with caution due to the inherent limitations of passive surveillance. VAERS is primarily a safety signal detection and hypothesis generating system. Generally, VAERS data cannot be used to determine if a vaccine caused an adverse event. VAERS data interpreted alone or out of context can lead to erroneous conclusions about cause and effect as well as the risk of adverse events occurring following vaccination. CDC makes VAERS data available to the public and readily accessible online.

We describe fundamental vaccine safety concepts, provide an overview of VAERS for healthcare professionals who provide vaccinations and might want to report or better understand a vaccine adverse event, and explain how CDC and FDA analyze VAERS data. We also describe strengths and limitations, and address common misconceptions about VAERS. Information in this review will be helpful for healthcare professionals counseling patients, parents, and others on vaccine safety and benefit-risk balance of vaccination.

Published by Elsevier Ltd.

1. Introduction

The Centers for Disease Control and Prevention (CDC) and the U.S. Food and Drug Administration (FDA) conduct post-licensure safety monitoring of U.S. licensed vaccines. This means that after a vaccine is approved, CDC and FDA continue to monitor safety while it is distributed in the marketplace for use. CDC and FDA co-administer the Vaccine Adverse Event Reporting System (VAERS), a spontaneous (or passive) reporting system [1]. Spontaneous surveillance means that no active effort is made to search for, identify and collect information, but rather information is passively received from those who choose to voluntarily report their experience. Therefore, VAERS relies on the intuition and experience of healthcare professionals in particular, but likewise for patients, parents and caregivers, to recognize and report unusual or unexpected events following vaccination or suspected vaccine safety problems. CDC and FDA also independently administer large-linked electronic health record-based surveillance systems [2,3]. Various methods and statistical techniques are used to analyze VAERS data, which CDC and FDA use to guide further safety evaluations and inform decisions around vaccine recommendations and regulatory action. Furthermore, VAERS transmits its vaccine adverse event reports to the Uppsala Monitoring Center, the World Health Organization collaborating center for international drug and vaccine safety monitoring [4,5], in order to contribute to the global pharmacovigilance effort along with other countries that employ passive vaccine safety monitoring systems. VAERS data must be interpreted with caution

Abbreviations: VAERS, Vaccine Adverse Event Reporting System; AEFI, adverse event following immunization; CDC, Centers for Disease Control and Prevention; FDA, U.S. Food and Drug Administration; MedDRA, Medical Dictionary for Regulatory Activities.

* Corresponding author at: Immunization Safety Office, Centers for Disease Control and Prevention, 1600 Clifton Road NE [redacted], Atlanta, GA 30329, Unites States. Tel.: [redacted]; fax: [redacted].
E-mail address: [redacted] (T.T. Shimabukuro).

http://dx.doi.org/10.1016/j.vaccine.2015.07.035
0264-410X/Published by Elsevier Ltd.

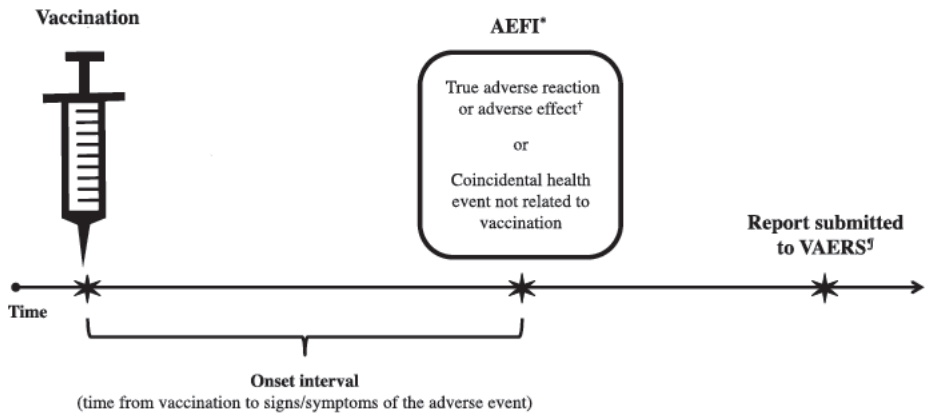


Fig. 1. Adverse event following immunization (AEFI) and the VAERS reporting timeline. *AEFI indicates only that the event happened after vaccination (i.e., a temporal association). †"Vaccine adverse reaction" and "vaccination adverse effect" are also AEFIs, but imply that the vaccine caused the event (i.e., a causal association). ¶There are no deadlines or time limits for the submission of a VAERS report, but reports should be submitted promptly after an adverse event occurs to facilitate surveillance and review. The National Vaccine Injury Compensation Program (VICP) is administered by the Health Resources and Services Administration (HRSA). The VICP is separate from the VAERS program and reporting an adverse event to VAERS does not constitute filing a claim for compensation to the VICP (see www.hrsa.gov/vaccinecompensation/index.html).

due to the inherent limitations of passive surveillance. VAERS is primarily a safety signal detection and hypothesis generating system. VAERS data interpreted alone or out of context can lead to erroneous conclusions about cause and effect or the risk of adverse events after vaccination.

We describe fundamental vaccine safety concepts, provide an overview of VAERS for healthcare professionals who provide vaccinations and might want to report or better understand a vaccine adverse event, and explain how CDC and FDA analyze VAERS data. We also describe strengths and limitations, and address common misconceptions about VAERS. Information in this review will be helpful for healthcare professionals counseling patients, parents, and others on vaccine safety and benefit-risk balance of vaccination.

2. What is a vaccine adverse event or adverse event following immunization?

A "vaccine adverse event," also referred to as an "adverse event following immunization" (AEFI), is an adverse health event or health problem that occurs following (Fig. 1) or during administration of a vaccine. Adverse events are temporally associated events, which might be caused by a vaccine or might be coincidental and not related to vaccination [6]. The Council for International Organizations of Medical Sciences (CIOMS) defines an AEFI as "... any untoward medical occurrence which follows immunization and which does not necessarily have a causal relationship with the usage of the vaccine. The adverse event may be any unfavourable or unintended sign, abnormal laboratory finding, symptom or disease" [7]. CIOMS also defines AEFI related to product quality defects, vaccination errors and anxiety-related reactions, in addition to those related to inherent properties of a vaccine. In contrast to the term "event", a vaccine adverse "reaction" and vaccination adverse "effect," like "adverse drug reaction" used in pharmacovigilance for drug safety monitoring [8], are synonymous terms that indicate a reasonable body of scientific evidence exists to suggest an adverse health event was caused by vaccination [6,9]. Examples of common vaccine adverse reactions are pain and redness at the injection site.

3. Why do the CDC and the FDA monitor vaccine safety?

The FDA requires extensive testing to evaluate safety and efficacy of a vaccine before granting licensure. The final phase of pre-licensure clinical trials might involve hundreds to thousands of volunteer study subjects [10]. Pre-licensure clinical trials are effective at identifying and characterizing the most common adverse events associated with a particular vaccine; examples include injection site reactions and post-vaccination fever. However, clinical trials might not be large enough to detect rare adverse events, which may be seen only after tens or hundreds of thousands of people are vaccinated. The limited patient follow-up period for clinical trials also constrains the ability to identify possible adverse events with delayed onset. Clinical trials generally conduct active follow-up on participants for up to a full year after vaccination, and often extended follow-up for periods beyond one a year. This level of follow-up is sufficient to assess most acute and delayed onset adverse events of interest for vaccine safety, but is not sufficient to assess conditions with onset multiple years following exposure. Additionally, clinical trials for initial licensure usually include only healthy individuals, so data on special populations, like those with chronic illnesses or pregnant women, are limited. Therefore, after a vaccine is licensed and distributed for widespread use it is necessary to conduct monitoring to further evaluate safety [11].

Apart from scientific and methodological issues, policy considerations also influence CDC and FDA determinations on vaccine safety monitoring. Vaccines are generally given to healthy individuals to prevent disease, whereas drugs are primarily given for treatment of illness. Sick patients, or parents of sick children, might be more willing to accept safety risks of drugs used to treat illnesses compared to vaccines used to prevent possible future illnesses. Furthermore, many state and local governments require vaccination for school attendance and healthcare facilities are increasingly requiring vaccination as a condition of employment [12,13]. These mandates place additional emphasis on vaccine safety and adverse event monitoring.

4. What is the Vaccine Adverse Event Reporting System (VAERS)?

VAERS is a national early warning system to detect possible safety problems in U.S. licensed vaccines. It is a spontaneous, voluntary reporting system for adverse events [1,14,15], and therefore no effort is made to search for individuals who experience adverse events and actively collect data, but rather VAERS passively receives information on adverse events from those who choose to report. VAERS is most useful as a hypothesis generating system with the primary goal to detect safety signals [9] that might be related to vaccination. The main objectives of VAERS are to: (1) detect new, unusual, or rare adverse events, (2) monitor reporting trends that

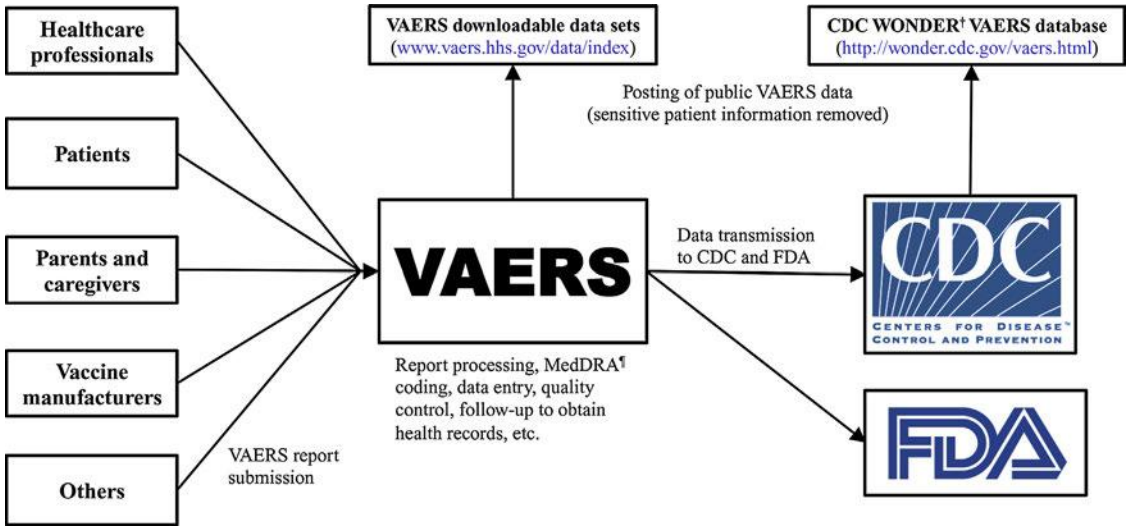


Fig. 2. Vaccine Adverse Event Reporting System (VAERS) report submission* and data flow. *During the time period 2011–2014, healthcare professionals submitted 38% of U.S. reports, patients and parents submitted 14%, vaccine manufacturers submitted 30%, and others (e.g., friends/acquaintances of the patient, 3rd party reporters who became aware of adverse events from the media, lawyers, etc.) submitted 12% (CDC unpublished data). There is variability in reporter type across different types and brands of vaccines. †Wide-ranging Online Data for Epidemiologic Research. ¶Medical Dictionary for Regulatory Activities.

might reflect true increases in known adverse events, (3) identify potential risk factors for particular types of adverse events, (4) assess the safety of newly licensed vaccines and new recommendations for existing vaccines, (5) detect and address possible reporting clusters (e.g., suspected localized [temporally or geographically] or product-/batch-/lot-specific adverse event reporting), (6) detect persistent safe-use problems and administration errors, and (7) provide a national safety monitoring system that extends to the entire general population for response to public health emergencies, such as a large-scale pandemic influenza vaccination program [16].

VAERS was established in 1990 [17,18] to fulfill a requirement of the National Childhood Vaccine Injury Act of 1986 [19]. By law, vaccine manufacturers are required to report adverse events that come to their attention, and healthcare professionals are required to report adverse events that are considered a contraindication to further doses of vaccine and those specified in the VAERS Table of Reportable Events Following Vaccination [20–23]. The National Childhood Vaccine Injury Act of 1986 also authorized establishment of the National Vaccine Injury Compensation Program [24]. Adverse events on the VAERS Table of Reportable Events Following Vaccination mirror the "illness, disability, injury or condition covered" conditions in the National Vaccine Injury Compensation Program's Vaccine Injury Table [25] used to help adjudicate petitioner claims of vaccine related injury.

Anyone can report an adverse event to VAERS, including healthcare professionals, vaccine manufacturers, patients, parents and caregivers, and others. Reports are submitted voluntarily either directly from individual reporters, who may be reporting for themselves or others, or secondarily from vaccine manufacturers, that also receive spontaneous reports and in turn submit them to VAERS. Reporting is encouraged for any clinically important or unexpected adverse event, even if the reporter is not sure if a vaccine caused the event [20]. VAERS accepts all reports without rendering judgment on clinical importance or whether vaccine(s) might have caused the adverse event.

5. How does VAERS work?

VAERS currently receives reports on a standard form via mail or fax, or through a secure online submission process (www.vaers.hhs.gov/esub/index). The VAERS form includes data fields for patient demographic information and medical history, information on the reporter and the facility where vaccine(s) were given, description of the adverse event and health outcomes, date of vaccination, vaccine(s) administered, onset of adverse event symptoms, recovery status, and other relevant information. VAERS reports are received at a central facility that is managed by a private contractor under the direction of CDC and FDA (Fig. 2). Here, staff specialized in coding case report information review reports and assign medical terms for adverse events using the Medical Dictionary for Regulatory Activities (MedDRA) [26], a widely used and accepted standardized medical terminology for adverse events. MedDRA terms are not confirmed medical diagnoses, but rather serve as the classification scheme to systematically encode information reported to VAERS. VAERS uses certified MedDRA coders and software programs to facilitate consistency in the capture and coding of signs and symptoms in reports. Reports are categorized as either serious or non-serious according to an FDA regulatory definition. Serious reports include at least one of the following: death following vaccination, life-threatening health event, hospitalization following vaccination or prolonged hospitalization if a vaccine was administered while the patient was already hospitalized, or lasting disability [21].

For VAERS reports submitted by the public, the primary reporter receives an acknowledgment letter or email and a request to provide additional information if there is missing or incomplete essential information on the report. For reports classified as serious, the VAERS contractor requests associated health records, including hospital discharge summaries, medical and laboratory results, and death certificates and autopsy reports for deaths. Additional MedDRA terms might be added based on information obtained through follow-up. Also, for serious reports where the patient has not recovered from the adverse event by the time the report was filed or recovery status was unknown, a follow-up letter is sent to the reporter at one year requesting information on recovery status if that information is still not known. Vaccine manufacturers are responsible for attempting to obtain follow-up information on serious and unexpected adverse event reports that they submit to VAERS [21].

Information in each report, along with assigned MedDRA terms, is entered into an electronic database and sent to CDC and FDA for analysis. Data are continuously updated as new reports come in and follow-up information for existing reports is received. CDC and

FDA receive a cumulative dataset every business day that contains all VAERS reports including recently entered reports and refreshed (or updated) reports. In addition, copies of original reports, any health records, and other associated documents are electronically maintained in an image database that CDC and FDA staff use to clinically review individual case reports. If errors or inconsistencies in reported information are detected during the course of follow-up or during routine analysis, corrections are made to the VAERS database. VAERS data from the primary reports, with sensitive patient information removed, are publicly available on the VAERS website (www.vaers.hhs.gov/data/index) and through CDC's Wide-ranging Online Data for Epidemiologic Research (WONDER) tool (http://wonder.cdc.gov/vaers.html) (Fig. 2). Due to patient privacy protections, additional information obtained during follow-up on individual VAERS reports is not included in the publicly available data.

During 2011–2014, VAERS averaged around 30,000 U.S. reports annually, with 7% classified as serious. Healthcare professionals submitted 38% of reports, vaccine manufacturers 30% and patients and parents 14%. Reporter type and percent of serious reports vary across vaccines, age of vaccine recipient and how long the vaccine has been in use. During this same time period VAERS averaged around 6000 foreign source reports annually. Vaccine manufacturers, which accounted for >99% of foreign source reporting, are required by law to submit foreign source adverse event reports that are both serious and unexpected [21], but not other types of foreign source reports. Given the vaccine manufacturer reporting requirements and the minimal amount of direct foreign source public reporting, it is not surprising that a relatively high percentage (48%) of foreign source reports are classified as serious. This likely represents selective reporting based on regulatory requirements rather than any substantial differences in safety profiles of foreign vaccines.

6. How do CDC and FDA analyze VAERS data?

CDC and FDA use several methods to analyze VAERS data to detect vaccine safety signals. CDC focuses on public health priority vaccines, like influenza vaccine which is given in large quantities during a compressed time period, and newly licensed and recommended vaccines during their initial uptake period. The data needs of the Advisory Committee on Immunization Practices (ACIP) [27] often drive CDC's monitoring priorities. FDA monitors all U.S. licensed vaccines and regularly submits mandated post-licensure safety reports to its advisory committees. When necessary, CDC, FDA and state and local health departments collaborate on investigations of unusual or unexpected reports or concerning patterns of reporting (e.g., clusters). The joint monitoring efforts of CDC and FDA ensure that U.S. licensed vaccines are continuously monitored, with emphasis on high use vaccines, new vaccines, and when new recommendations are implemented for existing vaccines.

6.1. Descriptive analysis, historical comparisons and reporting trends over time

The basic analyses of VAERS data are intended to detect concerning patterns or unusual and unexpected changes in adverse event reporting that might indicate a safety problem in a specific vaccine or vaccine type. CDC and FDA physicians, epidemiologists and statisticians assess numbers of reports, types of reports based on serious and non-serious status, the most common adverse events, current versus historical data, and reporting trends over time, such as comparisons of influenza vaccine reports across multiple consecutive influenza seasons. Analysis also includes evaluation of reporting rates of adverse events in the context of vaccine doses distributed for use in the U.S. marketplace. Vaccine doses distributed provides a proxy measure of persons vaccinated. Reporting rates enable comparison with background rates of adverse events from the literature or other sources, but they must be interpreted cautiously since vaccine doses distributed might not all be administered. Even if they do not exceed known background rates, reporting rates for specific adverse events that approach the background rates might indicate a safety problem due to the known underreporting of adverse events to VAERS.

6.2. Disproportionality analysis

Disproportionality analysis involves statistical techniques like empirical Bayesian data mining and the proportional reporting ratio to assess for disproportional reporting of specific vaccine-adverse event combinations [28–30]. VAERS is not able to provide incidence of adverse events. As a passive, numerator-only surveillance system, VAERS lacks information on total number of individuals vaccinated and total number who experience an adverse event, as well as incidence of adverse events in unvaccinated individuals. However, the proportion of reports involving a specific adverse event and a specific vaccine can be compared to the proportion of reports involving the same adverse event and other vaccines. An example would be comparing the proportion of live attenuated influenza vaccine (LAIV)-nasal congestion reports (a known causal association [31]) to the proportion of inactivated influenza vaccine-nasal congestion reports. Here we might expect to see a higher proportion of LAIV reports with nasal congestion than for inactivated influenza vaccine, for which there is no known causal association. In this case, disproportional reporting observed in post-licensure surveillance would not be considered a safety signal because nasal congestion is already a known, well characterized adverse reaction that was observed in clinical trials. A mathematical representation of the proportional reporting ratio illustrates the concept:

	Adverse event of interest	All other adverse events
Vaccine of interest	V_iAE_i	V_iAE_x
Comparator vaccine(s)	V_xAE_i	V_xAE_x

$$\text{Proportional reporting ratio} = \frac{V_iAE_i/(V_iAE_i + V_iAE_x)}{V_xAE_i/(V_xAE_i + V_xAE_x)}$$

In this equation, the proportion of reports involving the vaccine of interest and the adverse event of interest in relation to all adverse event reports involving the vaccine of interest is divided by the proportion of reports involving comparator vaccine(s) with the adverse event of interest in relation to all adverse event reports for comparator vaccine(s). The mathematical criteria used for a statistical signal is a proportional reporting ratio ≥2, chi-square ≥4 and number of reports in a cell ≥3 [30].

Disproportionality analysis complements clinical reviews and other analyses to identify adverse events that may be more frequently associated with a particular vaccine. A result that exceeds a pre-specified statistical alerting threshold might warrant further evaluation, such as clinical review of reports, but does not definitively demonstrate a true increased incidence of an adverse event, a causal association, or a safety problem. If, after an initial evaluation, CDC and FDA determine that a safety signal requires further assessment, epidemiologic studies can be conducted using other, more robust data sources to assess for causality [2,3]. An illustrative example of signal detection in VAERS using disproportionality analysis for febrile seizures in young children following inactivated influenza vaccine, with follow-on assessment using clinical review

of VAERS reports and an epidemiologic study in another data source is described in the final section of this paper.

6.3. Clinical review of reports

CDC and FDA physicians review serious reports, selected reports based on results of descriptive analysis and disproportionality analysis, and reports for selected conditions of interest. Clinical reviews are conducted to characterize the completeness and quality of reports, verify diagnoses if possible, characterize clinical and laboratory features, assess other potential risk factors (e.g., co-administration of vaccines, underlying health conditions), and evaluate the interval between vaccination and the adverse event. Reviewers use clinical judgment to detect concerning patterns or unusual and unexpected adverse events. CDC physicians generally conduct clinical reviews of selected types of vaccines and conditions of interest for particular vaccines (e.g., serious and pregnancy-related reports for influenza vaccines). FDA physicians structure clinical reviews of serious reports around individual vaccine brands with a regulatory focus. CDC and FDA regularly share information on clinical review findings. For selected adverse events of interest that are the focus of enhanced surveillance (e.g., anaphylaxis following inactivated influenza vaccine in egg allergic patients), Brighton Collaboration case definitions [32] are used when available. The Brighton Collaboration is a global research network with a mission to "...enhance the science of vaccine research by providing standardized, validated, and objective methods for monitoring safety profiles and benefit to risk ratios of vaccines." (https://brightoncollaboration.org/public/who-we-are.html). The Brighton Collaboration generates standardized adverse event case definitions in order to enhance data consistency and comparability across systems and studies.

7. What are the strengths of VAERS?

VAERS is national in scope and is able to receive information from the entire U.S. population. Because of the large and diverse population available to report, VAERS is able to rapidly detect possible safety problems and rare adverse events [1,14,15]. VAERS reports often include detailed information on vaccines given, characteristics of the individual vaccinated, and the adverse event itself. Furthermore, follow-up to obtain health records, when necessary, is possible. Due to direct reporting capability and the speed at which reports and follow-up information can be processed and analyzed, VAERS can often provide the earliest information on potential vaccine safety problems. VAERS is less impacted by data lags and delayed access to health records than claims-based monitoring systems, although these types of systems often compliment VAERS by allowing for more sophisticated follow-on signal assessment due to availability of numerator and denominator data. Lastly, VAERS data are made available online to the public, which affords an important level of transparency. This service allows the public to see the amount and nature of spontaneous adverse event reporting data that CDC and FDA collect and analyze to guide further safety evaluations and inform decisions around vaccine recommendations and regulatory action.

8. What are the limitations of VAERS?

Like all spontaneous public health reporting systems, VAERS has limitations [1,14]. VAERS is subject to reporting bias, including underreporting of adverse events – especially common, mild ones [33,34] – and stimulated reporting, which is elevated reporting that might occur in response to intense media attention and increased public awareness, such as during the 2009 H1N1 pandemic influenza vaccination program [35]. Quality and completeness of VAERS reports are variable and many reports lack valid medical diagnoses. The amount of VAERS reporting (30,000 U.S. reports annually) makes it impractical to conduct detailed follow-up on all reports to obtain missing and incomplete information and correct inconsistencies and errors. Because VAERS data do not include an unvaccinated comparison group, it is not possible to calculate and compare rates of adverse events in vaccinated versus unvaccinated individuals and determine if vaccination is associated with an increased risk of an adverse event (Fig. 3). Reporting efficiency, which is the proportion of adverse events that actually get reported to VAERS, is unknown, but is believed to be higher for clinically serious conditions. In a 1995 study, reporting sensitivities ranged from 68% for vaccine-associated polio following oral poliovirus vaccine to <1% for rash following measles, mumps, and rubella (MMR) vaccine [33]. Although underreporting is a limitation, VAERS is capable of detecting possible safety problems through disproportionality analyses and the other methods previously described.

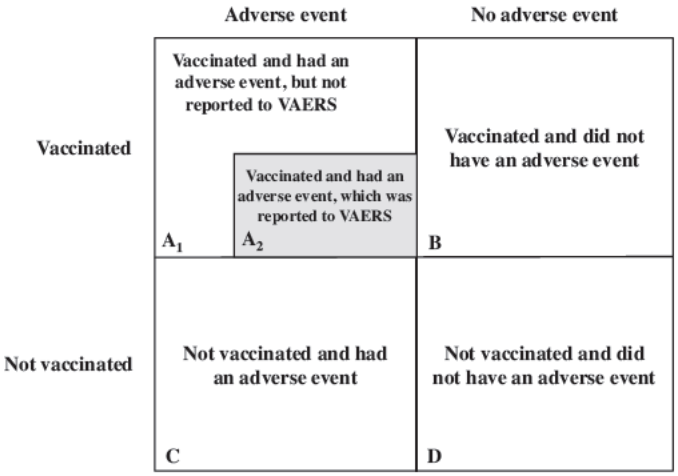


Fig. 3. 2 × 2 contingency table illustrating a hypothetical single vaccine and adverse event (AE) combination scenario. A_2 = VAERS database; incidence of AE in vaccinated individuals = $(A_1 + A_2)/((A_1 + A_2) + B)$; reporting efficiency to VAERS = $A_2/(A_1 + A_2)$; incidence of AE in unvaccinated individuals = $C/(C + D)$.

Except in unambiguous biologically plausible cases (like pain and redness at the injection site), it generally cannot be determined if a vaccine caused an adverse event using VAERS data [11,18]. On rare occasions, a detailed VAERS report with documentation of conclusive clinical or laboratory evidence might be sufficient to establish causality. For example, there have been case reports where vaccine strain rotavirus has been isolated from a stool specimen in a vaccinated infant experiencing severe gastroenteritis who was later diagnosed with severe combined immunodeficiency [36]. There have also been case reports documenting anaphylaxis occurring within an appropriate onset interval following vaccination with no other obvious environmental exposure triggers [37].

9. Misconceptions about VAERS

Perhaps the most common misconceptions about VAERS are that temporally associated reports represent true adverse reactions caused by vaccination, and that VAERS reports equate to rates of adverse events or indicate risk of adverse events associated with vaccination. The VAERS website has specific guidance on interpreting case report information, which includes the statement: "When evaluating data from VAERS, it is important to note that for any reported event, no cause-and-effect relationship has been established ... VAERS collects data on any adverse event

following vaccination, be it coincidental or truly caused by a vaccine" [38]. Despite this cautionary guidance, VAERS reports have been misinterpreted and erroneously communicated as definitive evidence of causally associated adverse events. For example, during the U.S. multi-state measles outbreak of 2015 [39], unsubstantiated claims of over 100 deaths caused by MMR vaccine in the United States during the previous decade began circulating on the Internet [40,41]. The claim was based on VAERS reports in the public data. The authors of the Internet articles further stated that no measles related deaths had been reported in the United States during the same time period, implying that MMR vaccine was doing more harm than good. In fact, many of the death reports after MMR vaccination involved children with serious preexisting medical conditions or were likely unrelated to vaccination (e.g., accidents). The complete VAERS reports and accompanying health records, autopsy reports and death certificates were reviewed in depth by CDC and FDA physicians and no concerning patterns emerged that would suggest a causal relationship with MMR vaccination and death [42].

The relatively rapid increase in numbers of reports to VAERS following the introduction and initial uptake of a new vaccine, an expected occurrence [43], has been misinterpreted as actual increases in incidence of adverse events and vaccine related risk. This has been the case with VAERS reports following quadrivalent human papillomavirus (HPV4) vaccination [44], which as expected, increased as uptake of HPV4 vaccine increased following licensure in 2006. However, post-licensure epidemiologic studies have consistently demonstrated the safety of HPV4 vaccine [45–51], confirming the limitations of passive surveillance systems like VAERS.

10. Closing thoughts

VAERS has been used to monitor adverse events since 1990 and continues to ably serve as the nation's frontline post-licensure vaccine safety monitoring system. VAERS has successfully detected safety signals that required further evaluation [36,52–59] and has also provided reassurance on the safety of vaccines [60–63]. One of the earliest successes in signal detection and assessment in VAERS involved the first rotavirus vaccine, RotaShield®. Within nine months of its licensure in the United States in August 1998, reports to VAERS raised suspicion of a possible safety problem with intussusception, a type of bowel obstruction, in infants [52]. Further evaluation of the signal, which combined estimated RotaShield® doses administered with known background rates of infant intussusception, indicated that the observed number of intussusception reports to VAERS within one week of receipt of RotaShield® was approaching what would be expected by chance alone. Given the known underreporting of adverse events to VAERS, these findings were concerning enough for CDC to suspend its recommendation for RotaShield® vaccination and initiate further investigation [64]; shortly thereafter the vaccine was withdrawn from the market by the manufacturer [65]. More recently, VAERS detected disproportional reporting for febrile seizures in young children following an inactivated influenza vaccine during the 2010–2011 influenza season [58,59]. Clinical review of the VAERS reports indicated the cases were typical of uncomplicated febrile seizures and all children fully recovered. A related finding was later detected using sequential monitoring methods in a separate CDC surveillance system that uses large-linked electronic health record databases, and the increased risk was assessed and quantified in an epidemiologic study [66]. The information was quickly communicated to the public along with reassurances on the benefit-risk balance of vaccinating children against influenza [67].

CDC and FDA are currently updating the VAERS reporting form and enhancing electronic methods for reporting to improve the public health and regulatory value of VAERS data. These data adjustments and system enhancements are necessary responses to changes in the U.S. immunization program that have made some VAERS data fields obsolete and have imposed other needs such as information on adverse events following maternal vaccination. Additionally, CDC and FDA are implementing processes to improve and facilitate online reporting and to transition vaccine manufacturers to reporting using standardized messages through electronic data interchange [68–71]. A major impetus for improving electronic reporting and increasing automation in VAERS was the 2009 influenza pandemic experience where 10,000 influenza A (H1N1) monovalent (pandemic) vaccine reports were submitted to VAERS during the 2009–2010 influenza season [72]. Other future initiatives might include incorporating adverse event reporting reminders [73] and VAERS reporting capability directly into the software of electronic health records systems [74].

While near real-time sequential monitoring using large-linked electronic health record databases has become increasingly prominent in post-licensure vaccine safety surveillance [75], VAERS will continue to remain a foundation of the U.S. vaccine safety monitoring infrastructure. Understanding the purpose, strengths, and limitations of VAERS is essential when interpreting VAERS data and when responding to concerns from patients, parents, and others about adverse event reports to VAERS and vaccine safety in general. Healthcare professionals reporting to VAERS is arguably the most broad-based, cost-effective, and timely way to obtain real world feedback on vaccine safety. Often healthcare professionals, relying on experience and intuition, are the first to suspect a medical product problem and bring it to the attention of public health and regulatory officials [76,77]. CDC and FDA encourage reporting of clinically important or unexpected adverse events to VAERS following any U.S. licensed vaccines.

Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention and the U.S. Food and Drug Administration.

Funding source

No external sources of funding.

Acknowledgments

The authors thank Paige Lewis from the Immunization Safety Office at the Centers for Disease Control and Prevention for her contributions to this paper.

Conflicts of interest: None of the authors have any conflicts of interest.

References

[1] Varricchio F, Iskander J, Destefano F, Ball R, Pless R, Braun MM, et al. Understanding vaccine safety information from the Vaccine Adverse Event Reporting System. Pediatr Infect Dis J 2004;23(4):287–94.
[2] Baggs J, Gee J, Lewis E, Fowler G, Benson P, Lieu T, et al. The Vaccine Safety Datalink: a model for monitoring immunization safety. Pediatrics 2011;127(Suppl 1):S45–53.
[3] Nguyen M, Ball R, Midthun K, Lieu TA. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. Pharmacoepidemiol Drug Saf 2012;21(Suppl 1):291–7.
[4] The Uppsala Monitoring Centre. Safer medicines, safer use of medicines, safer patients: what UMC is doing to help it happen. Available at: http://www.who-umc.org/graphics/27916.pdf [accessed 5.06.15].

[5] Uppsala Monitoring Centre (UMC). Report from the WHO Collaborating Centre for International Drug Monitoring: activities July 2013–June 2014. Available at: http://www.who-umc.org/graphics/28368.pdf [accessed 5.06.15].

[6] Centers for Disease Control and Prevention. Understanding the Vaccine Adverse Event Reporting System (VAERS). February 2013. Available at: http://www.cdc.gov/vaccines/hcp/patient-ed/conversations/downloads/vacsafe-vaers-color-office.pdf [accessed 5.06.15].

[7] Council for International Organizations of Medical Sciences (CIOMS). Definition and application of terms for vaccine pharmacovigilance: report of CIOMS/WHO working group on vaccine pharmacovigilance. Geneva, Switzerland: WHO Press, World Health Organization; 2012.

[8] Expert Working Group (Efficacy) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Guideline for industry. Clinical safety data management: definitions and standards for expedited reporting. March 1995. Available at: http://www.fda.gov/downloads/Drugs/Guidances/ucm073087.pdf [accessed 5.06.15].

[9] Council for International Organizations of Medical Sciences (CIOMS). Practical aspects of signal detection in pharmacovigilance: report of CIOMS working group VIII. Geneva, Switzerland: CIOMS; 2010.

[10] Marshall V, Baylor NW. Food and Drug Administration regulation and evaluation of vaccines. Pediatrics 2011;127(Suppl 1):S23–30.

[11] Chen RT, Davis RL, Rhodes PH. Special methodological issues in pharmacoepidemiology studies of vaccine safety. In: Strom BL, editor. Pharmacoepidemiology. 4th ed. Sussex: John Wiley & Sons; 2005.

[12] Centers for Disease Control and Prevention. State school immunization requirements and vaccine exemption laws. February–March 2015. Available at: http://www.cdc.gov/phlp/docs/school-vaccinations.pdf [accessed 5.06.15].

[13] Centers for Disease Control and Prevention. State immunization laws for healthcare workers and patients. November 2014. Available at: http://www2a.cdc.gov/vaccines/statevaccsApp/default.asp [accessed 5.06.15].

[14] Iskander JK, Miller ER, Chen RT. The role of the Vaccine Adverse Event Reporting system (VAERS) in monitoring vaccine safety. Pediatr Ann 2004;33(9):599–606.

[15] Singleton JA, Lloyd JC, Mootrey GT, Salive ME, Chen RT. An overview of the vaccine adverse event reporting system (VAERS) as a surveillance system. VAERS Working Group. Vaccine 1999;17(22):2908–17.

[16] Centers for Disease Control and Prevention. Surveillance for adverse events following immunization using the Vaccine Adverse Event Reporting System (VAERS). In: Manual for the surveillance of vaccine-preventable diseases. Atlanta, GA: Centers for Disease Control and Prevention; 2011, October.

[17] Centers for Disease Control and Prevention. Vaccine Adverse Event Reporting System – United States. MMWR Morb Mortal Wkly Rep 1990;39(41):730–3.

[18] Chen RT, Rastogi SC, Mullen JR, Hayes SW, Cochi SL, Donlon JA, et al. The Vaccine Adverse Event Reporting System (VAERS). Vaccine 1994;12(6):542–50.

[19] 42 U.S. Code §§ 300aa-1 to 300aa-34. National Childhood Vaccine Injury Act (1986).

[20] What can be reported to VAERS? Available at: https://vaers.hhs.gov/about/faqs#what [accessed 5.06.15].

[21] U.S. Code of Federal Regulations, 21 CFR 600.80. Postmarketing reporting of adverse experiences (2014). Available at: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=600.80 [accessed 5.06.15].

[22] 42 U.S. Code § 300aa-25. Recording and Reporting of Information (1999). Available at: http://www.fda.gov/RegulatoryInformation/Legislation/ucm180189.htm [accessed 5.06.15].

[23] VAERS Table of Reportable Events Following Vaccination. Available at: https://vaers.hhs.gov/resources/VAERS_Table_of_Reportable_Events_Following_Vaccination.pdf [accessed 5.06.15].

[24] Cook KM, Evans G. The National Vaccine Injury Compensation Program. Pediatrics 2011 May;127(Suppl 1):S74–7.

[25] Vaccine Injury Table. Available at: http://www.hrsa.gov/vaccinecompensation/vaccineinjurytable.pdf [accessed 5.06.15].

[26] Medical Dictionary for Regulatory Activities (MedDRA). Available at: http://www.meddra.org/ [accessed 5.06.15].

[27] Smith JC. The structure, role, and procedures of the U.S. Advisory Committee on Immunization Practices (ACIP). Vaccine 2010;28(April (Suppl 1)):A68–75.

[28] DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Am Stat 1999;53:177–90.

[29] Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. Drug Saf 2002;25(6):381–92.

[30] Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf 2001;10(6):483–6.

[31] Ambrose CS, Walker RE, Connor EM. Live attenuated influenza vaccine in children. Semin Pediatr Infect Dis 2006;17(October (4)):206–12.

[32] Kohl KS, Bonhoeffer J, Braun MM, Chen RT, Duclos P, Heijbel H, et al. The Brighton Collaboration: creating a global standard for case definitions (and guidelines) for adverse events following immunization. In: Henriksen K, Battles JB, Marks ES, Lewin DI, editors. Advances in patient safety: from research to implementation. Concepts and methodology, vol 2. Rockville, MD: Agency for Healthcare Research and Quality (US); 2005.

[33] Rosenthal S, Chen R. The reporting sensitivities of two passive surveillance systems for vaccine adverse events. Am J Public Health 1995;85(12):1706–9.

[34] Verstraeten T, Baughman AL, Cadwell B, Zanardi L, Haber P, Chen RT. Enhancing vaccine safety surveillance: a capture-recapture analysis of intussusception after rotavirus vaccination. Am J Epidemiol 2001;154(11):1006–12.

[35] Vellozzi C, Broder KR, Haber P, Guh A, Nguyen M, Cano M, et al. Adverse events following influenza A (H1N1) 2009 monovalent vaccines reported to the Vaccine Adverse Event Reporting System, United States, October 1, 2009–January 31, 2010. Vaccine 2010;28(45):7248–55.

[36] Bakare N, Menschik D, Tiernan R, Hua W, Martin D. Severe combined immunodeficiency (SCID) and rotavirus vaccination: reports to the Vaccine Adverse Events Reporting System (VAERS). Vaccine 2010;28(40):6609–12.

[37] Loughlin AM, Marchant CD, Adams W, Barnett E, Baxter R, Black S, et al. Causality assessment of adverse events reported to the Vaccine Adverse Event Reporting System (VAERS). Vaccine 2012;30(50):7253–9.

[38] VAERS data: guide to interpreting VAERS case report information. Available at: https://vaers.hhs.gov/data/index [accessed 5.06.15].

[39] Zipprich J, Winter K, Hacker J, Xia D, Watt J, Harriman K. Measles outbreak – California, December 2014–February 2015. MMWR Morb Mortal Wkly Rep 2015;64(February (6)):153–4.

[40] Shilhavy B. Zero US measles deaths in 10 years, but over 100 measles vaccine deaths reported. Health Impact News, February 12, 2015. Available at: http://healthimpactnews.com/2015/zero-u-s-measles-deaths-in-10-years-but-over-100-measles-vaccine-deaths-reported/ [accessed 5.06.15].

[41] Huff EA. Measles vaccines kill more people than measles, CDC data proves. Global research, February 5, 2015. Available at http://www.globalresearch.ca/measles-vaccines-kill-more-people-than-measles-cdc-data-proves/5429736 [accessed 5.06.15].

[42] Moro PL, Arana J, Cano M, Lewis P, Shimabukuro TT. Deaths reported to the Vaccine Adverse Event Reporting System (VAERS), United States, 1997–2013. Clin Infect Dis 2015;(May), pii: civ423.

[43] Weber JCP. Epidemiology of adverse reactions to nonsteroidal antiinflammatory drugs. In: Rainsford KD, Velo GP, editors. Advances in inflammation research, vol 6. New York: Raven Press; 1984. p. 1–6.

[44] Erickson N. How closely does the CDC monitor HPV vaccine safety? January 5, 2014. Available at: http://sanevax.org/closely-cdc-monitor-hpv-vaccine-safety/ [accessed 5.06.15].

[45] Gee J, Naleway A, Shui I, Baggs J, Yin R, Li R, et al. Monitoring the safety of quadrivalent human papillomavirus vaccine: findings from the Vaccine Safety Datalink. Vaccine 2011;29(October (46)):8279–84.

[46] Chao C, Klein NP, Velicer CM, Sy LS, Slezak JM, Takhar H, et al. Surveillance of autoimmune conditions following routine use of quadrivalent human papillomavirus vaccine. J Intern Med 2012;271(February (2)):193–203.

[47] Klein NP, Hansen J, Chao C, Velicer C, Emery M, Slezak J, et al. Safety of quadrivalent human papillomavirus vaccine administered routinely to females. Arch Pediatr Adolesc Med 2012;166(December (12)):1140–8.

[48] Arnheim-Dahlström L, Pasternak B, Svanström H, Sparén P, Hviid A. Autoimmune, neurological, and venous thromboembolic adverse events after immunisation of adolescent girls with quadrivalent human papillomavirus vaccine in Denmark and Sweden: cohort study. BMJ 2013;347(October):f5906.

[49] Grimaldi-Bensouda L, Guillemot D, Godeau B, Bénichou J, Lebrun-Frenay C, Papeix C, et al. Autoimmune disorders and quadrivalent human papillomavirus vaccination of young female subjects. J Intern Med 2014;275(April (4)):398–408.

[50] Scheller NM, Pasternak B, Svanström H, Hviid A. Quadrivalent human papillomavirus vaccine and the risk of venous thromboembolism. JAMA 2014;312(July (2)):187–8.

[51] Scheller NM, Svanström H, Pasternak B, Arnheim-Dahlström L, Sundström K, Fink K, et al. Quadrivalent HPV vaccination and risk of multiple sclerosis and other demyelinating diseases of the central nervous system. JAMA 2015;313(January (1)):54–61.

[52] Centers for Disease Control and Prevention. Intussusception among recipients of rotavirus vaccine – United States, 1998–1999. MMWR Morb Mortal Wkly Rep 1999;48(27):577–81.

[53] Centers for Disease Control and Prevention. Guillain-Barré syndrome among recipients of Menactra meningococcal conjugate vaccine – United States, June–July 2005. MMWR Morb Mortal Wkly Rep 2005;54(October (40)):1023–5.

[54] Centers for Disease Control and Prevention. Update: Guillain-Barré syndrome among recipients of Menactra meningococcal conjugate vaccine – United States, October 2005–February 2006. MMWR Morb Mortal Wkly Rep 2006;55(April (13)):364–6.

[55] Centers for Disease Control and Prevention. Update: Guillain-Barré syndrome among recipients of Menactra meningococcal conjugate vaccine – United States, June 2005–September 2006. MMWR Morb Mortal Wkly Rep 2006;55(October (41)):1120–4.

[56] Braun MM, Patriarca PA, Ellenberg SS. Syncope after immunization. Arch Pediatr Adolesc Med 1997;151(March (3)):255–9.

[57] Centers for Disease Control and Prevention. Syncope after vaccination – United States, January 2005–July 2007. MMWR Morb Mortal Wkly Rep 2008;57(May (17)):457–60.

[58] Leroy Z, Broder K, Menschik D, Shimabukuro T, Martin D. Febrile seizures after 2010–2011 influenza vaccine in young children, United States: a vaccine safety signal from the vaccine adverse event reporting system. Vaccine 2012;30(11):2020–3.

[59] Martin D, Menschik D, Bryant-Genevier M, Ball R. Data mining for prospective early detection of safety signals in the Vaccine Adverse Event Reporting System (VAERS): a case study of febrile seizures after a 2010–2011 seasonal influenza virus vaccine. Drug Saf 2013;36(7):547–56.

[60] Braun MM, Mootrey GT, Salive ME, Chen RT, Ellenberg SS. Infant immunization with acellular pertussis vaccines in the United States: assessment of the first two years' data from the Vaccine Adverse Event Reporting System (VAERS). Pediatrics 2000;106(October (4)):E51.
[61] Slade BA, Leidel L, Vellozzi C, Woo EJ, Hua W, Sutherland A, et al. Postlicensure safety surveillance for quadrivalent human papillomavirus recombinant vaccine. JAMA 2009;302(7):750–7.
[62] Centers for Disease Control and Prevention. Safety of influenza A (H1N1) 2009 monovalent vaccines – United States, October 1–November 24, 2009. MMWR Morb Mortal Wkly Rep 2009;58(48):1351–6.
[63] Moro PL, Broder K, Zheteyeva Y, Walton K, Rohan P, Sutherland A, et al. Adverse events in pregnant women following administration of trivalent inactivated influenza vaccine and live attenuated influenza vaccine in the Vaccine Adverse Event Reporting System, 1990–2009. Am J Obstet Gynecol 2011;204(2), 146.e1–7.
[64] Centers for Disease Control and Prevention. Suspension of rotavirus vaccine after reports of intussusception – United States, 1999. MMWR Morb Mortal Wkly Rep 2004;53(September (34)):786–9. Erratum in: MMWR Morb Mortal Wkly Rep 2004; 53(September (37)):879.
[65] Centers for Disease Control and Prevention. Withdrawal of rotavirus vaccine recommendation. MMWR Morb Mortal Wkly Rep 1999;48(November (43)):1007.
[66] Tse A, Tseng HF, Greene SK, Vellozzi C, Lee GM, VSD Rapid Cycle Analysis Influenza Working Group. Signal identification and evaluation for risk of febrile seizures in children following trivalent inactivated influenza vaccine in the Vaccine Safety Datalink Project, 2010–2011. Vaccine 2012;30(March (11)):2024–31.
[67] Centers for Disease Control and Prevention. Febrile seizures in children following vaccination with influenza vaccines and pneumococcal vaccines – 2010–2011 influenza season. Available at: http://www.cdc.gov/vaccinesafety/Concerns/FebrileSeizures-archived.html [accessed 5.06.15].
[68] Shimabukuro T. The Vaccine Adverse Event Reporting System (VAERS) form Version 2.0 (proposed). In: Presented at the National Vaccine Advisory Committee meeting. 2014.
[69] Department of Health and Human Services CfDCaP, editor. Request for comment on draft Vaccine Adverse Event Reporting System (VAERS) 2.0 form. 2014, November. p. 69853–4.
[70] Department of Health and Human Services FaDA, editor. Postmarketing safety reports for human drug and biological products; electronic submission requirements. 2014, June. p. 33072–92.
[71] Services DoHaH, editor. Draft guidance for industry on providing submissions in electronic format – postmarketing safety reports. Department of Health and Human Services, Food and Drug Administration; 2014, June. p. 33200–1.
[72] Vellozzi C, Broder KR, Haber P, Guh A, Nguyen M, Cano M, et al. Adverse events following influenza A (H1N1) 2009 monovalent vaccines reported to the Vaccine Adverse Event Reporting System, United States, October 1, 2009–January 31, 2010. Vaccine 2010;28(October (45)):7248–55.
[73] Hinrichsen VL, Kruskal B, O'Brien MA, Lieu TA, Platt R. Using electronic medical records to enhance detection and reporting of vaccine adverse events. J Am Med Inform Assoc 2007;14(November–December (6)):731–5.
[74] Baker MA, Kaelber DC, Bar-Shain DS, Moro PL, Zambarano B, Mazza M, et al. Advanced clinical decision support for vaccine adverse event detection and reporting. Clin Infect Dis 2015;(June), pii: civ430.
[75] Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K, et al. Real-time vaccine safety surveillance for the early detection of adverse events. Med Care 2007;45(October (10 Suppl 2)):S89–95.
[76] Connolly HM, Crary JL, McGoon MD, Hensrud DD, Edwards BS, Edwards WD, et al. Valvular heart disease associated with fenfluramine-phentermine. N Engl J Med 1997;337(August (9)):581–8. Erratum in: N Engl J Med 1997; 337(December 24):1783.
[77] Centers for Disease Control and Prevention. Cardiac valvulopathy associated with exposure to fenfluramine or dexfenfluramine: U.S. Department of Health and Human Services interim public health recommendations, November 1997. MMWR Morb Mortal Wkly Rep 1997;46(November (45)): 1061–6.