From: "Marks, Peter" <[redacted]>
To: "Forshee, Richard" <[redacted]>
Cc: "Anderson, Steven" <[redacted]>, "Witten, Celia (CBER)" <[redacted]>
Subject: RE: Contact from Ana Szarfman
Date: Tue, 13 Jul 2021 18:14:15 +0000
Importance: Normal
Inline-Images: image001.png; image002.jpg; image003.jpg; image004.jpg; image005.jpg; image006.jpg

---

Dear Rich,

Thanks so much for documenting this. I will follow up appropriately.

Best Regards,
Peter

From: Forshee, Richard <[redacted]>
Sent: Tuesday, July 13, 2021 1:55 PM
To: Marks, Peter <[redacted]>
Cc: Anderson, Steven <[redacted]>; Witten, Celia (CBER) <[redacted]>
Subject: Contact from Ana Szarfman

Dear Peter,

Ana Szarfman called me at about 4:15pm on Friday 7/9. She said that she and Bill DuMouchel had found an increased risk of mortality following COVID-19 vaccination using data mining methods. I asked her to send me the analysis and promised to review it, and I've attached the email she sent on Monday 7/12. It has very little information on the methods. I've pasted my reply to her below.

I am very concerned that whatever association they think they have identified is spurious based on the way the COVID-19 vaccination program prioritized individuals and the required and stimulated reporting we are seeing with the COVID-19 vaccines. Ana said that she had taken her name off a publication that is being prepared.

Please let me know how you would like us to proceed.

Best Regards,
--Rich

----------Response to Ana----------
Hi Ana,

Thanks for sharing this, and my team will review it. Do you have any more details on the new methods that Bill DuMouchel is using? That would be helpful in our evaluation.

In the email thread, Bill asked, "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."

During the time frame of this analysis, people with high-risk medical risk conditions were prioritized for vaccination. This included people in Nursing Homes early in the rollout. In later phases, most people needed to demonstrate that they had a high-risk medical condition to qualify for a vaccination. These included conditions like diabetes, chronic lung diseases, hypertension, heart conditions, obesity, and liver disease. Unless Bill has controlled for this somehow, I worry that the disproportionality in these results is caused by selection bias.

What are your thoughts on this possibility?

Best regards,
--Rich

----------End of Response----------

Richard Forshee, Ph.D.
Acting Deputy Office Director, CBER/OBE

Center for Biologics Evaluation and Research
Office of Biostatistics and Epidemiology
Analytics and Benefit-Risk Assessment Team
U.S. Food and Drug Administration
[redacted]

FDA U.S. FOOD & DRUG ADMINISTRATION

From: "Szarfman, Ana" <[redacted]>
To: "Forshee, Richard" <[redacted]>
Cc: "Stockbridge, Norman L" <[redacted]>, "Thompson, Aliza" <[redacted]>, "Unger, Ellis" <[redacted]>
Subject: RE: Findings of interest to the Division
Date: Fri, 11 Jun 2021 14:22:41 +0000
Importance: Normal
Attachments: CDER_Clearance_Request_for_Articles_Speeches_and_Other_Publications__Masking_Associated_with_Early_COVID-19_Vaccine_Safety_Surveillance.pdf; PDS_-_COVID19_safety_surveillance_and_masking.docx
Inline-Images: image001.jpg; image002.jpg

Hi Richard,

Please refer to the feedback by Ellis regarding the FDA/CDC AC on covid-19 vaccines held yesterday, and to the RGPS finding of AMI in the public domain data of 2 months ago, as well as to a methodological paper on RGPS that I am forwarding.

We would love to have you as a collaborator.

Many thanks, Ana

From: Unger, Ellis <[redacted]>
Sent: Friday, June 11, 2021 9:52 AM
To: Szarfman, Ana <[redacted]>
Cc: Stockbridge, Norman L <[redacted]>; Thompson, Aliza <[redacted]>
Subject: RE: Findings of interest to the Division

Impressive, Ana! Did you send this to Richard Forshee in CBER? He's involved (and may lead) CBER's post-marketing/pharmacovigilance office/division. He's mathematically oriented (I think he's a statistician by training). He'd be extremely interested in this.

From: Szarfman, Ana <[redacted]>
Sent: Friday, June 11, 2021 9:16 AM
To: Unger, Ellis <[redacted]>
Cc: bill.dumouchel ([redacted]) <[redacted]>; Stockbridge, Norman L <[redacted]>; Thompson, Aliza <[redacted]>
Subject: RE: Findings of interest to the Division

Hi Dear Ellis,

You may be already aware that the main topic of discussion of FDA/CDC AC on covid-19 vaccines held yesterday was myocardial events and the lack of signals in VAERS and other data resources.

I am not astonished that MGPS was unable to detect these signals.

In contrast, in the email that I am forwarding, that I originally send over a month ago, we documented that RGPS signals AMI. (We also detected clear signals for other similar events). Notice that we used public domain data posted about 2 months ago.

From: Szarfman, Ana
Sent: Thursday, May 6, 2021 6:52 PM
To: Stockbridge, Norman L <[redacted]>; Thompson, Aliza <[redacted]>
Cc: bill.dumouchel <[redacted]>
Subject: Findings of interest to the Division

Hi Norman and Aliza,

Two issues of interest to the Division:

First, A signal of Acute Myocardial Infarction with COVID-19 vaccines in VAERS that was unmasked by RGPS.

Refer to the first screen shot that I pasted below. The analysis was performed by Bill using public domain data.

Note that the ER05 signals of RGPS are higher than the EB05 signals of MGPS.

This is, of course an early signal that we are monitoring.

Second, the lower rate of coverage of the Pfizer vaccine against symptomatic infection in patients with CHF or CKD

Refer to the second screen shot.

Ran Balicer, the director of the COVID-19 vaccine in Israel described these findings during the Parallel Accelerator meeting of today. These are findings published in the New Engl J Med 2021. DOI: 10.1056/NEJMc2104281.

Vaccine Name	Event: PT	N	E	ER05	EB05	ROR05	ERAM	EBGM	ROR	PRR	PRR_B
COVID19 (COVID19 (JANSSEN))	**Acute myocardial infarction**	5	1.935	**0.97**	**0.533**	0.581	2.385	1.064	1.23	1.23	154
COVID19 (COVID19 (MODERNA))	**Acute myocardial infarction**	49	7.331	**5.037**	**2.212**	3.107	6.477	2.805	4.165	4.152	110
COVID19 (COVID19 (PFIZER-BIONTECH))	**Acute myocardial infarction**	61	6.712	**7.01**	**3.031**	4.699	8.767	3.76	6.192	6.165	98
COVID19 (COVID19 (JANSSEN))	Facial discomfort	3	0.698	**1.037**	**0.49**	0.591	3.615	1.139	1.575	1.575	43
COVID19 (COVID19 (MODERNA))	Facial discomfort	22	2.35	**5.942**	**1.93**	3.507	8.761	2.748	5.815	5.81	24
COVID19 (COVID19 (PFIZER-BIONTECH))	Facial discomfort	19	2.32	**5.031**	**1.632**	2.506	7.665	2.383	4.12	4.117	27
COVID19 (COVID19 (JANSSEN))	Illness	22	2.444	**5.85**	**1.908**	2.427	8.632	2.716	3.512	3.498	590
COVID19 (COVID19 (MODERNA))	Illness	48	11.368	**3.248**	**1.302**	1.447	4.189	1.654	1.88	1.878	564
COVID19 (COVID19 (PFIZER-BIONTECH))	Illness	129	9.406	**11.656**	**3.968**	6.186	13.557	4.607	7.466	7.418	483

Predictor	Event: PT	N	E	Coef	Base Prob	**Odds Ratio**
ANTHRAX (NO BRAND NAME)	**Acute myocardial infarction**	5	0.318	1.275		**3.579**
COVID19 (COVID19 (MODERNA))	**Acute myocardial infarction**	49	14.572	1.795		**6.019**
COVID19 (COVID19 (PFIZER-BIONTECH))	**Acute myocardial infarction**	61	13.442	2.0883		**8.071**
INFLUENZA (SEASONAL) (FLUZONE HIGH-DOSE)	**Acute myocardial infarction**	0	12.633	-1.9224		0.146
PNEUMO (PNEUMOVAX)	**Acute myocardial infarction**	1	17.929	-1.5909		0.204
SMALLPOX (ACAM2000)	**Acute myocardial infarction**	19	0.94	2.6097		**13.595**
ZOSTER (SHINGRIX)	**Acute myocardial infarction**	1	20.68	-1.8385		0.159
Group 1	**Acute myocardial infarction**	26	2.591	-6.8207	0.001090	
Group 2	**Acute myocardial infarction**	25	4.762	-6.9896	0.000921	
Group 3	**Acute myocardial infarction**	17	4.139	-7.3055	0.000671	
Group 4	**Acute myocardial infarction**	16	3.461	-7.7204	0.000443	
Group 5	**Acute myocardial infarction**	14	4.527	-7.5685	0.000516	
Group 6	**Acute myocardial infarction**	15	5.667	-8.2038	0.000274	
Group 7	**Acute myocardial infarction**	16	9.305	-8.1326	0.000294	
Group 8	**Acute myocardial infarction**	16	24.682	-9.0167	0.000121	
Group 9	**Acute myocardial infarction**	14	99.868	-10.5956	0.000025	

Predictor	Event: PT	N	E	Coef	Base Prob	Odds Ratio
COVID19 (COVID19 (MODERNA))	Facial discomfort	22	6.261	1.9273		**6.871**
COVID19 (COVID19 (PFIZER-BIONTECH))	Facial discomfort	19	6.199	1.7963		**6.027**
Group 1	Facial discomfort	17	2.337	-8.6942	0.000168	
Group 2	Facial discomfort	12	4.329	-9.1636	0.000105	
Group 3	Facial discomfort	10	5.673	-9.6851	0.000062	
Group 4	Facial discomfort	7	33.662	-11.2595	0.000013	

Predictor	PredType	N	E	Coef	Base Prob	Odds Ratio
COVID19 (COVID19 (JANSSEN))	Illness	22	6.599	1.9062		6.727
COVID19 (COVID19 (MODERNA))	Illness	48	26.941	1.3822		3.984
COVID19 (COVID19 (PFIZER-BIONTECH))	Illness	129	25.776	2.5381		12.656
HEP A (VAQTA)	Illness	0	18.081	-2.4797		0.084
HEP B (RECOMBIVAX HB)	Illness	0	20.674	-2.6091		0.074
HIB (PEDVAXHIB)	Illness	0	10.468	-1.976		0.139
HPV (GARDASIL 9)	Illness	1	23.284	-1.9985		0.136
HPV (GARDASIL)	Illness	1	36.322	-2.3904		0.092
INFLUENZA (SEASONAL) (FLUZONE HIGH-DOSE)	Illness	0	13.931	-1.7044		0.182
INFLUENZA (SEASONAL) (NO BRAND NAME)	Illness	369	22.201	3.5744		35.673
MEASLES + MUMPS + RUBELLA (MMR II)	Illness	2	27.081	-1.7004		0.183
SLES + MUMPS + RUBELLA + VARICELLA (PROQ	Illness	0	22.841	-2.6923		0.068
PNEUMO (PNEUMOVAX)	Illness	1	48.689	-2.8518		0.058
ROTAVIRUS (ROTATEQ)	Illness	0	27.787	-2.9018		0.055
VARICELLA (VARIVAX)	Illness	0	45.055	-3.3202		0.036
ZOSTER LIVE (ZOSTAVAX)	Illness	0	54.822	-3.4857		0.031
Group 1	Illness	322	75.702	-5.7714	0.003106	
Group 2	Illness	85	40.663	-7.256	0.000705	
Group 3	Illness	40	22.234	-7.4062	0.000607	
Group 4	Illness	34	33.26	-7.8602	0.000386	
Group 5	Illness	25	36.923	-8.5048	0.000202	
Group 6	Illness	45	82.613	-8.6076	0.000183	
Group 7	Illness	26	66.953	-8.878	0.000139	
Group 8	Illness	27	103.643	-9.5566	0.000071	
Group 9	Illness	8	150.009	-9.8717	0.000052	

Table 1. Estimated Vaccine Effectiveness in Subgroups.*

Subgroup and Period†	No. of Persons	Vaccine Effectiveness (95% CI)			
		Against Documented Infection	Against Symptomatic Infection	Against Hospitalization	Against Severe Disease
			percent		
Full study population					
14–20 days after first dose	693,814	48 (42 to 52)	56 (51 to 61)	70 (52 to 82)	65 (45 to 80)
21–27 days after first dose	480,438	65 (60 to 69)	71 (66 to 75)	78 (61 to 90)	77 (56 to 91)
7–28 days after second dose	310,696	93 (91 to 94)	96 (94 to 97)	92 (85 to 97)	95 (89 to 99)
Age group					
16–39 yr	115,916	94 (92 to 96)	97 (95 to 99)	NA	NA
40–69 yr	166,592	92 (90 to 94)	95 (93 to 97)	98 (94 to 100)	100 (35 vs. 0)‡
≥70 yr	28,318	91 (86 to 95)	92 (83 to 97)	81 (57 to 94)	86 (63 to 97)
No. of coexisting conditions					
0	179,448	94 (92 to 96)	97 (95 to 99)	98 (91 to 100)	100 (15 vs. 0)‡
1 or 2	103,792	95 (93 to 96)	95 (93 to 97)	92 (80 to 100)	95 (78 to 100)
≥3	27,478	81 (69 to 90)	88 (79 to 95)	88 (71 to 97)	93 (78 to 100)
Overweight§	79,440	94 (91 to 97)	96 (93 to 98)	97 (89 to 100)	100 (16 vs. 0)‡
Obesity§	40,164	93 (89 to 97)	93 (86 to 97)	97 (90 to 100)	100 (12 vs. 0)‡
Hypertension	35,056	85 (75 to 92)	90 (84 to 95)	94 (84 to 99)	93 (80 to 100)
Type 2 diabetes mellitus	20,370	86 (76 to 93)	86 (76 to 94)	85 (65 to 100)	91 (71 to 100)
Heart disease	9,144	82 (62 to 93)	80 (60 to 93)	89 (69 to 100)	97 (84 to 100)
Chronic kidney disease	8,212	79 (60 to 92)	80 (57 to 94)	76 (14 to 100)	74 (–40 to 100)
Neurologic disease	3,464	81 (37 to 100)	84 (42 to 100)	69 (–27 to 100)	100 (5 vs. 0)‡
Cerebrovascular disease	2,762	53 (–42 to 92)	75 (29 to 96)	85 (26 to 100)	91 (43 to 100)
Immunodeficiency	1,674	90 (49 to 100)	84 (19 to 100)	100 (2 vs. 0)‡	100 (1 vs. 0)‡

* The number of persons represents the number of persons who contributed follow-up time in each analysis. Estimates and 95% confidence intervals (CIs) were calculated as one minus the risk ratio for different outcomes over different time periods in the different subgroups. Confidence intervals were estimated with the use of the percentile bootstrap method with 500 repetitions. Estimates were calculated only for cells with more than 10 instances of an outcome across the two groups. NA denotes not available.

† In all the subgroups except the full study population, the analysis period was 7 to 28 days after the receipt of the second dose of vaccine.

‡ No events were recorded in the vaccinated group, which thus precluded the estimation of the confidence interval by the nonparametric percentile bootstrap method. We instead show the raw count of events in the unvaccinated group as compared with the vaccinated group.

§ Overweight was defined as a body-mass index (BMI; the weight in kilograms divided by the square of the height in meters) of 25.0 to 29.9, and obesity as a BMI of 30.0 to 39.9.

New Engl J Med 2021. DOI: 10.1056/NEJMc2104281

CDER Clearance Request for Articles, Speeches, and Other Publications	Date of Request 06/03/2021	If clearance is requested to meet a deadline, please provide deadline date. 06/08/2021

1. Person to Contact	2. Phone Number	3. Email address
Ana Szarfman	[redacted]	[redacted]

4. Title of Article, Speech, or Other Publication

Masking Associated with Early COVID-19 Vaccine Safety Surveillance

5. Authors

DuMouchel W.1, Harpaz R.1*, Szarfman A.2, Van Manen R.1, Nip A.1, Bright S.1, Al-Ansari M1.

6. Author Affiliations

1. Oracle Health Sciences, Bedford, MA, United States; 2. Division of Cardiology and Nephrology, OCHEN, OND, CDER, FDA

7. Details of Article or Speech

☑ **Journal Article**	☐ Regulatory Summary ☐ Letter ☐ Review Article ☐ Editorial ☑ Peer Reviewed Research	Journal 'Pharmacoepidemiology and Drug Safety'		
☐ **Book Chapter**	Chapter Title	Book Title		
☐ **Meeting Abstract Symposium/ Workshop**	Meeting Title	Meeting Sponsor	Meeting Date	Meeting Location
☐ **Speech**	Talk Title	Meeting Sponsor	Meeting Date	Meeting Location

Submitter Assurances

8. This article or speech was completed as: Assigned Work

9. The article or speech is a result of research involving human subjects, specimens, or subject level data (e.g., NDA data)? ○ Yes ◉ No

If yes – Please provide Research Involving Human Subjects Committee (RIHSC) Protocol Number

10. If the article or speech is reporting the results of CDER research, all of the methodological details, analytical procedures, and underlying source data supporting the conclusions are documented and available for inspection? ◉ Yes ○ No

11. To the best of my knowledge, this article or speech NOT contains non-public information.

12. To the best of my knowledge, the statements and conclusions in this article or speech conform with FDA policy. ◉ Yes ○ No

13. The following divisions and offices have reviewed this article or speech:

14. Submitter Signature (digital) Ana Szarfman -S

Digitally signed by Ana Szarfman -S
DN: c=US, o=U.S. Government, ou=HHS, ou=FDA, ou=People, cn=Ana Szarfman -S, 0.9.2342.19200300.100.1.1=1300048298
Date: 2021.06.03 14:28:16 -04'00'

Review and Clearance

		Name	Signature (digital)	Date
14. First Line Reviewer (optional - by office)	Free of non-public information Select Conforms with FDA policy Select			
15. Second Line Reviewer (optional - by office)	Free of non-public information Select Conforms with FDA policy Select	Name	Signature (digital)	Date
16. Clearance Official *Article or Speech is:*	Free of non-public information Select Conforms with FDA policy Select	Comments		

☐ **Cleared** ☐ **Not Cleared** ☐ **Returned for Revisions** Is a disclaimer required? Select One (If Yes, ☐ Disclaimer 1 ☐ Disclaimer 2)

Name	Signature (digital)	Date

CLEAR FORM SAVE FORM PRINT FORM

Masking Associated with Early COVID-19 Vaccine Safety Surveillance

DuMouchel W.[1], Harpaz R.[1], Szarfman A.[2], Van Manen R.[1], Nip A.[1], Bright S.[1], Al-Ansari M[1].*

[1] Oracle Health Sciences, Bedford, MA, United States
[2] U.S. FDA, Silver Spring, MD, United States
[*] Corresponding author

Abstract:

Purpose*: raise awareness to the problem of masking associated with early stage VAERS COVID-19 vaccine surveillance, which can lead to missed signals. Provide a preliminary investigation of these masking effects.*

Methods*: three signal detection methodologies: MGPS, PRR, and a new methodology called RGPS that can control masking were applied to six years of VAERS reports, consisting of 17 weeks of COVID-19 vaccine reports.*

Results*: several statistically masked associations are identified, some of which are also listed in the product labels and reported in clinical trials. The most extreme case of masking is statistically verified by removing reports containing the 'offending' masker. RGPS appears to provide a reasonable middle ground between MGPS and PRR with respect to the veracity and number of signals produced by each methodology.*

Conclusions*: statistical masking is present and should be considered in the context of early stage COVID-19 signal detection. RGPS can address masking and confounding effects, which cannot be properly controlled by conventional signal detection methodologies.*

Purpose:

As the world contends with rolling out massive scale vaccination programs to end the COVID-19 pandemic, identifying and studying adverse events related to these vaccines is critically urgent. The Vaccine Adverse Event Reporting System (VAERS), co-administered by the US Food and Drug Administration and the Centers for Disease Control and Prevention (CDC), is one of several systems used to monitor adverse events that occur after vaccination, including the COVID-19 vaccine. Like other safety surveillance systems, VAERS offers the opportunity to rapidly flag potential safety issues related to vaccines–a process usually known as signal detection.

Computational methodologies for signal detection have been routinely applied to safety surveillance systems for over 20 years and have become a de facto standard[1]. Given its impact on public health, signal detection is still an active area of research, and since its conception multiple guidances[2-5] have been published with practice recommendations as well as admonitions concerning data and methodological limitations.

In particular, 'masking' is a problem that may result in missed signals. Masking[3, 6, 7] is an artifact of conventional disproportionality statistics used for signal detection that are based on 2x2 contingency tables. A masked relationship between a target product and target adverse event can emerge when another product/s is frequently reported with the target event while making the background rate for the target event considerably large. This larger background rate can then make the relationship between the

target product and the target event appear less unusual, hence masking the true relationship. Conditions that make signal detection especially vulnerable to masking effects include: smaller volume of cases such as in VAERS, relationships involving rare events, and relationships involving newer products (i.e., emerging signals). As such, the early stages of COVID-19 vaccine surveillance makes signal detection especially susceptible to masking.

Masking effects can be ameliorated by removing cases containing the 'offending' product, by using stratification, and by employing regression techniques, all of which require to some extent identifying masking sources prior to signaling.

Regression-Adjusted GPS (RGPS)[8] is a signal detection methodology that is designed to combine the application simplicity of conventional signal detection methodologies with the power of regression to produce disproportionately statistics with adjusted background rates. Among other, these adjusted background rates can control masking effects.

The purpose of this report to raise awareness to masking in the context of early stage VAERS COVID-19 vaccine surveillance and provide a preliminary investigation of these masking effects using the RGPS methodology.

Methods:

Data: The investigation was done using VAERS reports from Jan. 01, 2015 to May 01, 2021–the latest public release of VAERS data available at the time of writing. This data represents 17 weeks of COVID-19 vaccine reports, and 6 years of other vaccine reports used as background. In total 466,401 reports were used. Of those 145,300 reports included the COVID-19 vaccine from three manufacturers: Pfizer/BionTech (39.6%), Moderna (45.9%), and Janssen (14.5%). Events were represented at the MedDRA Preferred Term (PT) level. To investigate adverse events commonly reported across the three manufacturers, products were represented at the 'vaccine type', e.g., 'COVID19' rather than at the 'manufacturer' level, e.g., 'COVID19_PFIZER/BIONTECH'.

Methods: The RGPS methodology operates by fitting separate Bayesian logistic regression models to each target adverse event. RGPS automatically selects two types of predictors to be included in each regression model: (1) products that are statistically associated with the target event, which are represented as indicator variables, and (2) stratification categories grouped by target event rates, which are represented as multiple regression intercepts. Rather than using the fitted regression coefficients to compute signal scores (disproportionalities), RGPS computes observed to expected ratios of counts similar to conventional methodologies. The expected counts are computed by summing the regression predicted probabilities of the target event across all reports mentioning the target product under the null hypothesis of no association between the target product and target event. The null hypothesis probabilities are computed by setting the coefficient of the target product to zero if selected as a model predictor. This results in adjusted expected counts (background rates) that can address masking. The final signal score (and its intervals) are computed using Bayesian shrinkage of observed to expected counts similar to MGPS[9]. Complete details of the RGPS methodology are presented in reference[8].

The results of RGPS were compared against those of the MGPS and the PRR[10] methodologies. Stratification categories used for RGPS and MGPS were: age (10 levels) and gender (3 values). Stratification by 'report year' could not be applied because COVID-19 VAERS reports dominate reporting from

December 2020. We applied the canonical version of PRR, which does not require stratification. For RGPS and MGPS we generated both the point estimates, labeled ERAM and 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 three methodologies and analysis thereof was done using Oracle Empirica Signal 9.1[11].

Results:

Figure 1 provides a high-level comparison of signal scores generated by RGPS, MGPS, and PRR for COVID-19 vaccine-related adverse events. The comparison is illustrated by means of sector (heat) maps.

The figure positions RGPS as a compromise between PRR and MGPS with respect to the number of signals produced by each methodology. Setting the signal score cutoff to 2, PRR produced 3,695 signals, whereas RGPS 467 signals, and MGPS 74. Setting the cutoff to 5, PRR produced 1,492 signals, whereas RGPS 74 signals, and MGPS only 7. The canonical version of PRR is generally known to produce more signals at the expense of possibly a larger number of false alerts compared to MGPS, but as a result may miss fewer signals than MGPS. Which of the methods provides a better tradeoff is still a subject of debate, and beyond the scope of this report.

We conjecture that the larger number of signals produced by RGPS compared to MGPS is due to masking. Both RG*PS and PRR show a 'hot' zone of cardiac events some of which were recently highlighted by the CDC[12].

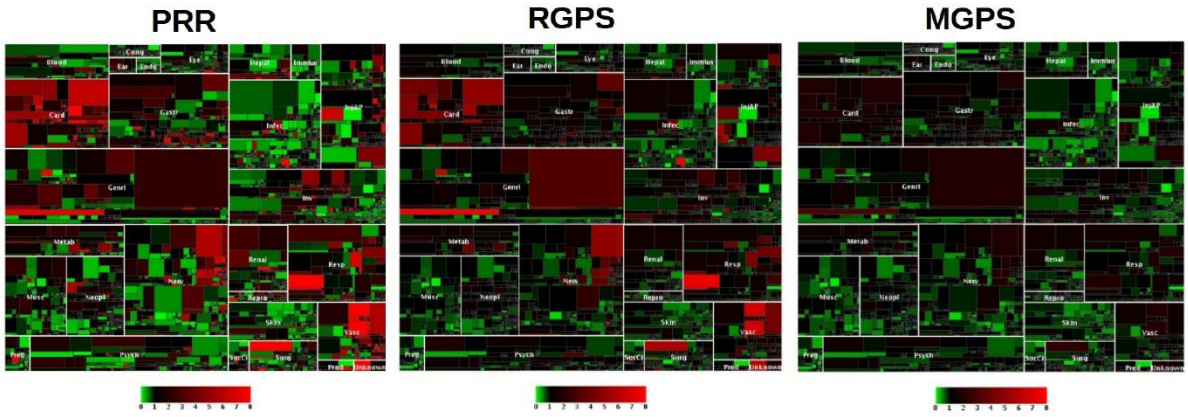


Figure 1. Sector-map comparison of signal scores generated by RGPS, MGPS, and PRR for COVID-19 vaccine-related adverse events. The major rectangular areas represent system-organ-classes (SOCs) and the smaller interior rectangles represent preferred terms (PTs). The box sizes represent the number of cases, and colors represent the size/scale of signal scores, ranging from green-smaller scores to red-larger scores.

Table 1 provides signal statistics for the top 20 masked COVID-19 vaccine associations flagged by RGPS. A candidate association for masking is defined as one whose signal statistics satisfy the following condition:

$$ER05 > EB95 \text{ and } ER05 > 1 \text{ and } EB05 < 1$$

That is, an association where RGPS and MGPS disagree by producing non-overlapping credible intervals (*ER05 > EB95*) with RGPS' interval above the boundary of no association (*ER05>1*) and that of MGPS below

or including the boundary of no association (*EB05 < 1*). The associations are ranked by the magnitude of the masking effect, which we define as the ratio of RGPS' and MGPS' signal scores, i.e., *ERAM/EBGM*. The table also provides signal scores for the IC statistic[13], which were close to those of MGPS.

Several of the adverse events listed in Table 1 have been listed in the product labels and reported in clinical trials[14], and others are known issues associated with vaccine administration. While the size of the signal scores in Table 1 would normally not be considered large enough to warrant immediate action, they nonetheless clearly illustrate the masking effect. Herpes zoster (shingles, 903 reports) was ranked highest in terms of the masking ratio equal to 1.20/0.31=3.91. PRR produced a signal score of 0.17 for the same association. RGPS automatically selected 39 product predictors and 16 stratification groups for the Herpes zoster regression model (the 'COVID19' vaccine type was not one of the predictors). The strongest predictor and likely the culprit for masking was the VARZOS vaccine--a combination vaccine of Varicella and Zoster. Upon removal of all reports containing the VARZOS vaccine (the alternative approach for controlling masking) the PRR and EBGM signal scores reverted to 1.76 and 1.12 respectively, supporting RGPS' finding of the offending masker. Herpes zoster has also been reported in an observation study[15].

Table 1.Top 20 masked COVID-19 vaccine associations flagged by RGPS. N: number of reports with the adverse event and COVID-19. E_RGPS: adjusted expected counts produced by RGPS. ERAM: signal score produced by RGPS. ER05/ER95: lower and upper bounds of RGPS' credible intervals. E_MGPS: adjusted expected counts produced by MGPS. EBGM: signal score produced by MGPS. EB05/EB95: lower and upper bounds of MGPS' credible intervals. PRR: signal score produced by PRR. 2^IC: 2 to the power of the signal score produced by the IC calculation. Masking ratio: ERAM/EBGM.

Event (MedDRA PT)	N	E_RGPS	ER05	ERAM	ER95	E_MGPS	EB05	EBGM	EB95	PRR	2^IC	Masking Ratio
Herpes zoster	903	752.2	1.14	1.20	1.27	2940.2	0.29	0.31	0.33	0.17	0.31	3.91
Numb chin syndrome	10	1.8	2.63	4.79	7.46	5.2	0.96	1.58	2.49	41.28	1.85	3.03
Cholecystitis acute	10	1.8	2.64	4.80	7.48	5.1	0.96	1.58	2.50	41.28	1.86	3.03
Influenza virus test negative	354	303.6	1.07	1.17	1.27	497.4	0.65	0.71	0.78	0.52	0.71	1.64
Pneumonia	612	462.3	1.24	1.32	1.41	712.6	0.80	0.86	0.92	0.60	0.86	1.54
White blood cell count increased	491	432.6	1.05	1.14	1.22	619.3	0.74	0.79	0.85	0.62	0.79	1.43
Underdose	468	345.8	1.25	1.35	1.46	488.1	0.89	0.96	1.03	0.70	0.96	1.41
Syringe issue	313	209.9	1.35	1.49	1.63	289.0	0.98	1.08	1.18	1.11	1.08	1.38
Oral herpes	247	188.0	1.18	1.31	1.45	259.1	0.86	0.95	1.06	1.28	0.95	1.38
Sepsis	192	127.5	1.33	1.50	1.68	175.3	0.97	1.09	1.23	0.82	1.10	1.37
Product administered to patient of inappropriate age	1132	908.9	1.19	1.25	1.31	1158.6	0.93	0.98	1.03	0.38	0.98	1.27
Abdominal discomfort	1262	1038.9	1.16	1.21	1.27	1272.1	0.95	0.99	1.04	1.41	0.99	1.22
Head injury	344	274.0	1.15	1.25	1.37	333.3	0.94	1.03	1.13	0.83	1.03	1.22
Seizure	884	818.0	1.02	1.08	1.14	986.4	0.85	0.90	0.95	0.42	0.90	1.21
Disturbance in attention	431	382.2	1.04	1.13	1.22	454.8	0.87	0.95	1.02	0.95	0.95	1.19
Pallor	1171	1029.9	1.08	1.14	1.19	1219.7	0.92	0.96	1.01	0.58	0.96	1.18
Vaccination site swelling	569	516.3	1.03	1.10	1.18	609.5	0.87	0.93	1.00	0.96	0.93	1.18
Syncope	2666	2292.0	1.13	1.16	1.20	2681.5	0.96	0.99	1.03	0.87	0.99	1.17
Rash pruritic	1838	1640.8	1.08	1.12	1.16	1880.6	0.94	0.98	1.02	1.39	0.98	1.15
Loss of consciousness	2259	1947.6	1.12	1.16	1.20	2213.4	0.99	1.02	1.06	0.97	1.02	1.14

Conclusion:

The novelty and early stages of COVID-19 vaccine surveillance, compounded by the size of the VAERS database predisposes signal detection to the notorious masking effect. Left unattended masking could lead to costly missed signals. This report demonstrates that statistical masking is indeed present and at the very least should be considered in the context of COVID-19 signal detection. The report also demonstrates the potential utility of a new signal detection methodology called RGPS that can address masking and confounding effects that cannot be properly controlled by conventional signaling methodologies. Signal detection is inherently exploratory, therefore the associations named in this report do not imply causal relationships.

Disclaimer:

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

Competing financial interests:

The authors declare no competing financial interests. DuMouchel W., Harpaz R., Van Manen R., Nip A., Bright S., Al-Ansari M. are employed by Oracle Health Sciences.

References:

1. Szarfman, A., S.G. Machado, and R.T. O'Neill, *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): p. 381-392.
2. Wisniewski, A.F.Z., et al., *Good Signal Detection Practices: Evidence from IMI PROTECT.* Drug Safety, 2016. **39**(6): p. 469-490.
3. Almenoff, J., et al., *Perspectives on the Use of Data Mining in Pharmacovigilance.* Drug Safety, 2005. **28**(11): p. 981-1007.
4. *CIOMS Working Group VIII. Practical Aspects of Signal Detection in Pharmacovigilance*. 2010: CIOMS.
5. *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.
6. 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): p. 195-207.
7. Juba, K.M., R.P. van Manen, and S.E. Fellows, *A Review of the Food and Drug Administration Adverse Event Reporting System for Tramadol-Related Hypoglycemia.* Ann Pharmacother, 2020. **54**(3): p. 247-253.
8. DuMouchel, W. and R. Harpaz, *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.
9. Dumouchel, W., *Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System.* The American Statistician, 1999. **53**(3): p. 177-190.

10. Evans, S.J.W., P.C. Waller, and S. Davis, *Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports.* Pharmacoepidemiology and Drug Safety, 2001. **10**(6): p. 483-486.
11. *Oracle Empirica Signal.* https://docs.oracle.com/en/industries/health-sciences/empirica-signal/9.1/index.html. 2021.
12. *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.
13. Bate, A., et al., *A Bayesian neural network method for adverse drug reaction signal generation.* Eur J Clin Pharmacol, 1998. **54**(4): p. 315-21.
14. *DailyMed, COVID-19 Vaccine Product Labels.* https://dailymed.nlm.nih.gov/dailymed/search.cfm?labeltype=all&query=COVID-19+vaccine. 2021.
15. Furer, V., et al., *Herpes zoster following BNT162b2 mRNA Covid-19 vaccination in patients with autoimmune inflammatory rheumatic diseases: a case series.* Rheumatology, 2021.

From: "Marks, Peter" <[redacted]>
To: "Anderson, Steven" <[redacted]>, "Forshee, Richard" <[redacted]>, "Szarfman, Ana" <[redacted]>, "Stockbridge, Norman L" <[redacted]>, "Dal Pan, Gerald" <[redacted]>, "Ball, Robert" <[redacted]>, "Witten, Celia (CBER)" <[redacted]>
Cc: "Jenkins, Charlene" <[redacted]>
Subject: FW: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running
Date: Sun, 14 Feb 2021 19:01:07 +0000
Importance: Normal
Attachments: NYT_--_As_Millions_Get_Covid_Vaccine_Shots,_F.D.A._Struggles_With_Safety_Monitoring.pdf
Inline-Images: image001.png

Dear Al,

Based on Ana's incoming, Charlene will set up an hour sometime in the next few weeks for us to discuss her concerns and recommendations. Thanks.

Best Regards,
Peter

From: Szarfman, Ana <[redacted]>
Sent: Sunday, February 14, 2021 9:31 AM
To: Marks, Peter <[redacted]>
Cc: Stockbridge, Norman L <[redacted]>
Subject: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running

Hi Peter.

Thanks so much for your work.

Regarding the NYT article, I am quite concerned that the distributed network of EHRs that the Sentinel System uses does not have mortality data, and that they (and we) are not clamoring to fix this problem. It still takes several years to collect mortality data. Indeed there are no incentives to update electronic health records with mortality data.

https://www.nytimes.com/2021/02/12/health/covid-vaccine-how-safe.html?referringSource=articleShare

→ Let me know if you want Bill DuMouchel and I to discuss our proposal for more effective monitoring, including the need to use an updated algorithm by DuMouchel for data mining spontaneous reports at CDER and CBER.

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]

FDA U.S. FOOD & DRUG ADMINISTRATION

As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running

For now, the government has been relying on a patchwork of programs that officials say are hampered by limited size and gaps in data collection.

A drive-through mass vaccination site at Coors Field baseball stadium in Denver last month. Chet Strange/Agence France-Presse — Getty Images

By Sheila Kaplan

Feb. 12, 2021

More than 35 million Americans have received Covid vaccines, but the much-touted system the government designed to monitor any dangerous reactions won't be capable of analyzing safety data for weeks or months, according to numerous federal health officials.

For now, federal regulators are counting on a patchwork of existing programs that they acknowledge are inadequate because of small sample size, missing critical data or other problems.

Clinical trials have shown both of the vaccines authorized in the United

States — one from Pfizer-BioNTech and the other from Moderna — to be highly protective and safe against the coronavirus.

But even the best trials have limited ability to detect adverse reactions that are rare, those that only occur in certain population groups, or which happen beyond the three-month period studied in the trials. Tracking adverse events once the vaccines are administered to the public at large is essential not just to detect problems but to build confidence in the safety of vaccines.

In interviews, F.D.A. officials acknowledged that a promised monitoring system, known as BEST, is still in its developmental stages. They expect it to start analyzing vaccine safety data sometime soon — but likely not until after the Biden administration reaches its goal of vaccinating 100 million people.

“I’m concerned about this disjointed tracking system,” said Dr. Ashish K. Jha, dean of the Brown University School of Public Health. “We knew these vaccines were coming for at least several months before they got authorized, so we really should have had a well-developed system.”

Dr. Jha and others believe that with all the public attention on the vaccines, any serious adverse reactions will likely be reported somewhere. But, they say, a more systematic approach is crucial.

“It’s critical to track, because it will help build confidence,” Dr. Jha said.

Monitoring is all the more important because the vaccines were developed and approved in record time, with the goal of inoculating most of the U.S. population as quickly as possible.

“It’s the right thing to do, but the fact of the matter is we don’t have enough information and we’re desperately in need of post-market information and monitoring,” said a high-ranking F.D.A. official, who asked not to be named because he was not authorized to discuss the matter publicly.

The government is now relying most on a 30-year-old safety monitoring

system that the F.D.A. shares with the Centers for Disease Control and Prevention, and a new smartphone app that people who get vaccinated can download and use to report problems if they wish. The C.D.C. also runs the Vaccine Safety Datalink, a collaboration between the agency and nine health systems that collects vaccine data and electronic medical records of roughly 12 million patients. Although it is well-regarded, it is of limited use because of its small size.

Boxes of the Pfizer-BioNTech vaccine were prepared for shipment at a facility in Portage, Mich., in December. Pool photo by Morry Gash

“It’s great for routine stuff, but when it comes to safety surveillance, it’s all about size,” said Dr. Daniel Salmon, director of the Institute for Vaccine Safety at Johns Hopkins University, and a former federal vaccine official. “The bigger it is, the faster you get an answer. Eventually the VSD will get a really good answer — probably one of the best answers of anybody out there because they are so good at doing it. But in a pandemic, time isn’t on our side.”

So far, few serious problems have been reported through these channels and no deaths have conclusively been linked to the vaccines. The 30-year-old initiative, known as the Vaccine Adverse Event Reporting System, or VAERS, relies on self-reported cases from patients and health care providers.

Health officials say that so far, the two vaccines already authorized for use appear to be quite safe. There have been a few severe allergic reactions, including anaphylaxis, but they are treatable and considered rare. The rate at which anaphylaxis has occurred so far — 4.7 cases in every million doses of the vaccine by Pfizer and BioNTech, and 2.5 cases per million for the vaccine by Moderna — are in line with what happens with other widely used vaccines.

The Coronavirus Outbreak ›

Latest Updates ›

Updated

The televangelist Frederick K.C. Price has died at 89 of complications from Covid-19.

F.D.A. officials say their 'flawed' policy led to a flood of unreliable antibody tests early in the pandemic.

Peru has a new health minister after a vaccine scandal forced the previous one to resign.

Bruising and bleeding caused by lowered platelet counts have also been reported, though it is not known if they are linked to the vaccines, or coincidental. In total, 9,000 adverse events were reported, with 979 serious and the rest classified as nonserious, according to the most recent C.D.C. report available.

In interviews, public health experts, including current and former officials at the F.D.A. and the C.D.C., expressed a need to improve upon old "passive" surveillance, which depends on self-reporting. They said that

funding shortages, turf wars and bureaucratic hurdles had slowed preparing BEST, formally called the Biologics Evaluation Safety Initiative, to monitor the Covid vaccines.

An earlier version of BEST was started in 2017, to improve the F.D.A.'s tracking of new blood products and vaccines, but the agency has only used it on a limited basis. It is considered an "active" surveillance system because scientists can use data collected from clinical care to hunt for safety problems, rather than rely on individuals to report health problems that they believe — but often without proof — were caused by the vaccine. BEST is part of the agency's move toward using more real-world evidence to vet new products or monitor them after approval. The F.D.A. has done some preliminary studies using BEST to evaluate the safety of shingles and flu vaccines.

When the monitoring system is fully up and running, the F.D.A. expects to have access to more than 100 million individual medical records, and will be able to look for signs of safety problems, and then determine whether they are real. But critics say it is folly for the F.D.A. to be launching a new system in the midst of a pandemic. And several C.D.C. officials said the F.D.A. was not giving them a real sense of when the complex system would begin to work.

"It's been a puzzle to me," said one C.D.C. official who was not authorized to discuss the issue and asked not to be identified. "F.D.A. talks about this in a way that is really unclear as to what is up and ready to go and what isn't."

The headquarters of the F.D.A. in Silver Spring, Md. Jim Lo Scalzo/EPA, via Shutterstock

But even BEST will suffer from a data problem that is already hindering existing systems: the dearth of health insurance claims to show who got which vaccine, and when. Typically health care providers and patients submit such claims to insurers, but with the vaccines being given at no charge, often at government-sponsored events, few are bothering to file claims. Critics say that federal health officials should have predicted this glitch and prepared for it.

"The current safety surveillance system in the U.S. is dependent on health insurance claims data and electronic health records," said Dr. Salmon. "If the vaccine data information doesn't get into the safety system, then that safety system is unable to function."

In December, the C.D.C. launched V-safe, a smartphone-based system that checks in with individuals who get the Covid vaccine to monitor for side

effects. Roughly two million people who have been vaccinated have enrolled, a small fraction of the total number, and of those, one million have responded to text queries and surveys about their post-vaccine health.

At a recent C.D.C. advisory meeting, Dr. Tom Shimabukuro, who oversees Covid-19 vaccine safety for the agency, said he was pleased that the new app had enrolled so many users, but he also acknowledged problems like errors that indicated men and older women to be listed as pregnant.

It's also unclear how heartily vaccine providers are promoting V-safe. Some health care providers send post-vaccine emails to patients noting its availability, and others merely put a stack of C.D.C. fact-sheets about V-safe in the vaccination room and hope patients pick it up. Even Dr. Jha said he didn't sign up for it.

Still, Dr. Shimabukuro said he was confident in the current surveillance system. "For the national Covid-19 vaccination program, we have implemented the most intense safety monitoring in the history of the United States," he said. "We have multiple systems that are complementary to each other, that are able to rapidly collect information, that are able to rapidly assess the safety of immunizations."

Medical workers filled doses of Moderna's vaccine at a a drive-through site in Robstown, Texas. Go Nakamura/Reuters

One factor slowing down BEST is that the F.D.A. has not yet calculated what are called background rates, the levels of certain health problems that normally occur in the non-vaccinated population. These are critical for determining whether the vaccine is actually causing a spike in certain problems, such as heart attacks, strokes, and other issues that the F.D.A. and C.D.C. consider adverse events of special interest, which require close monitoring.

Rather than calculate them on its own, as the C.D.C. does, the F.D.A. sent a proposal out for public comment, in which it detailed how it planned to compute the background rates. They plan to start working on it in the next few weeks. This delay strikes some public health experts as unnecessary.

“It’s a little bit surprising,” said Dr. Peter Lurie, president of the Center for Science in the Public Interest, and a former associate commissioner at the F.D.A. “That doesn’t feel like a mechanism appropriate to the urgency of a pandemic. It seems to me that a few well-placed phone calls to key people in the field would provide as much information as a request for comment.”

Dr. Peter Marks, the director of the F.D.A.’s Center for Biologics Evaluation and Research, which oversees vaccine approval and safety, said the agency needed outside input.

“The background rates are a critical input for our rapid cycle analysis, so we followed a deliberative and transparent process,” he said in an interview. “We needed to develop an approach that could be used in several health care claims data systems and we needed to account for the possibility that health care utilization may have changed during the pandemic.”

Jeffrey Brown, an associate professor at Harvard Medical School and a leader of the F.D.A. program that monitors adverse reactions to drugs, said he is concerned about the lack of insurance claims data and other holes in the vaccine safety surveillance systems.

“It is imperative to have policies that ensure vaccination data are submitted to insurers to enable effective use of the nation’s investment in active safety

monitoring," said Dr. Brown. "It is not only critical to get needles into arms, but also to get data into databases. We still have a chance to get it done well."

Denise Grady contributed reporting.

Jessica Hill/Associated Press

From: "Forshee, Richard" <[redacted]>
To: "Forshee, Richard" <[redacted]>
Subject: FW: Summary of DE issue and concerns
Date: Wed, 31 Aug 2022 12:13:26 +0000
Importance: Normal

From: Anderson, Steven <[redacted]>
Sent: Thursday, September 16, 2021 12:33 PM
To: Marks, Peter <[redacted]>; Forshee, Richard <[redacted]>
Subject: RE: Summary of DE issue and concerns

Dear Peter,

Thank you for raising this with CDER we appreciate the help with this. We'll keep you updated on any further developments with Ana.

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: Marks, Peter <[redacted]>
Sent: Thursday, September 16, 2021 4:54 AM
To: Anderson, Steven <[redacted]>; Forshee, Richard <[redacted]>
Subject: FW: Summary of DE issue and concerns

Dear Steve and Rich,

FYI in follow up.

Best Regards,
Peter

From: Stein, Peter <[redacted]>
Sent: Wednesday, September 15, 2021 8:49 PM
To: Marks, Peter <[redacted]>

Cc: Cavazzoni, Patrizia <[redacted]>
Subject: FW: Summary of DE issue and concerns

Peter, thanks for flagging this – we've made it clear to her that she should not be discussing or providing internal analyses externally, and needs to focus on her assigned work. My apologies that this was thrown into the mix at a challenging time – from Steve's note, it's evident that your team is aware of the issues, and planning on looking into the approach your taking at an appropriate time.

Hopefully, you won't have further surprises….
Peter

From: Marks, Peter <[redacted]>
Sent: Wednesday, September 15, 2021 4:55 AM
To: Cavazzoni, Patrizia <[redacted]>
Cc: Walinsky, Sarah <[redacted]>
Subject: FW: Summary of DE issue and concerns

Dear Patrizia,

I really am sorry to bother you with this, but issue is become a major distraction. One of the CDER statisticians, Ana Szarfman, has decided on her own to do vaccine analyses using VAERS as part of her work at FDA. She is, however, not doing this in collaboration with our CBER statisticians, and quite to the contrary, has been asked to cease and desist, because the strategy that she is using could create erroneous conflicts that feed in to anti-vaccination rhetoric. This is creating an issue, as documented below by our office director, Steve Anderson.

This issue came up previously during the pandemic and working with Gerald it seemed to go away, but it is now back. Can we catch up about this sometime?

Thanks,
Peter

From: Anderson, Steven <[redacted]>
Sent: Tuesday, September 14, 2021 10:59 PM
To: Marks, Peter <[redacted]>
Cc: Forshee, Richard <[redacted]>; Nair, Narayan <[redacted]>
Subject: Summary of DE issue and concerns

Dear Peter,

Below is a summary of an issue and concerns expressed by the OBE Division of Epidemiology and IOD:

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.

As background, when the CBER Division of Epidemiology leadership began planning its approach to passive surveillance for the COVID-19 vaccines in the summer of 2020, the overarching strategy was to build on existing, established systems

whenever possible. With regard to Data Mining, DE feels it is important to utilize the standardized, established system that has been in use for other vaccines for the past several years. The Division is concerned that use of a novel approach that had not been validated (and to our knowledge, has not been adopted by other medical product centers) would add another layer of uncertainty in the context of an EUA during the COVID-19 pandemic when rapid retrieval and interpretation of data would be imperative. DE recognizes as with all passive surveillance our current data mining process has limitations. In particular, DE is 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.

DE is planning to re-evaluate the data mining approach (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 DE believes 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). This retrospective approach is more downstream than what Ana is proposing but would be preferable to shifting midstream.

Let us know if you have questions or wish to discuss. Sorry for the hassles with this.

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: "Szarfman, Ana" <[redacted]>
To: "Califf, Robert" <[redacted]>, "Robert Califf, M.D." <[redacted]>
Subject: Interesting paper with very similar ideas to our paper
Date: Thu, 4 Aug 2022 01:14:43 +0000
Importance: Normal
Attachments: Perspective_Hamburg_et_al_nejmp2207374_(2).pdf
Inline-Images: image001.png

Good Night Dear Dr. Califf,

I wanted to let you know that today, the NEJM published an interesting perspective paper with very similar ideas to our paper in Nature pj Journal.

Please refer to the attachment and link below:

Building a National Public Health System in the United States | NEJM
https://www.nejm.org/doi/full/10.1056/NEJMp2207374?query=TOC&cid=NEJM%20eToc,%20August%204,%202022%20DM1314883_NEJM_Non_Subscriber&bid=1099325007

Ana

From: Califf, Robert <[redacted]>
Sent: Sunday, July 17, 2022 12:13 PM
To: Szarfman, Ana <[redacted]>
Subject: Re: New papers FYI

Thanks. These are good.

rmc

From: "Szarfman, Ana" <[redacted]>
Date: Sunday, July 17, 2022 at 8:38 AM
To: FDA Commissioner <[redacted]>, "Robert Califf, M.D." <[redacted]>
Subject: New papers FYI

Good Morning Dear Dr. Califf,

From the trenches:

Please refer to our first attached paper (***currently under an embargo by Nature pj Communications Medicine***) entitled: ***Recommendations for achieving interoperable and shareable medical data in the USA***, that will be published tomorrow Mon July 18, 2022, and be available at the following site: https://www.nature.com/articles/s43856-022-00148-x.

I have previously sent you a draft copy of our recommendations.

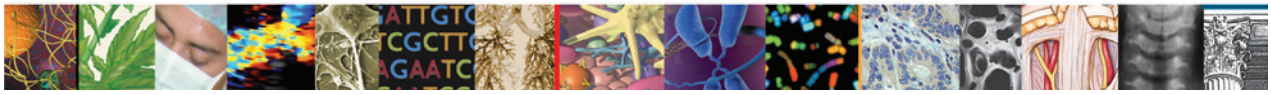
Please also note attached another paper recently published describing advances in the data mining (DM) methodology, especially the capacity to unmask hidden signals due to previously unaccountable confounders in the denominators available in the following site: https://rdcu.be/cQhDH.

I think this new data mining technology by Bill DuMouchel can be applied to address some additional pressing needs besides safety. They include to understand and quickly address the data quality problems generated by too prevalent mapping and remapping routines w/o traceability to the factual data.

Warmest regards,

--Ana

Ana Szarfman, MD, PhD, FAMIA, Medical Officer
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)
Division of Cardiology and Nephrology, OCHEN, Center for Drug Evaluation and Research, Food and Drug Administration
[redacted]



The NEW ENGLAND JOURNAL of MEDICINE

Perspective

AUGUST 4, 2022

Building a National Public Health System in the United States

Margaret A. Hamburg, M.D., Mandy Cohen, M.D., M.P.H., Karen DeSalvo, M.D., M.P.H., Julie Gerberding, M.D., M.P.H., Joneigh Khaldun, M.D., M.P.H., David Lakey, M.D., Ellen MacKenzie, Ph.D., Sc.M., Herminia Palacio, M.D., M.P.H., and Nirav R. Shah, M.D., M.P.H.

Over the past 2 years, as 1 million lives have been lost in the United States, the coronavirus pandemic has laid bare the shortcomings of the country's haphazard approach to public health.[1] Recent reports on the pandemic response have identified major gaps in leadership, coordination, communications, testing, and attention to critical issues of equity.[2,3]

These problems are not new; analyses after the 2001 anthrax attacks, the 2009 H1N1 influenza pandemic, and the 2015 Ebola outbreak also identified weaknesses. Major challenges are not limited to infectious diseases; the spiraling epidemic of overdose deaths, growing burden of diabetes, and dramatic increases in maternal mortality, particularly among Black women, also reflect the inadequacy of the public health enterprise. According to the World Bank, the United States ranks below more than 60 other countries in life expectancy, with major disparities according to race, ethnicity, and geography.

What will it take for the country to do better? In March 2022, the Commonwealth Fund convened the nine of us as the Commission on a National Public Health System to propose urgent, necessary, and realistic reforms. Over 90 days, we reviewed reports and recommendations spanning the past two decades, consulted with dozens of stakeholder groups, experts, and government officials, and reviewed more than 100 public comments.

It was a clarifying process. We heard time and again that readiness is much more than a plan on the shelf and countermeasures in a storage facility. Effective responses depend on strong routine public health efforts, grounded in a core set of capabilities that protect health and save lives, every day. Moreover, there was broad agreement that the United States can no longer rely on a discoordinated collection of nearly 3000 public health agencies without consistent standards and accountability for health improvement.

Accordingly, in our final report published on June 21, 2022,[4] we call for the development of a national public health system to promote and protect the health of all people, regardless of who they are and where they live; to implement effective prevention and response strategies with partners in the public and private sectors; and to earn public trust.

Immediate Steps to Build a National Public Health System.

For Congress:
- Establish a position of undersecretary or deputy secretary for health to provide leadership and accountability for developing the national public health system.
- Provide adequate and reliable funding to states, localities, tribes, and territories to support building the core public health capabilities in all health departments, in exchange for results, certified through accreditation.
- Provide the necessary funding to support a modern public health information technology system and provide the Department of Health and Human Services and its agencies the necessary authority to establish and enforce standards and implementation for that system.
- Target funding already appropriated ($3.7 billion from the American Rescue Plan Act for fiscal year 2022) for public health workforce development and infrastructure as down payments toward building core public health capabilities and modern information systems throughout the country.
- Direct the Centers for Medicare and Medicaid Services to start regularly mapping and sharing deidentified data to inform communities about health conditions and identify inequities and to start providing incentives for sharing health care data at the state level.
- Direct efforts to improve and modernize the accreditation for health departments.
- Accelerate efforts to modernize public health communications, including by expanding on the effort led by the Office of the Surgeon General to counter misinformation and disinformation.

For states, localities, tribes, and territories:
- Examine their health departments' abilities to meet core public health capabilities and, where health departments are very small, consider regional service-sharing approaches to providing essential public health protections.
- Create requirements for health care and public health to work together to support achievement of critical public health goals.
- Actively engage community residents in decision making regarding public health priorities.

For health care organizations:
- Jointly conduct community needs assessments with public health agencies and implement needed follow-on activities.
- Implement mechanisms to share data with public health agencies in support of community health improvement.
- Develop opportunities for cross-training and exchanges with public health.

For community organizations:
- Participate in planning activities of health agencies, including by setting key priorities for funding.
- Partner with local health agencies to provide accurate and effective messaging, including that needed to counter misinformation and disinformation.
- Demand ethical standards, integrity, and transparency from health agencies at all levels.

Building this national public health system starts with federal leadership. The Department of Health and Human Services (HHS) is home to multiple agencies with major roles in public health, including the Centers for Disease Control and Prevention, but their efforts are often insufficiently coordinated. Moreover, no single office or person has the dedicated responsibility for public health. HHS also lacks the authority to require collection of information essential to monitoring threats to health and has little flexibility to call upon and move resources urgently.

Addressing these gaps should be an urgent national priority, and legislation introduced by Senators Patty Murray (D-WA) and Richard Burr (R-NC) is an opportunity to move forward rapidly (see box). Among other steps, we argue that Congress should establish a new undersecretary or deputy secretary for public health in HHS. This empowered role would be a focal point for accountability and responsibility, coordinating the responses of the nation's health agencies to major, ongoing public health challenges, as well as leading the long-overdue modernization of public health data and surveillance systems, workforce, and laboratories.

Under the U.S. Constitution, states have the primary responsibility for protecting the health of the public. At the same time, there is a strong federal interest in states doing this job well, especially since threats in one area easily spread to others. Thus, Congress should provide adequate and reliable support for the health departments of states, localities, tribes, and territories. Upgrading public health agencies, including their workforce and information technology, would cost approximately $8 billion more annually than current expenditures, with potential savings as illnesses are averted and crises mitigated; the pandemic is estimated to have cost the U.S. economy more than $16 trillion.[5]

As former federal, state, and local health officials, we recognize that in exchange for this new funding, there must be standards and expectations to meet these standards. HHS should condition public health infrastructure grants and flexibility in the use of federal funds on progress toward meeting core public health capa-

bilities, as defined by the Public Health National Center for Innovations and as certified through a revised accreditation process that builds on the work of the Public Health Accreditation Board. Such an approach would encourage states to consider regional service sharing or consolidating some of our 2800 local health departments. Some states, including Indiana and Missouri, have already begun to reassess their organization of public health efforts.

During the pandemic, many health care delivery systems and clinicians rallied to support local public health efforts, sharing workforce, communications platforms, and data. Such collaborations should not disappear as the pandemic recedes, only to be reinvented for the next crisis. Rather, they should pivot to day-to-day health challenges, such as preventing complications of chronic illness, addressing urgent mental health concerns, or identifying emerging outbreaks of disease. Integration of the health care system in public health efforts will require both new funding and heightened accountability tied to ongoing federal support.

Greater sharing of health care data with public health agencies can start today. The Centers for Medicare and Medicaid Services can regularly share deidentified data and maps to inform local public health efforts. Then, with new authority from Congress, HHS can condition new infrastructure funding on states' developing and using near-real-time, statewide, all-payer databases, which can support collaborative efforts such as tracking cases of childhood asthma. HHS should also be able to require standardized reporting on hospital resources and supplies, such as bed and ICU capacity and availability of personal protective equipment, ventilators, and staffing. These capabilities, which can inform local efforts every day, are the foundation for robust responses to disasters.

Mobilizing the health care workforce for public health is another major opportunity. HHS can support public health training for staff in community health centers and health systems, as well as opportunities for rotating through state and local health departments. With more than 80 million people covered, Medicaid and the Children's Health Insurance Program present a special opportunity for joint action. The staff of state Medicaid offices and managed care plans can work closely with public health agencies, including through contractual arrangements, to improve health outcomes.

In our deliberations, we came to recognize that new structures and policies — however important — are limited in what they can accomplish. To succeed, public health agencies must earn and maintain public trust. The pandemic has revealed two profound challenges: long-standing suspicion on the part of many people who experience racism, discrimination, and marginalization in health care and public health, and strident opposition from those who rejected evidence-informed restrictions that were imposed to reduce the risk of illness and death.

We recognize that addressing these gaps requires a long-term commitment. Public health agencies will have to work closely with community residents, fund local organizations, and share decision making regarding local priorities. This work should involve collaborations with businesses and agencies in other sectors to address fundamental drivers of poor health, which communities often prioritize over traditional public health activities.

To blunt the corrosive effects of misinformation and disinformation, HHS should lead a major upgrade of public health communications on multiple platforms and for multiple audiences. To help the public understand the reasons for difficult decisions, HHS should also develop and implement model standards for ethics, transparency, and integrity.

It will be no small task to pull the U.S. approach to public health into the 21st century, but there are outsized reasons to do so. Establishing a national public health system will save lives from ongoing health challenges, protect our economy during future public health crises, and respect the extraordinary sacrifices of pub-

It will be no small task to pull the U.S. approach to public health into the 21st century, but there are outsized reasons to do so.

lic health and health care workers during the pandemic. If not now, when?

Disclosure forms provided by the authors are available at NEJM.org.

From the Commonwealth Fund Commission on a National Public Health System, New York (M.A.H., M.C., K.D., J.G., J.K., D.L., E.M., H.P., N.R.S.); Aledade (M.C.) and the Foundation for the National Institutes of Health (J.G.), Bethesda, and Johns Hopkins Bloomberg School of Public Health, Baltimore (E.M.) — all in Maryland; Google, Mountain View (K.D.), and the Clinical Excellence Research Center, Department of Medicine, Stanford University, Palo Alto (N.R.S.) — both in California; CVS Health, Woonsocket, RI (J.K.); the Department of Medicine, University of Texas at Tyler Health Science Center, Tyler (D.L.); the Guttmacher Institute, Washington, DC (H.P.); and American Health Associates, Davie, FL (N.R.S.).

This article was published on June 21, 2022, at NEJM.org.

1. Wallace M, Sharfstein JM. The patchwork U.S. public health system. N Engl J Med 2022;386:1-4.
2. DeSalvo K, Hughes B, Basset M, et al. Public health COVID-19 impact assessment: lessons learned and compelling needs. Washington, DC: National Academy of Medicine. April 7, 2021 (https://nam.edu/public-health-covid-19-impact-assessment-lessons-learned-and-compelling-needs/).
3. Oladele CR, McKinnsey TL, Tolliver D, Tuckson R, Dawes D, Nunez-Smith M. The state of Black America and COVID-19: a two-year assessment. Black Coalition Against Covid, 2022. (https://blackcoalitionagainstcovid.org/the-state-of-black-america-and-covid-19/).
4. Hamburg MA, Cohen M, DeSalvo K, et al. Meeting America's public health challenge: recommendations for building a national public health system that addresses ongoing and future health crises, advances equity, and earns trust. New York: The Commonwealth Fund. June 21, 2022 (https://www.commonwealthfund.org/publications/fund-reports/2022/jun/meeting-americas-public-health-challenge).
5. Cutler DM, Summers LH. The COVID-19 pandemic and the $16 trillion virus. JAMA 2020;324:1495-6.

DOI: 10.1056/NEJMp2207374

A Preview of the Dangerous Future of Abortion Bans — Texas Senate Bill 8

Whitney Arey, Ph.D., Klaira Lerma, M.P.H., Anitra Beasley, M.D., M.P.H., Lorie Harper, M.D., M.S.C.I., Ghazaleh Moayedi, D.O., M.P.H., and Kari White, Ph.D., M.P.H.

When the U.S. Supreme Court issues its decision in *Dobbs v. Jackson Women's Health Organization*, the abortion care landscape will most likely be changed for at least a generation. Even before a draft opinion was leaked, many experts anticipated that the Court would overturn *Roe v. Wade*, and nearly half the states are poised to ban or dramatically limit abortion care when that occurs.[1] These state laws criminalizing abortion may allow for very narrow exemptions, and anyone who violates the law could be subject to civil penalties, criminal fines, or imprisonment.[2]

Health systems and clinicians planning their responses[3] can look to Texas, where we have already witnessed the impact of strict abortion bans on the provision of evidence-based, essential health care for pregnant people. Since September 1, 2021, Texas Senate Bill 8 (SB8) has prohibited abortions after the detection of embryonic cardiac activity, which occurs around 6 weeks after a person's last menstrual period. After that point, SB8 allows abortions only in physician-documented medical emergencies. Anyone suspected of violating the law or aiding and abetting a prohibited abortion can face a civil lawsuit with monetary penalties of at least $10,000.

We interviewed 25 clinicians from across Texas about how SB8 has affected their practice in general obstetrics and gynecology, maternal and fetal medicine (MFM), or genetic counseling. We concurrently interviewed 20 Texans who had medically complex pregnancies and sought care either in Texas or out of state after September 1, 2021. Although aimed at clinicians who provide abortion care, SB8 has had a chilling effect on a broad range of health care professionals, adversely affecting patient care and endangering people's lives.

Some Texas clinicians still provide abortion counseling and referrals, believing that the law does not limit their free speech, while also noting that such freedom depends on a clinician's willingness to assume possible legal risk. On the basis of legal guidance, other Texas clinicians believe they are not even allowed to counsel patients regarding the availability of abortion in cases of increased maternal risks or poor fetal prognosis, although before SB8 they would have done so. Many clinicians have also been advised that they cannot provide

From: "Anderson, Steven" <[redacted]>
To: "Marks, Peter" <[redacted]>, "Forshee, Richard" <[redacted]>
Subject: RE: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running
Date: Mon, 01 Mar 2021 20:48:02 -0000
Importance: Normal
Inline-Images: image002.png

Dear Peter,

Sounds good. We will talk then.

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]

From: Marks, Peter <[redacted]>
Sent: Monday, March 1, 2021 3:45 PM
To: Forshee, Richard [redacted]>; Anderson, Steven <[redacted]>
Subject: RE: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running

Dear Steve and Rich,

Thanks for participating. I look forward to catching up at our 2:2 on how to proceed.

Best Regards,
Peter

From: Szarfman, Ana <[redacted]>
Sent: Monday, March 1, 2021 2:08 PM
To: Marks, Peter <[redacted]>; Anderson, Steven <[redacted]>; Forshee, Richard <[redacted]>; Dal Pan, Gerald <[redacted]>; Witten, Celia (CBER) <[redacted]>; Stockbridge, Norman L <[redacted]>
Subject: RE: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running

Hi All,

I enjoyed the opportunity to have this discussion with you.

I have attached an updated presentation and the WSJ article on IBM retreating from Watson.

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]

U.S. FOOD & DRUG ADMINISTRATION

-----Original Appointment-----
From: Marks, Peter <[redacted]>
Sent: Wednesday, February 17, 2021 6:50 PM
To: Marks, Peter; Anderson, Steven; Forshee, Richard; Szarfman, Ana; Dal Pan, Gerald; Ball, Robert; Witten, Celia (CBER); Stockbridge, Norman L
Subject: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running
When: Monday, March 1, 2021 1:00 PM-2:00 PM (UTC-05:00) Eastern Time (US & Canada).
Where: WebEX
Importance: High

From: Szarfman, Ana <[redacted]>
Sent: Sunday, February 14, 2021 9:31 AM
To: Marks, Peter <[redacted]>
Cc: Stockbridge, Norman L <[redacted]>
Subject: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running

Hi Peter.

Thanks so much for your work.

Regarding the NYT article, I am quite concerned that the distributed network of EHRs that the Sentinel System uses does not have mortality data, and that they (and we) are not clamoring to fix this problem. It still takes several years to collect mortality data. Indeed there are no incentives to update electronic health records with mortality data.

https://www.nytimes.com/2021/02/12/health/covid-vaccine-how-safe.html?referringSource=articleShare

→ Let me know if you want Bill DuMouchel and I to discuss our proposal for more effective monitoring, including the need to use an updated algorithm by DuMouchel for data mining spontaneous reports at CDER and CBER.

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]

[redacted] US Toll
US Toll Free
Global call-in numbers | Toll-free calling restrictions

If you are a host, click here to view host information.

Need help? Go to https://help.webex.com

From: "Niu, Manette" <[redacted]>
To: "Anderson, Steven" <[redacted]>
Subject: RE: OBE Activities
Date: Wed, 18 Nov 2020 19:01:31 +0000
Importance: Normal

I am also working with Diane Gubernot on the Background rate project manuscript (in collaboration with CDC) – this has a quick timeline with the due date for a draft manuscript being mid-Dec. 2020.

I have a meeting with Ana Szarfman tomorrow to discuss her proposal on instituting new methods to improve data mining.

Thank you!

Manette

From: Niu, Manette
Sent: Wednesday, November 18, 2020 1:58 PM
To: Anderson, Steven <[redacted]>
Subject: OBE Activities

Steve,

I wanted to check in with you now that I'm back from filling in for AEB.

Please let me know if there is anything else you would like me to attend to besides the BEST and surveillance meetings I am currently on? I thought it would be informative if I continue to attend/follow DE COVID-19 surveillance meetings/activities, including those in which DE collaborates with CDC.

Thank you!

Manette

From: "Forshee, Richard" <[redacted]>
To: "Marks, Peter" <[redacted]>
Cc: "Anderson, Steven" <[redacted]>, "Witten, Celia (CBER)" <[redacted]>
Subject: Contact from Ana Szarfman
Date: Tue, 13 Jul 2021 17:55:16 +0000
Importance: Normal
Embedded: Issue__1_--_Death_signal_-->_WVAERS_2021W21_data_loaded_on_slc06lhx
Inline-Images: image001.png; image002.jpg; image003.jpg; image004.jpg; image005.jpg; image006.jpg

---

Dear Peter,

Ana Szarfman called me at about 4:15pm on Friday 7/9. She said that she and Bill DuMouchel had found an increased risk of mortality following COVID-19 vaccination using data mining methods. I asked her to send me the analysis and promised to review it, and I've attached the email she sent on Monday 7/12. It has very little information on the methods. I've pasted my reply to her below.

I am very concerned that whatever association they think they have identified is spurious based on the way the COVID-19 vaccination program prioritized individuals and the required and stimulated reporting we are seeing with the COVID-19 vaccines. Ana said that she had taken her name off a publication that is being prepared.

Please let me know how you would like us to proceed.

Best Regards,
--Rich

----------Response to Ana----------
Hi Ana,

Thanks for sharing this, and my team will review it. Do you have any more details on the new methods that Bill DuMouchel is using? That would be helpful in our evaluation.

In the email thread, Bill asked, "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."

During the time frame of this analysis, people with high-risk medical risk conditions were prioritized for vaccination. This included people in Nursing Homes early in the rollout. In later phases, most people needed to demonstrate that they had a high-risk medical condition to qualify for a vaccination. These included conditions like diabetes, chronic lung diseases, hypertension, heart conditions, obesity, and liver disease. Unless Bill has controlled for this somehow, I worry that the disproportionality in these results is caused by selection bias.

What are your thoughts on this possibility?

Best regards,
--Rich

----------End of Response----------

Richard Forshee, Ph.D.
Acting Deputy Office Director, CBER/OBE

Center for Biologics Evaluation and Research
Office of Biostatistics and Epidemiology
Analytics and Benefit-Risk Assessment Team
U.S. Food and Drug Administration
[redacted]

FDA U.S. FOOD & DRUG ADMINISTRATION

From: "Szarfman, Ana" <[redacted]>
To: "Forshee, Richard" <[redacted]>
Cc: "Stockbridge, Norman L" <[redacted]>, "Weichold, Frank" <[redacted]>
Subject: Issue #1 -- Death signal --> WVAERS 2021W21 data loaded on slc06lhx
Date: Mon, 12 Jul 2021 20:36:06 +0000
Importance: Normal
Attachments: VaccineHLT.xlsx
Inline-Images: image003.png

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]

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: "Forshee, Richard" <[redacted]>
To: "Marks, Peter" <[redacted]>
Cc: "Anderson, Steven" <[redacted]>
Subject: FYI: Ana Szarfman publication on COVID-19 vaccine safety
Date: Wed, 31 Aug 2022 12:28:10 +0000
Importance: Normal
Embedded: FW:_Summary_of_DE_issue_and_concerns
Inline-Images: image001.png; image002.jpg; image003.jpg; image004.jpg; image005.jpg; image006.jpg

---

Dear Peter,

I believe you will recall our conversations about the work that Ana Szarfman (CDER) was doing on COVID-19 vaccine safety. I have attached a copy of an email from about a year ago where you shared our concerns with CDER.

We have just learned that Dr. Szarfman is a co-author on a recently published paper based on COVID-19 vaccines and VAERS. We believe there are a number of issues with the paper and its findings, and we are discussing how best to respond. I don't recall receiving any prior notification from CDER about this publication.

Here is a link to the paper:
https://link.springer.com/article/10.1007/s40264-022-01186-z

We can discuss this at our next scheduled meeting, or we can schedule an *ad hoc* meeting this week. Please let us know if you have any other instructions.

Best Regards,
--Rich

Richard Forshee, Ph.D. (he/him/his)
Deputy Director, CBER/OBPV

Center for Biologics Evaluation and Research
Office of Biostatistics and Pharmacovigilance
Analytics and Benefit-Risk Assessment Team
U.S. Food and Drug Administration
[redacted]

FDA U.S. FOOD & DRUG ADMINISTRATION

From: "Marks, Peter" <[redacted]>
To: "Anderson, Steven" <[redacted]>, "Forshee, Richard" <[redacted]>
Subject: FW: Summary of DE issue and concerns
Date: Thu, 16 Sep 2021 08:53:34 +0000
Importance: Normal

Dear Steve and Rich,

FYI in follow up.

Best Regards,
Peter

From: Stein, Peter <[redacted]>
Sent: Wednesday, September 15, 2021 8:49 PM
To: Marks, Peter <[redacted]>
Cc: Cavazzoni, Patrizia <[redacted]>
Subject: FW: Summary of DE issue and concerns

Peter, thanks for flagging this – we've made it clear to her that she should not be discussing or providing internal analyses externally, and needs to focus on her assigned work. My apologies that this was thrown into the mix at a challenging time – from Steve's note, it's evident that your team is aware of the issues, and planning on looking into the approach your taking at an appropriate time.

Hopefully, you won't have further surprises….
Peter

From: Marks, Peter <[redacted]>
Sent: Wednesday, September 15, 2021 4:55 AM
To: Cavazzoni, Patrizia <[redacted]>
Cc: Walinsky, Sarah <[redacted]>
Subject: FW: Summary of DE issue and concerns

Dear Patrizia,

I really am sorry to bother you with this, but issue is become a major distraction. One of the CDER statisticians, Ana Szarfman, has decided on her own to do vaccine analyses using VAERS as part of her work at FDA. She is, however, not doing this in collaboration with our CBER statisticians, and quite to the contrary, has been asked to cease and desist, because the strategy that she is using could create erroneous conflicts that feed in to anti-vaccination rhetoric. This is creating an issue, as documented below by our office director, Steve Anderson.

This issue came up previously during the pandemic and working with Gerald it seemed to go away, but it is now back. Can we catch up about this sometime?

Thanks,
Peter

From: Anderson, Steven <[redacted]>
Sent: Tuesday, September 14, 2021 10:59 PM
To: Marks, Peter <[redacted]>
Cc: Forshee, Richard <[redacted]>; Nair, Narayan <[redacted]>
Subject: Summary of DE issue and concerns

Dear Peter,

Below is a summary of an issue and concerns expressed by the OBE Division of Epidemiology and IOD:

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.

As background, when the CBER Division of Epidemiology leadership began planning its approach to passive surveillance for the COVID-19 vaccines in the summer of 2020, the overarching strategy was to build on existing, established systems whenever possible. With regard to Data Mining, DE feels it is important to utilize the standardized, established system that has been in use for other vaccines for the past several years. The Division is concerned that use of a novel approach that had not been validated (and to our knowledge, has not been adopted by other medical product centers) would add another layer of uncertainty in the context of an EUA during the COVID-19 pandemic when rapid retrieval and interpretation of data would be imperative. DE recognizes as with all passive surveillance our current data mining process has limitations. In particular, DE is 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.

DE is planning to re-evaluate the data mining approach (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 DE believes 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). This retrospective approach is more downstream than what Ana is proposing but would be preferable to shifting midstream.

Let us know if you have questions or wish to discuss. Sorry for the hassles with this.

Regards,

Steve

Steve Anderson, Ph.D., M.P.P.
Director
Office of Biostatistics and Epidemiology
Center for Biologics Evaluation and Research
[redacted]

From: "Dal Pan, Gerald" <[redacted]>
To: "Marks, Peter" <[redacted]>, "Szarfman, Ana" <[redacted]>, "Anderson, Steven" <[redacted]>, "Forshee, Richard" <[redacted]>, "Witten, Celia (CBER)" <[redacted]>, "Stockbridge, Norman L" <[redacted]>
Subject: Follow up from today's meeting on mortality data and data mining
Date: Tue, 2 Mar 2021 02:27:08 +0000
Importance: Normal
Attachments: Trinidad_et_al_-_National_Vital_Statistics_Reports_-_2016.pdf; Hedegaard_et_al_-_National_Vital_Statistics_Report_-_2018.pdf; FAERS_Reports.xlsx
Inline-Images: image002.png

Peter and others,

Thank you for inviting me to participate in today's meeting. I have two follow-up items:

1. I mentioned that OSE collaborated with NCHS to examine literal texts from death certificates. I have attached two papers that describe how this was done (Trinidad et al) and some results (Hedegaard et al).
2. Ana mentioned that adverse event reports for AstrZeneca SARS-CoV-2 vaccine are in FAERS and not VAERS. Our team worked with Craig Zinderman in OBE to understand this. Since the AstraZeneca vaccine is neither authorized nor approved, the company as no postmarketing adverse event reporting requirements. The 200+ reports that we have in FAERS (Excel spreadsheet) list the AstraZeneca vaccine as a "suspect" drug but not as the "primary suspect" drug. Nearly all of these reports came from companies other than AstraZeneca, since those are companies were submitting adverse event reports for their dugs that are approved in the US. It is not unusual for reports for a drug to appear in FAERS before their US approval date for this reason.

This is all FYI only.

Please let me know if you have any questions.

Thanks.

Gerald

From: Marks, Peter <[redacted]>
Sent: Monday, March 1, 2021 3:44 PM
To: Szarfman, Ana <[redacted]>; Anderson, Steven <[redacted]>; Forshee, Richard <[redacted]>; Dal Pan, Gerald <[redacted]>; Witten, Celia (CBER) <[redacted]>; Stockbridge, Norman L <[redacted]>
Subject: RE: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running

Dear Ana,

Thanks so much for taking the time to go over everything so carefully with us. We will work through the issues that you presented.

Best Regards,
Peter

From: Szarfman, Ana <[redacted]>
Sent: Monday, March 1, 2021 2:08 PM
To: Marks, Peter <[redacted]>; Anderson, Steven <[redacted]>; Forshee, Richard <[redacted]>; Dal Pan, Gerald <[redacted]>; Witten, Celia (CBER) <[redacted]>; Stockbridge, Norman L <[redacted]>
Subject: RE: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running

Hi All,

I enjoyed the opportunity to have this discussion with you.

I have attached an updated presentation and the WSJ article on IBM retreating from Watson.

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]

U.S. FOOD & DRUG ADMINISTRATION

-----Original Appointment-----
From: Marks, Peter <[redacted]>
Sent: Wednesday, February 17, 2021 6:50 PM
To: Marks, Peter; Anderson, Steven; Forshee, Richard; Szarfman, Ana; Dal Pan, Gerald; Ball, Robert; Witten, Celia (CBER); Stockbridge, Norman L
Subject: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running
When: Monday, March 1, 2021 1:00 PM-2:00 PM (UTC-05:00) Eastern Time (US & Canada).
Where: WebEX
Importance: High

From: Szarfman, Ana <[redacted]>
Sent: Sunday, February 14, 2021 9:31 AM
To: Marks, Peter <[redacted]>
Cc: Stockbridge, Norman L <[redacted]>
Subject: NYT - As Millions Get Shots, F.D.A. Struggles to Get Safety Monitoring System Running

Hi Peter.

Thanks so much for your work.

Regarding the NYT article, I am quite concerned that the distributed network of EHRs that the Sentinel System uses does not have mortality data, and that they (and we) are not clamoring to fix this problem. It still takes several years to collect mortality data. Indeed there are no incentives to update electronic health records with mortality data.

https://www.nytimes.com/2021/02/12/health/covid-vaccine-how-safe.html?referringSource=articleShare

→ Let me know if you want Bill DuMouchel and I to discuss our proposal for more effective monitoring, including the need to use an updated algorithm by DuMouchel for data mining spontaneous reports at CDER and CBER.

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]

Meeting password: [redacted]

Tap to join from a mobile device (attendees only)
[redacted], [redacted] US Toll
US Toll Free

Join by phone
[redacted] US Toll
US Toll Free
Global call-in numbers | Toll-free calling restrictions

If you are a host, click here to view host information.

Need help? Go to https://help.webex.com

National Vital Statistics Reports

Volume 65, Number 9 December 20, 2016

Using Literal Text From the Death Certificate to Enhance Mortality Statistics: Characterizing Drug Involvement in Deaths

by James P. Trinidad, M.P.H., M.S., U.S. Food and Drug Administration; Margaret Warner, Ph.D., Brigham A. Bastian, B.S., Arialdi M. Miniño, M.P.H., and Holly Hedegaard, M.D., M.S.P.H., National Center for Health Statistics

Abstract

Objectives—This report describes the development and use of a method for analyzing the literal text from death certificates to enhance national mortality statistics on drug-involved deaths. Drug-involved deaths include drug overdose deaths as well as other deaths where, according to death certificate literal text, drugs were associated with or contributed to the death.

Methods—The method uses final National Vital Statistics System–Mortality files linked to electronic files containing literal text information from death certificates. Software programs were designed to search the literal text from three fields of the death certificate (the cause of death from Part I, significant conditions contributing to the death from Part II, and a description of how the injury occurred from Box 43) to identify drug mentions as well as contextual information. The list of drug search terms was developed from existing drug classification systems as well as from manual review of the literal text. Literal text surrounding the identified drug search terms was analyzed to ascertain the context. Drugs mentioned in the death certificate literal text were assumed to be involved in the death unless contextual information suggested otherwise (e.g., "METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS INFECTION"). The literal text analysis method was assessed by comparing the results from application of the method with results based on ICD–10 codes, and by conducting a manual review of a sample of records.

Keywords: text analysis • drug-involved death • drug overdose • National Vital Statistics System

Introduction

Recent mortality trends in the United States show a substantial increase in the rate of drug overdose deaths. From 2000 to 2014, the mortality rate for drug overdose more than doubled from 6.2 to 14.7 per 100,000 population (1). To address this public health concern, many researchers use National Vital Statistics System mortality data (NVSS–M) to describe these trends and to monitor the populations most at risk (1–4).

The NVSS–M data are based on information from the death certificates filed in the 50 states and the District of Columbia. The data set includes cause-of-death, demographic, and geographic information extracted from death certificates for all decedents in the United States (5). The NVSS–M data are coded using a standardized classification system, the *International Classification of Diseases and Related Health Problems, Tenth Revision* (ICD–10) (6). While this classification system allows for consistency in identifying the underlying and contributory causes of death, there are limitations in the use of ICD–10-coded data to study drug-involved mortality. Specifically, in the ICD–10 classification system, only a few drugs (e.g., heroin, methadone, and cocaine) are assigned a unique classification code (T40.1, T40.3, and T40.5, respectively) under certain circumstances (e.g., when the death is an overdose). Most drugs, however, are assigned to broad categories (e.g., both oxycodone and morphine are categorized to T40.2, Poisoning: Other opioids) (7). The use of broad categories in ICD–10 makes it difficult to use ICD–10 coded data to monitor trends in deaths involving specific drugs that are not already uniquely classified in ICD–10.

Analysis of literal text has been used to enhance mortality statistics in investigations of sudden infant death syndrome, Creutzfeldt-Jakob disease, influenza and pneumonia, cancer, and drug poisonings (8–13). The literal text often includes information beyond the general classification captured in an ICD–10 code description. For example, researchers have examined the literal

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics
National Vital Statistics System

text to better understand the circumstances (e.g., unsafe sleep environments) contributing to sudden infant death syndrome (9,12). Literal text can also be analyzed to identify a specific subset of deaths coded to a broad ICD–10 classification. For example, reseachers have examined the literal text to identify deaths from Creutzfeldt-Jakob disease among decedents with an ICD–10 underlying cause of death of B94.8, Sequelae of other specified infectious diseases (13). Similarly, researchers have explored literal text analysis methods to better understand the contribution of specific drugs in drug-poisoning deaths, and found that the literal text data provided more information on specific drugs than the ICD–10-coded data (11). These previous literal text analyses involving information on specific drugs did not consider literal text information other than drug mentions, were limited by causes of death, and assessed only records from a single state. Further development and use of literal text analysis methodology provides an opportunity for an enhanced understanding of the national picture of drug involvement in deaths in the United States (11).

This report describes the collaborative efforts of the National Center for Health Statistics (NCHS) and the U.S. Food and Drug Administration (FDA) to develop and assess a method for using literal text from death certificates to identify specific drugs involved in deaths, that is, drug overdose deaths and deaths with other types of drug involvement. This report accompanies a study that highlights the specific drugs most frequently involved in drug overdose deaths from 2010 through 2014 (14).

Methods Development

Overview

The analysis method uses search terms to identify drugs mentioned in electronic death certificate literal text (i.e., the cause-of-death statements on the death certificate). Unless contextual information suggested otherwise, drugs mentioned in the death certificate literal text were assumed to be involved in the death. Therefore, the method also analyzes literal text surrounding the identified search terms to determine whether the drugs mentioned were not involved in death (e.g., "METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS INFECTION"). The processed data resulting from applying the method includes all identified drug mentions and contextual information on drug involvement.

The following sections describe the data source for the literal text analysis methodology; some issues considered during the methods development; an assessment of the quality of the literal text data; the approach used to optimize the efficiency of literal text analysis; the development of lists of terms and phrases that were used in the processing of literal text; the steps of the literal text analysis methodology; and the data produced by applying the literal text analysis methodology.

Data source

The literal text analysis methodology was developed using final NVSS–M data linked to literal text data. Both NVSS–M data and literal text data are derived from information on death certificates (5).

In NVSS–M, the coded causes of death are assigned based on information written in the cause-of-death section on the death certificate (Figure 1). The information written on the death certificate by the medical certifier on the cause, manner, circumstances, and other factors contributing to the death is referred to as the literal text fields. The literal text fields of the cause-of-death section on the U.S. Standard Certificate of Death (15,16) include:

- The chain of events leading to death (from Part I)
- Other significant conditions that contributed to the death (from Part II)
- How the injury occurred (in the case of deaths due to injuries [from Box 43])

NCHS uses a software program to code the literal text from the death certificate according to the rules of ICD–10 (17). These processes involve the identification of statements from death certificate literal text, such as "MYOCARDIAL INFARCTION" and "DIABETES MELLITUS." Some statements, such as "METHADONE INTOXICATION," refer to drug-involved mortality. The identified statements are translated into ICD–10 codes. For example, the identified statement "OXYCODONE POISONING" is coded to ICD–10 codes T40.2, Poisoning: other opioids, and X42, Accidental poisoning by and exposure to narcotics and psychodysleptics (hallucinogens), not elsewhere classified. Note that throughout this report, text from death certificates is indicated in quotes and uppercase letters.

ICD–10 codes reflect the conditions reported on the death certificate. During the coding process, the software program assigns ICD–10 codes to 1 underlying cause and up to 20 multiple causes of death. Records rejected by the software program are reviewed by trained nosologists, and ICD–10 codes are manually assigned. In general, nosologists manually code about one-fifth of the death records. For deaths with an underlying cause of drug overdose (deaths with an underlying cause code of X40–X44, X60–X64, X85, or Y10–Y14), about two-thirds are coded manually (18). Entity axis ICD–10 codes include the ICD–10 code and information on the placement of the coded condition on the death certificate.

NCHS maintains the coded NVSS–M final mortality file and the literal text data separately, and linkage between the NVSS–M and literal text data leverages the information from both data sets. To link the data, NVSS–M and literal text files were merged on year of death, state of occurrence, and death certificate number.

U.S. STANDARD CERTIFICATE OF DEATH

LOCAL FILE NO. STATE FILE NO.

NAME OF DECEDENT
For use by physician or institution

To Be Completed/ Verified By: FUNERAL DIRECTOR:

1. DECEDENT'S LEGAL NAME (Include AKA's if any) (First, Middle, Last) | 2. SEX | 3. SOCIAL SECURITY NUMBER

4a. AGE Last Birthday (Years) | 4b. UNDER 1 YEAR Months Days | 4c. UNDER 1 DAY Hours Minutes | 5. DATE OF BIRTH (Mo/Day/Yr) | 6. BIRTHPLACE (City and State or Foreign Country)

7a. RESIDENCE STATE | 7b. COUNTY | 7c. CITY OR TOWN

7d. STREET AND NUMBER | 7e. APT. NO. | 7f. ZIP CODE | 7g. INSIDE CITY LIMITS? □ Yes □ No

8. EVER IN US ARMED FORCES? □ Yes □ No | 9. MARITAL STATUS AT TIME OF DEATH □ Married □ Married, but separated □ Widowed □ Divorced □ Never Married □ Unknown | 10. SURVIVING SPOUSE'S NAME (If wife, give name prior to first marriage)

11. FATHER'S NAME (First, Middle, Last) | 12. MOTHER'S NAME PRIOR TO FIRST MARRIAGE (First, Middle, Last)

13a. INFORMANT'S NAME | 13b. RELATIONSHIP TO DECEDENT | 13c. MAILING ADDRESS (Street and Number, City, State, Zip Code)

14. PLACE OF DEATH (Check only one: see instructions)

IF DEATH OCCURRED IN A HOSPITAL: □ Inpatient □ Emergency Room/Outpatient □ Dead on Arrival

IF DEATH OCCURRED SOMEWHERE OTHER THAN A HOSPITAL: □ Hospice facility □ Nursing home/Long term care facility □ Decedent's home □ Other (Specify):

15. FACILITY NAME (If not institution, give street & number) | 16. CITY OR TOWN , STATE, AND ZIP CODE | 17. COUNTY OF DEATH

18. METHOD OF DISPOSITION: □ Burial □ Cremation □ Donation □ Entombment □ Removal from State □ Other (Specify): | 19. PLACE OF DISPOSITION (Name of cemetery, crematory, other place)

20. LOCATION CITY, TOWN, AND STATE | 21. NAME AND COMPLETE ADDRESS OF FUNERAL FACILITY

22. SIGNATURE OF FUNERAL SERVICE LICENSEE OR OTHER AGENT | 23. LICENSE NUMBER (Of Licensee)

ITEMS 24-28 MUST BE COMPLETED BY PERSON WHO PRONOUNCES OR CERTIFIES DEATH

24. DATE PRONOUNCED DEAD (Mo/Day/Yr) | 25. TIME PRONOUNCED DEAD

26. SIGNATURE OF PERSON PRONOUNCING DEATH (Only when applicable) | 27. LICENSE NUMBER | 28. DATE SIGNED (Mo/Day/Yr)

29. ACTUAL OR PRESUMED DATE OF DEATH (Mo/Day/Yr) (Spell Month) | 30. ACTUAL OR PRESUMED TIME OF DEATH | 31. WAS MEDICAL EXAMINER OR CORONER CONTACTED? □ Yes □ No

CAUSE OF DEATH (See instructions and examples)

32. **PART I.** Enter the chain of events diseases, injuries, or complications that directly caused the death. DO NOT enter terminal events such as cardiac arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add additional lines if necessary.

Approximate interval: Onset to death

IMMEDIATE CAUSE (Final disease or condition ---------> resulting in death) a. _____ Due to (or as a consequence of):

Sequentially list conditions, if any, leading to the cause listed on line a. Enter the **UNDERLYING CAUSE** (disease or injury that initiated the events resulting in death) **LAST**

b. _____ Due to (or as a consequence of):

c. _____ Due to (or as a consequence of):

d. _____

PART II. Enter other significant conditions contributing to death but not resulting in the underlying cause given in PART I

33. WAS AN AUTOPSY PERFORMED? □ Yes □ No

34. WERE AUTOPSY FINDINGS AVAILABLE TO COMPLETE THE CAUSE OF DEATH? □ Yes □ No

To Be Completed By: MEDICAL CERTIFIER

35. DID TOBACCO USE CONTRIBUTE TO DEATH? □ Yes □ Probably □ No □ Unknown

36. IF FEMALE: □ Not pregnant within past year □ Pregnant at time of death □ Not pregnant, but pregnant within 42 days of death □ Not pregnant, but pregnant 43 days to 1 year before death □ Unknown if pregnant within the past year

37. MANNER OF DEATH □ Natural □ Homicide □ Accident □ Pending Investigation □ Suicide □ Could not be determined

38. DATE OF INJURY (Mo/Day/Yr) (Spell Month) | 39. TIME OF INJURY | 40. PLACE OF INJURY (e.g., Decedent's home; construction site; restaurant; wooded area) | 41. INJURY AT WORK? □ Yes □ No

42. LOCATION OF INJURY: State: City or Town: Street & Number: Apartment No.: Zip Code:

43. DESCRIBE HOW INJURY OCCURRED:

44. IF TRANSPORTATION INJURY, SPECIFY: □ Driver/Operator □ Passenger □ Pedestrian □ Other (Specify)

45. CERTIFIER (Check only one):
□ Certifying physician To the best of my knowledge, death occurred due to the cause(s) and manner stated.
□ Pronouncing & Certifying physician To the best of my knowledge, death occurred at the time, date, and place, and due to the cause(s) and manner stated.
□ Medical Examiner/Coroner On the basis of examination, and/or investigation, in my opinion, death occurred at the time, date, a nd place, and due to the cause(s) and manner stated.

Signature of certifier: _____

46. NAME, ADDRESS, AND ZIP CODE OF PERSON COMPLETING CAUSE OF DEATH (Item 32)

47. TITLE OF CERTIFIER | 48. LICENSE NUMBER | 49. DATE CERTIFIED (Mo/Day/Yr) | 50. **FOR REGISTRAR ONLY** DATE FILED (Mo/Day/Yr)

To Be Completed By: FUNERAL DIRECTOR

51. DECEDENT'S EDUCATION Check the box that best describes the highest degree or level of school completed at the time of death.
□ 8th grade or less
□ 9th 12th grade; no diploma
□ High school graduate or GED completed
□ Some college credit, but no degree
□ Associate degree (e.g., AA, AS)
□ Bachelor's degree (e.g., BA, AB, BS)
□ Master's degree (e.g., MA, MS, MEng, MEd, MSW, MBA)
□ Doctorate (e.g., PhD, EdD) or Professional degree (e.g., MD, DDS, DVM, LLB, JD)

52. DECEDENT OF HISPANIC ORIGIN? Check the box that best describes whether the decedent is Spanish/Hispanic/Latino. Check the "No" box if decedent is not Spanish/Hispanic/Latino.
□ No, not Spanish/Hispanic/Latino
□ Yes, Mexican, Mexican American, Chicano
□ Yes, Puerto Rican
□ Yes, Cuban
□ Yes, other Spanish/Hispanic/Latino (Specify) _____

53. DECEDENT'S RACE (Check one or more races to indicate what the decedent considered himself or herself to be)
□ White
□ Black or African American
□ American Indian or Alaska Native (Name of the enrolled or principal tribe) _____
□ Asian Indian
□ Chinese
□ Filipino
□ Japanese
□ Korean
□ Vietnamese
□ Other Asian (Specify) _____
□ Native Hawaiian
□ Guamanian or Chamorro
□ Samoan
□ Other Pacific Islander (Specify) _____
□ Other (Specify) _____

54. DECEDENT'S USUAL OCCUPATION (Indicate type of work done during most of working life. DO NOT USE RETIRED).

55. KIND OF BUSINESS/INDUSTRY

REV. 11/2003

SOURCE: NCHS, National Vital Statistics System.

Figure 1. U.S. standard death certificate

Considerations in developing methods to process death certificate literal text

In developing the analysis methodology, several characteristics and limitations of the literal text needed to be considered.

Availability of literal text information–Deaths may have no literal text data or only literal text mentions regarding the status of the death investigation (e.g., mentions of "PENDING" or "UNDER INVESTIGATION"). For these deaths, there are no mentions of drugs in the literal text.

Syntax of literal text–The syntax of the death certificate literal text generally consists of a few words or simple phrases (e.g., "DRUG TOXICITY") rather than clauses or sentences (e.g., "DECEDENT DIED OF DRUG POISONING"). The literal text analysis methods were developed by imitating the software program and processes that extract and assign ICD–10 codes to the literal text information as described above. These processes identify statements in the text.

Four text fields in Part I–The text fields constituting Part I of the death certificate have an assumed interpretation: The cause of death listed in the first text field is due to (or a consequence of) the cause of death (if any) listed in the second text field, which is due to (or a consequence of) the cause of death (if any) listed in the third text field, which is due to (or a consequence of) the cause of death (if any) listed in the fourth text field. The first cause of death listed in this sequence is the immediate cause of death, and the cause of death on the lowest-used line in Part I is the underlying cause of death. The assumed interpretation works well for some deaths. However, the assumption does not work well for other deaths. For example, medical certifiers may list multiple causes of death on a single line, may list a single cause of death on multiple lines, or may not write the causes in the appropriate sequential order. To simplify analyses, the assumed interpretation in Part I was ignored, and the text fields constituting Part I were concatenated as a single text field.

Case, symbols, and numbers–Use of uppercase and lowercase characters, symbols, and numbers varies across deaths. Some death certificate literal text may be in uppercase only, others in lowercase only, and others in a mixture of uppercase and lowercase. Literal text may contain symbols, such as hyphens. While the names of some drugs have hyphens (e.g., "GAMMA-HYDROXYBUTYRIC ACID"), use of hyphens in drug names can vary across death certificates, which complicates the identification of mentions of these drugs in literal text. Drug names (particularly generic drug names) generally do not include numbers, although numbers may be informative in clarifying the extent of drug exposure (such as in the phrase "BLOOD LEVEL ≥ 20 MG/DL"). To simplify analyses, all text was converted to uppercase, and symbols and numbers were removed.

Specificity of drug information–The specificity of drug information varies across death certificates. Death certificates may have mentions of specific drugs in the literal text (e.g., "OXYCODONE" or "FENTANYL"), mentions of drug classes (e.g., "OPIOID"), or exposures not otherwise specified (NOS) (e.g., "DRUG," "CHEMICAL," or "POLYPHARMACY"). Death certificates may have a mixture of mentions of specific drugs, drug classes, and exposures NOS. When a specific drug is mentioned alongside mentions of drug classes or exposures NOS, the mentions are sometimes referential (e.g., heroin is assumed to be the opioid in the phrase "OPIOID (HEROIN) OVERDOSE").

Synonyms–A specific drug may be referenced by various terms that are synonymous. For example, acetaminophen (generic name), paracetamol (generic name), and APAP (abbreviation) all refer to the same drug. When referring to a single-ingredient product, Tylenol (brand name) is also synonymous with acetaminophen. Brand names can refer to products with one or more drug ingredients. Literal text can have plural forms of drug mentions (e.g., mentions of "DRUGS," the plural form of drug). Literal text can also include misspellings. While drug metabolites are not synonymous with the parent drug products, drug metabolites may appear in the literal text, and are assumed to be the same. For example, a literal text mention of a toxicological finding of desmethyldiazepam (metabolite) would indicate exposure to diazepam (the parent drug).

Contextual information–Mentions of drugs are often accompanied by contextual information, which are other words in the literal text that either describe the drug(s) or provide information on how it was involved in mortality, if at all. The words in proximity to the drug mentions provide more informative contextual information than words that are distant.

Contextual information can provide details on drug characteristics or characteristics of drug exposure, such as the number of drugs (e.g., "MULTIPLE DRUGS"), extent of drug exposure (e.g., "FATAL LEVEL OF DRUG" or "THERAPEUTIC AMOUNT OF DRUG"), drug formulation (e.g., "DRUG TABLET"), the type of drug (e.g., "ILLICIT DRUG" or "DRUGS WHICH WERE PRESCRIBED"), and possession or ownership of the drug (e.g., "HIS DRUG" or "HER DRUG"). These descriptions can be complex and use conjunctions, such as the word "AND" (e.g., "FATAL LEVEL OF PRESCRIPTION DRUGS ILLEGALLY OBTAINED" and "ILLEGAL AND PRESCRIPTION DRUGS").

Contextual information can also explicitly describe how the drug exposure was involved in the death (e.g., "HEROIN POISONING" and "ANAPHYLAXIS DUE TO ANTIBIOTIC") or other aspects of drug involvement. Other aspects of drug involvement include route of administration (e.g., "DRUG INJECTION"), medical history with drug exposure (e.g., "HISTORY OF DRUG ABUSE" or "THERAPEUTIC USE OF METHADONE"), and other complications with drug exposure (e.g., "DRUG-DRUG INTERACTION"). Contextual information can also indicate drug exposure, either explicitly (e.g., "USE OF DRUG") or implicitly (e.g., "DRUG BLOOD LEVEL 20 MG/DL").

The contextual information can also be used to determine whether the drug mentioned in the literal text was not involved in mortality. For example, the drug "METHICILLIN" in the phrase "METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS INFECTION" does not suggest drug involvement in mortality, but rather a type of bacterial infection. Similarly, the phrase "NOT DRUG RELATED" clearly indicates that a death did not involve drugs. This report distinguishes between a drug mention, a drug mentioned with involvement (DMI), and a DMI death.

- A drug mention is any mention of a drug, a drug class, or exposure NOS in the literal text fields.
- A DMI is defined as a mention of a drug, a drug class, or exposure NOS in the literal text fields, excluding mentions where the contextual information suggested that the drug was not involved in the death.
- A DMI death is defined as a death having at least one DMI.

Information in the literal text can be contextual in that it provides information about drug characteristics or characteristics of drug exposure (i.e., descriptors), or contextual in that it describes whether and how a drug was involved in mortality (i.e., contextual phrases). Although descriptors provide some detail about the drugs mentioned in the literal text, they provide little or no information about drug involvement and, therefore, for the purposes of developing the literal text methodology, are less important than contextual phrases.

Multiple drugs–Deaths may involve multiple drugs. Medical certifiers may list these drugs consecutively, but not necessarily in order of importance to the cause of death (e.g., alphabetical order). These sequential drug mentions may be written with conjunctions, such as the word "AND" in the phrase "METHICILLIN AND VANCOMYCIN." Other sequential drug mentions do not contain conjunctions, such as in the phrase "OVERDOSE (HEROIN, COCAINE)."

While keyword searches can be performed to identify drug mentions, keyword searches are not efficient in identifying the contextual information associated with each drug mention. This is because the same contextual information may relate to more than one drug. For example, an infection that is resistant to both methicillin and vancomycin is inferred in the phrase "METHICILLIN AND VANCOMYCIN RESISTANT INFECTION." In the example, a search for "METHICILLIN RESISTANT INFECTION" would not identify the mention of methicillin, and a search for "METHICILLIN" would fail to identify that methicillin was not involved in mortality.

Searching for statements that incorporate both drug mentions and contextual information (e.g., searching for the statement "METHICILLIN AND VANCOMYCIN RESISTANT INFECTION") is the most direct approach for simultaneously identifying drug mentions and associated contextual information. However, this approach would require a vast number of statements due to the large number of drugs that can be mentioned, variability in the order of the drug mentions, and variability in the contextual information. In summary, there is an inexhaustible variety of combinations of statements consisting of drug mentions and contextual information.

Assessment of the presence of uninformative literal text

The quality of the literal text and its potential utility in identifying drug mentions was assessed by determining the percentage of records with no information that could be used to assign the cause(s) of death. The literal text was considered uninformative if: 1) there was no text in any of the literal text fields (i.e., the fields were blank) or 2) the fields only contained descriptive words or phrases about the status of the investigation (e.g., mentions of "PENDING" or "UNDER INVESTIGATION"). In most cases, when all the literal text is uninformative, an underlying cause-of-death of ICD–10 code R99 (Other ill-defined and unspecified causes of mortality) is assigned. Figure 2 contains all the terms considered to be uninformative for the purposes of identifying drug mentions.

Among NVSS–M records merged with literal text data for year 2013 (the most recent year of data at the time of the assessment), a small minority (less than 1%) had blank or uninformative literal text (Table A). Most of these were assigned an underlying cause-of-death code R99 (Other ill-defined and unspecified causes of mortality). Therefore, a small minority (less than 1%) of all records have literal text fields and ICD–10 coding that provide no information on specific causes of death.

Exchangeability: Optimizing efficiency of processing literal text information

Manual review of the literal text revealed that drug mentions are exchangeable (i.e., conceptually similar) when contextual information is fixed. For example, the word "HEROIN" in the phrase "HEROIN OVERDOSE" could be replaced (i.e., exchanged) with the word "OPIOID," with no change in the broad interpretation of the literal text (i.e., the cause of death was a drug overdose). Combinations of drug mentions are also exchangeable. For example, "METHICILLIN AND VANCOMYCIN" is exchangeable with the word "ANTIBIOTIC" in the phrase "ANTIBIOTIC RESISTANT INFECTION." Descriptors are also exchangeable. For example, the word "RX" can replace the descriptor "MULTIPLE PRESCRIPTION" in the phrase "MULTIPLE PRESCRIPTION DRUGS."

CAUSE UNDER INVESTIGATION, DEFERRED, PENDING, PENDING ADDITIONAL STUDIES, PENDING ADDITIONAL STUDY, PENDING AUTOPSY, PENDING AUTOPSY AND HISTOLOGY, PENDING AUTOPSY AND TOXICOLOGY, PENDING AUTOPSY HISTOLOGY, PENDING AUTOPSY TOXICOLOGY, PENDING AUTOPSY FINDING, PENDING AUTOPSY FINDINGS, PENDING AUTOPSY STUDIES, PENDING AUTOPSY STUDY, PENDING FURTHER INVESTIGATION, PENDING FURTHER STUDIES, PENDING FURTHER STUDY, PENDING HISTOLOGY, PENDING HISTOLOGY AND AUTOPSY, PENDING HISTOLOGY AND TOXICOLOGY, PENDING HISTOLOGY AUTOPSY, PENDING HISTOLOGY STUDIES, PENDING HISTOLOGY STUDY, PENDING HISTOLOGY TOXICOLOGY, PENDING LABORATORY STUDIES, PENDING LABORATORY STUDY, PENDING INVESTIGATION, PENDING TOXICOLOGY, PENDING TOXICOLOGY AND AUTOPSY, PENDING TOXICOLOGY AND HISTOLOGY, PENDING TOXICOLOGY AUTOPSY, PENDING TOXICOLOGY HISTOLOGY, PENDING TOXICOLOGY STUDIES, PENDING TOXICOLOGY STUDY, PENDING STUDIES, PENDING STUDY, PENDING TOX UNDER INVESTIGATION.

SOURCE: NCHS, National Vital Statistics System, death certificate literal text.

Figure 2. Literal text strings considered uninformative for assigning cause of death and identifying drug mentions

Table A. Deaths having no informative literal text on cause of death: U.S. residents, 2013

Characteristics	Number of deaths	Percent of deaths
All deaths	2,596,993	100.00
Deaths having no informative literal text	3,831	0.15
Deaths having no informative literal text and ICD–10 code R99 as underlying cause of death[1]	3,421	0.13

[1]The ICD–10 code R99 indicates Other ill-defined and unspecified causes of mortality.

NOTE: ICD–10 is the *International Classification of Diseases and Related Health Problems, Tenth Revision.*

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

The exchangeability of drug mentions enables the DMI programs to more efficiently process data on drug mentions and their associated contextual information. For example, replacing sequential drug mentions identified in the literal text (e.g., "METHICILLIN AND VANCOMYCIN" or "VANCOMYCIN AND METHICILLIN") with the word "DRUG" greatly simplifies the processing steps for the DMI program.

A stepwise approach was used to enhance the DMI program efficiency in extracting information from the literal text. This stepwise approach leverages the exchangeability of drug mentions and the exchangeability of descriptors. In other words, contextual information on drug involvement can be most efficiently identified and processed using the computer algorithms when the variability in drug mentions and associated descriptors is reduced. This stepwise approach required the development of lists of drugs, descriptors, and joining phrases that link search terms or descriptors together, and the development of contextual phrases.

Developing a search term list for drugs

A list of search terms was developed to identify drug mentions. This list was developed using a two-phase approach. The final list of search terms included single words (e.g., "HEROIN") and combinations of words (e.g., "CRACK COCAINE") for specific drugs, drug classes, and drug exposures NOS.

In the first phase, the search term list was constructed from single-word generic names listed in the Substance Abuse and Mental Health Services Administration's (SAMHSA) Drug Abuse Warning Network (DAWN) Drug Reference Vocabulary (DRV), published in 2012 (19). DAWN DRV is a drug vocabulary and classification system based on the Multum Lexicon database from Cerner Multum, Inc. Its structure is hierarchical with generic drugs categorized under higher-level groupings (e.g., drug class). For use with the DAWN data system, SAMHSA added substances that are misused and abused (e.g., illicit drugs and inhalants) that were not included in the Multum Lexicon database.

During this first phase of generating the search term lists, the following DAWN DRV categories were excluded: major substances of abuse, nutritional products, alternative medicines, medical gases, biologicals, immune globulins, immunostimulants, sterile irrigating solutions, and drugs unknown. Products in these categories had generic names that were difficult to condense into a single word denoting a drug product. The list also excluded combination products, nearly all of which could be identified by their components. The search term list that resulted from the first phase of development did not include names of drug classes or drug exposures NOS.

In the second phase, the search term list was expanded by adding terms for specific drugs not identified in the first phase, including illicit drugs; drug classes; drug exposures NOS; terms containing more than one word; brand names; and obvious, frequently occurring misspellings. Most of the search terms added during the second phase were identified through nonsystematic manual reviews and queries of the 2003–2014 literal text.

Methods development was focused on literal text data from 2007, the first year of literal text data that was available during methods development, and from 2013, the most recent year of data available at the time when assessments were conducted. Additional search terms for brand names of prescription drugs were identified using the Drugs@FDA website, and search terms for misspelled drugs were identified using FDA Adverse Event Reporting System data (20). A few search terms were also identified using other approaches, including comparison with ICD–10 codes.

The search term list that resulted from the second phase excluded foods and food additives (e.g., starch), excipients, gases (e.g., helium and carbon monoxide), airborne contaminants (e.g., soot), industrial chemicals (e.g., ethylene glycol), periodic table elements (e.g., lithium and iodine), and substances with unknown industrial or pharmaceutical applications. Although therapeutic uses of some of these substances is possible, these substances were not included because it proved difficult to determine whether the exposures to the substances were therapeutic, for misuse or abuse, or environmental.

Study team members trained in pharmacy and pharmacoepidemiology categorized search terms by various characteristics, including whether the terms referred to specific drugs, drug classes, or exposures NOS. Search terms were also classified by whether they represented generic drug names or other variants, such as brand names, common use or street names, abbreviations, metabolites, and misspellings. Most search terms were mapped to a single "principal variant," the overarching label assigned to a drug, a drug class, or exposure NOS. In general, the principal variant was the generic drug name. Some search terms—mostly for combination drug products—were mapped to two or more principal variants. The use of principal variants made it possible to identify all deaths that involved the same drug.

The development of the search term list involved various efforts to create a comprehensive list of all drugs mentioned in literal text. Although many methods were used to develop the list, the list might not contain all possible search terms for all possible drugs. The assessments conducted during methods development were based on a June 2015 list of 2,865 search terms representing 1,649 principal variants (see Table I–1). This list was updated in November 2015 to include 3,116 search terms representing 1,643 principal variants (see Table I–2).

Developing lists of contextual information

Three lists of contextual information were developed using iterative manual reviews and queries of literal text for data years 2003 through 2014. The three lists consisted of descriptors, joining phrases, and contextual phrases.

The list of descriptors included a word or words that provide information on drug characteristics or characteristics of drug exposure, such as "MULTIPLE," "PRESCRIPTION," and "NON PRESCRIPTION." The list classified whether the descriptor should be identified before a drug mention, after a drug mention, or either before or after a drug mention (as would be the case for the descriptor "PRESCRIPTION" in the phrases "PRESCRIPTION DRUG" and "DRUG PRESCRIPTION"). The list also classified the descriptors by the characteristic(s) they aim to describe (e.g., "TABLET" and "TRANSDERMAL" describe type of drug formulation).

The list of joining phrases included words and asterisks that acted as conjunctions. For this list, each joining phrase was comprised of 1) two asterisks that indicate exchangeability of either drug mentions or descriptors and 2) potentially other words that indicate linkage. Examples of words that indicate linkage include "AND" and "AS WELL AS." Bookending these words were asterisks, as in the case of the joining phrases "* AND *" and "* AS WELL AS *." These asterisks were exchangeable with drug mentions or descriptors, as in the phrases "METHICILLIN AND VANCOMYCIN" and "ILLICIT AS WELL AS PRESCRIPTION." The simplest joining phrase was "* *," indicating two adjacent drug mentions or two adjacent descriptors.

The list of contextual phrases included words and asterisks that, altogether, describe drug involvement (if any). Examples of contextual phrases include "* TOXICITY" and "ABUSED *." Like the asterisks in joining phrases, asterisks in contextual phrases indicate exchangeability of mentions. However, while the asterisks in joining phrases refer to either drug mentions or descriptors, the asterisks in contextual phrases simultaneously refer to drug mentions, any associated descriptors, and joining phrases. In addition, while there are only two asterisks in joining phrases, contextual phrases may have one or more asterisks, as in the case of "ACCIDENTAL * TOXICITY WITH *," which could refer to the phrase "ACCIDENTAL DRUG TOXICITY WITH HEROIN AND OTHER ILLICIT DRUGS." The simplest contextual phrase was "*," indicating the mention of one or more drugs and associated descriptors, but no other contextual information.

Study team members classified the contextual phrases by various characteristics. The most important characteristic was whether the contextual phrase did not suggest drug involvement. Contextual phrases that suggested no drug involvement generally referred to health conditions or disease states. For example, when the word "INSULIN" replaces "*" in the contextual phrase "* DEPENDENT DIABETES," the resulting text refers to a health condition. Similarly, when the word "METHICILLIN" replaces "*" in the contextual phrase "* RESISTANT STAPHYLOCOCCUS AUREUS INFECTION," the resulting text refers to a type of bacterial infection. Other contextual phrases clearly indicated no drug involvement, which would be the case for the contextual phrase "NO * INVOLVED."

The drugs mentioned in the death certificate literal text were assumed to be involved in the death unless contextual information suggested otherwise.

Contextual phrases that described similar ideas (such as "* TOXICITY" and "TOXICITY FROM *") were classified under a common category. Some phrases were classified under more than one category; for example, "TOXICITY FROM * INJECTION" was classified under the category for toxicity and the category for injection.

The assessments conducted during methods development were based on 527 descriptors, 22 joining phrases, and 1,641 contextual phrases that were listed as of June 2015. These lists were updated in November 2015.

Identifying mentions of drugs and ascribing context

Using SAS Version 9.3 (21), a suite of software programs (referred to as the DMI programs) was developed to automate the identification of drug mentions in the literal text and to determine possible involvement of the drug in the death based on contextual information.

Figure 3 provides an example of the application of the DMI program logic to the following death certificate literal text: "INGESTED ILLICIT AND RX DRUGS (HEROIN AND METHADONE); HX OF OPIOID ABUSE." Leveraging the exchangeability of drug mentions and the exchangeability of descriptors, the DMI programs use five steps to identify drug mentions and ascribe context to each drug mention (Figure 3).

The first step prepares the literal text, resulting in text that does not have symbols, numbers, and double spaces, and is formatted in uppercase letters.

The second step uses the list of search terms to identify drug mentions in the literal text. During this step, a new record is generated for every search term (i.e., drug mention) identified in the literal text. The DMI programs also identify simple plural forms (i.e., search term plus the letter "S"). In the example in Figure 3, the DMI programs generate four records for the mentions of "DRUGS," "HEROIN," "METHADONE," and "OPIOID."

Using the list of descriptors, the DMI programs iteratively identify descriptors for each drug mention in the third step. In the first iteration, the DMI programs identify and map descriptors (such as "RX") to adjacent drug mentions (such as the mention of "DRUGS"), resulting in a drug mention with a simple description (e.g., "RX DRUGS"). Subsequent iterations use the list of joining phrases and list of descriptors to form more complex descriptions. In the example, the DMI programs link the descriptor "ILLICIT" and the descriptor "RX" with the

NOTE: In this example, the DMI (drug mentioned with involvement) programs identify three drug mentions ("DRUGS," "HEROIN," "METHADONE") in the literal text and map these drug mentions to one contextual phrase ("INGESTED *"). The DMI programs also identify one drug mention ("OPIOID") and map this drug mention to one contextual phrase ("HX OF * ABUSE").
SOURCE: NCHS, Division of Vital Statistics.

Figure 3. Example of the application of the DMI program logic to the literal text

joining phrase "* AND *" to form the more complex description "ILLICIT AND RX." The resultant drug mention and associated descriptors are subsequently more complex (e.g., "ILLICIT AND RX DRUGS").

The fourth step replaces drug mentions and associated descriptors with a single asterisk "*" and also replaces consecutive drug mentions and associated descriptors with a single asterisk "*." This step also uses joining phrases to determine whether drug mentions are consecutive. For example, using the joining phrase "* AND *," the mention of "HEROIN" and the mention of "METHADONE" are consecutive in the text "HEROIN AND METHADONE." Similarly, with the joining phrase "* *," the mention of "ILLICIT AND PRESCRIPTION DRUGS" and "HEROIN" are consecutive mentions. In the example, the mention of "OPIOID" is not listed consecutively with other drug mentions.

Using the list of contextual phrases, the fifth step identifies and maps contextual phrases to the appropriate drug mention(s), that is, the drug mentions that were replaced in step 4. In the example, the mentions of "DRUGS," "HEROIN," and "METHADONE" are mapped to the contextual phrase "INGESTED *," while the mention of "OPIOID" is mapped to the contextual phrase "HX OF * ABUSE."

Each search term is mapped to only one contextual phrase. To optimize the mapping procedures, contextual phrases with asterisks located between other words (e.g., "HX OF * ABUSE") are mapped before contextual phrases with asterisks located at the end of the contextual phrase (e.g., "INGESTED *").

Data produced by applying the literal text analysis methodology

Application of the literal text analysis methodology results in a data set of decedents, drug mentions, and contextual information associated with each drug mention (Figure 3). The drug mentions are categorized by principal variant and whether the drug mentions refer to specific drugs, drug classes, or exposures NOS. The contextual phrases are categorized by indication of involvement of drugs in death. When the processed literal text data are linked with NVSS–M data, the ICD–10 underlying and multiple cause-of-death codes, demographic information, geographic information, and other information in the multiple cause-of-death file are also available.

In this report, the literal text analysis methodology was applied to NVSS–M data linked with literal text for year 2013 as an example. For this analysis, mentions of alcohols, tobacco, and nicotine were excluded, as they are involved in many deaths that do not involve other drugs.

Table B shows the number of U.S. resident deaths with drug mentions and DMIs based on the 2013 literal text. Of the approximately 2.6 million deaths in 2013, 114,621 had at least one drug, alcohol, tobacco, or nicotine mention. The number of deaths with a drug mention was 72,518. The number of deaths with at least one drug mention and no contextual information indicating that the drug was not involved in the death (DMI) was 65,062. Among these deaths, there were 150,342 DMIs, for an average of 2.3 DMIs per death.

Table C shows the level of specificity of the drug mentions (i.e., whether the drug mention was a specific drug, a drug class, or an exposure NOS) for the 65,062 DMI deaths in 2013. The majority of DMIs referred to a specific drug (58%). Most of the specific drug mentions were generic names (82,895 DMIs,

Table B. Deaths with drug mentions and mentions of drug involvement: U.S. residents, 2013

Characteristics	Number of deaths	Number of mentions
Deaths among U.S. residents	2,596,993	…
Deaths with at least one drug, alcohol, tobacco, or nicotine mention	114,621	216,361
Deaths with at least one drug mention	72,518	158,104
Deaths with at least one DMI (drug mentioned with involvement in death)	65,062	150,342

… Category not applicable.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

Table C. Number and percentage of DMIs, by level of specificity of the drug mention: U.S. residents, 2013

Type of DMI	Number	Percent
All DMIs	150,342	100.0
Specific drug	87,764	58.4
Drug class	8,979	6.0
Exposure not otherwise specified[1]	53,599	35.7

[1]Category includes nonspecific references to drugs (e.g., mention of "POLYPHARMACY" or "DRUG").

NOTES: Mentions of alcohols, tobacco, and nicotine were excluded from the analyses. DMI is a drug mentioned with involvement in the death.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with literal text data.

Table D. Number and percentage of DMI deaths, by level of specificity of the DMI: U.S. residents, 2013

Type of DMI	Number	Percent
All DMI deaths	65,062	100.0
Deaths with mention of at least one specific drug	45,035	69.2
Deaths with mention of a drug class only	4,560	7.0
Deaths without mention of a drug class or specific drug[1]	15,467	23.8

[1]Category includes DMI deaths with mentions of nonspecific drug references (e.g., mention of "POLYPHARMACY" or "DRUG").
NOTES: Mentions of alcohols, tobacco, and nicotine were excluded from the analyses. DMI is a drug mentioned with involvement in the death.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with literal text data.

or 95%), while the remainder of the specific drug mentions were other variants, such as brand names and misspellings. Slightly more than one-third of all DMIs (36%) were nonspecific references to drugs.

Similarly, DMI deaths can be categorized by the highest level of specificity of the drugs involved. Table D shows the number of DMI deaths from 2013. Of the 65,062 DMI deaths, 69% had mentions of at least one specific drug, while 7% had mentions of a drug class but not a specific drug. For 24% of the deaths, only nonspecific drug references were found (i.e., neither a drug class nor specific drug were mentioned).

Assessments of the literal text analysis methodology

Two assessments examined the performance of the literal text analysis methodology in identifying DMIs and DMI deaths. The assessments were conducted using NVSS–M data linked with literal text for year 2013.

The first assessment examined the agreement between data produced by the DMI programs and ICD–10 coded data for three selected drugs. In the ICD–10 classification system, there are a few T codes, F codes, and R codes that identify deaths with poisonings, mental and behavioral disorders, and toxicological findings related to specific drugs, respectively. These specific drugs include cocaine, heroin, and methadone. The ICD–10 rules for assigning codes in mortality can be found elsewhere (17). Comparisons were made between the numbers of DMI deaths identified by the DMI programs and the numbers of deaths identified as having one of the specific T, F, or R codes for cocaine, heroin, or methadone (Table E). Considering the differences between the DMI definition (i.e., a drug was mentioned and there was no contextual information indicating that the drug was not involved in the death) and the ICD–10 definitions and rules for assigning T, F, and R codes, there was high agreement (greater than 90%) between the DMI programs and the ICD–10 codes in the identification of deaths involving cocaine, heroin, and methadone (Table E).

The second assessment examined the accuracy of the DMI programs in identifying DMIs and DMI deaths. This assessment was based on two subsets of mortality records that were likely to have DMIs: 1) deaths selected by the application of the DMI programs to mortality records and 2) deaths with no uninformative literal text fields and selected using ICD–10 entity axis codes that likely pertained to a drug-involved mortality. These codes included ICD–10 codes referring to mental or behavioral disorders due to psychoactive substance use, poisonings, adverse effects due to drugs and alcohol, and ICD–10 codes whose title or definition explicitly indicated drug involvement (e.g., P04.4 Fetus and newborn affected by maternal use of drugs of addiction) (Figure 4). ICD–10 codes that only indicated alcohol, tobacco, or nicotine involvement were excluded from the list of selected ICD–10 codes. In summary, the codes used in the analysis included those typically used to identify drug overdose deaths and those that indicated other drug involvement (e.g., anaphylaxis) (2,22).

From the pool of mortality records identified by either of the two selection methods, a simple random sample of 2,000 records was taken and manually reviewed to determine whether drug mentions in the literal text (if any) met the definition of a DMI

Table E. Agreement between DMI programs and selected ICD–10 codes: U.S. residents, 2013

Referent drug	ICD–10 code(s) that apply to referent drug[1]	Deaths with DMI of referent drug[2] [A]	Deaths with ICD–10 code(s) that apply to referent drug[3] [B]	Deaths with either DMI of referent drug or ICD–10 code(s) that apply to referent drug [C]	Deaths with both referent drug mention and ICD–10 code(s) that apply to referent drug [D]	*D/A* x 100	*D/B* x 100	*D/C* x 100
Cocaine	T40.5, F14.–, R78.2	7,324	7,176	7,361	7,139	97.4	99.5	97.0
Heroin	T40.1	8,924	8,360	8,968	8,316	93.2	99.5	92.7
Methadone	T40.3	4,005	3,737	4,029	3,713	92.7	99.4	9.2

[1]ICD–10 codes used in this analysis were entity axis codes.
[2]The DMI programs identify deaths with mention of the referent drug in the literal text fields, excluding mentions where the contextual information suggested that the drug was not involved in the death.
[3]The listed T codes, F codes, and R codes identify deaths due to poisonings, mental and behavioral disorders, and toxicological findings related to the referent drug, respectively.
NOTES: DMI is a drug mentioned with involvement in the death. ICD–10 is *International Classification of Diseases and Related Health Problems, Tenth Revision.*

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

A80.0, D52.1, D59.0, D59.2, D61.1, D64.2, D68.3, E03.2, E06.4, E16.0, E23.1, E24.2, E27.3, E66.1, F11–F16, F19, F55, G21.1, G24.0, G25.1, G25.4, G25.6, G44.4, G62.0, G72.0, H26.3, H40.6, I42.7, I95.2, J70.2, J70.3, J70.4, K85.3, L10.5, L23.3, L24.4, L25.1, L27[.0–.1], L27[.8–.9], L43.2, L56[.0–.1], L64.0, M10.2, M32.0, M34.2, M80.4, M81.4, M83.5, M87.1, N14[.0–.2], O35.5, P04[.0–.1], P04.4, P04[.8–.9], P58.4, P93, P96[.1–.2], Q86[.1–.2], R50.2, R78[.1–.6], R78[.8–.9], R82.5, R83[.2–.3], R84[.2–.3], R85[.2–.3], R86[.2–.3], R87[.2–.3], R89[.2–.3], T36, T37, T38, T39, T39[.1–.4, .8–.9], T42, T43–T50, T57[.8–.9], T65[.5, .8–.9], T88[.0–.1, .6–.7], T96, T97, X4T400–X44, X49, X60–X64, X69, X85, X89–X90, Y10–Y14, Y19, Y40–Y47, Y49–Y59, Y88.0, Z03.6, Z72.2, Z91.0, Z92[.1–.2]

SOURCE: *International Classification of Diseases and Related Health Problems, Tenth Revision* (ICD–10).

Figure 4. ICD–10 entity axis codes likely pertaining to a drug-involved mortality

and whether the sampled record reflected a true DMI death. The results from the manual review served as the "gold standard."

The performance of the DMI programs to identify DMIs and DMI deaths was quantified using the following measures: true positives, false positives, false negatives, true negatives (only calculated for deaths, not drug mentions), and positive predictive values (PPVs). Each drug mention was categorized as either a true-positive mention (identified by both the DMI programs and by manual review), a false positive mention (identified by the DMI program but not the manual review), or a false negative mention (identified by manual review but not the DMI program). Reasons that a mention was categorized as false positive or false negative were described.

Similarly, each death was categorized as either a true-positive death, false positive death, or false negative death. True-negative deaths were identified only by ICD–10 codes, but were not categorized as DMI deaths according to manual review. PPVs quantified the percentage of DMIs or DMI deaths correctly identified as such by the DMI programs (i.e., true positives/[true positives + false positives]). Measures of sensitivity and specificity could not be calculated because the selected records were not randomly sampled from all mortality records.

From the application of the DMI programs, 65,062 deaths were identified as possible DMI deaths in 2013. From selection based on ICD–10 codes (Figure 4), 61,282 deaths were identified as likely pertaining to drug involvement. Combined, the two methods identified 69,493 unique deaths with possible drug involvement. A majority of these deaths (56,851 or 81.8%) was identified by both methods, while a minority of these deaths (4,431 or 6.4%) was only identified using ICD–10 codes. The remaining deaths (8,211 or 11.8%) were identified by the DMI programs only.

The 2,000 randomly sampled deaths included 1,808 deaths identified using ICD–10 codes, of which 1,691 deaths were also identified by the DMI programs. The remaining 192 deaths in the sample were only identified by the DMI programs.

The DMI programs identified DMIs with high accuracy (Table F). According to manual review of literal text, 4,357 (97%) of the 4,487 mentions identified by the DIM programs were true-positive mentions, while the remaining 130 mentions (3%) were categorized as false positive. The DMI programs failed to identify 52 mentions of drugs involved in mortality. Some deaths may have a mixture of true-positive, false-positive, and false-negative mentions in their literal text.

The DMI programs also identified DMI deaths with high accuracy (Table G). According to manual review of literal text for deaths identified using ICD–10 codes, 1,804 of the 1,883 deaths (96%) identified by the DMI programs were true-positive deaths, while the remaining 79 deaths (4%) were categorized as false positive. The DMI programs did not identify 100 deaths that did not have drug involvement (true-negative deaths), but failed to identify 17 deaths that did have drug involvement (false-negative deaths). All 117 of these deaths were identified using ICD–10 codes.

The false-positive and false-negative mentions fell into nine categories (Table H). In a few instances, the DMI programs identified more text than should have been identified. For example, the DMI programs identified a mention of "PAIN NARCOTIC" instead of "NARCOTIC" in the literal text "BACK PAIN NARCOTIC DEPENDENT." In contrast, the DMI programs sometimes identified one or more search terms that were nested in a longer drug name, resulting in false-positive and false-negative mentions. The DMI programs also identified false-positive mentions for other reasons, including: search terms were not drugs, search terms were used to describe health conditions and disease states, or contextual information indicated no drug involvement. Manual review of literal text also identified other reasons for false-negative mentions: drugs mentioned in the literal text were not search terms, or a drug mention was not separated by a space from other words in literal text.

The findings from the assessment were used to update and improve the lists of search terms and contextual information.

Table F. DMI programs' ability to identify DMIs among a random sample of 2,000 deaths having one or more ICD–10 entity axis codes or identified using the DMI programs: U.S. residents, 2013

Evaluation	DMIs identified from the manual review		
	Yes	No	Total
DMIs identified by the DMI programs	4,357	130	4,487
DMIs not identified by the DMI programs	52	…	…

… Category not applicable.

NOTES: See Figure 4 for list of entity axis codes. Positive predictive value calculated as: 4,357 mentions/4,487 mentions = 97.1%. DMI is a drug mentioned with involvement in the death.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

Table G. DMI programs' ability to identify DMI deaths among a random sample of 2,000 deaths having one or more ICD–10 entity axis codes or identified using the DMI programs: U.S. residents, 2013

Evaluation	DMI deaths identified from the manual review		
	Yes	No	Total
DMI deaths identified by the DMI programs	1,804	79	1,883
DMI deaths not identified by the DMI programs	17	100	117

NOTES: See Figure 4 for list of entity axis codes. Positive predictive value calculated as: 1,804 deaths/1,883 deaths = 95.8%. DMI is a drug mentioned with involvement in the death.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

Table H. Reasons for false-positive and false-negative mentions in the assessment of the DMI programs to identify DMIs and DMI deaths

Reason for false-positive or false-negative mention	Example	Result of assessment
Search term was not a drug	DMI program identified "DIFLUOROETHANE," which is not a drug	Identified a false-positive mention
Search term used to describe health condition or disease state	DMI program identified "FOLIC ACID" in text "FOLIC ACID DEFICIENCY," or identified "PCP," referring to pneumocystis pneumonia	Identified a false-positive mention
Drug mention nested in an identified search term	DMI program identified "PAIN NARCOTIC" instead of "NARCOTIC" in text "BACK PAIN NARCOTIC DEPENDENT"	Identified a false-positive mention and a false-negative mention
Search term was nested in a longer drug name	DMI program identified "DRUG" in text "NON STEROIDAL ANTIINFLAMMATORY DRUG"	Identified a false-positive mention and a false-negative mention
Context adjacent to search term indicated no drug involvement	DMI program identified "DRUG" in text "NO DRUG INVOLVEMENT"	Identified a false-positive mention
Drug was not a search term	"CONTRAST DYE" was not identified because it was not a search term	Identified a false-negative mention
Drug name was not separated from other words in literal text	DMI program failed to identify "ALPRAZOLAM" in text "OPIOID ANDALPRAZOLAM OVERDOSE"	Identified a false-negative mention

NOTES: A false-positive mention indicates that the drug was identified by the DMI programs but not during the manual review. A false-negative mention indicates that the drug was identified during the manual review but not by the DMI programs. DMI is drug mentioned with involvement.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text.

Discussion

New method for identifying drug involvement in death

The application of the literal text analysis methodology described in this report can be used to enhance mortality statistics by facilitating the identification of specific drugs involved in drug overdose deaths and deaths with other drug involvement. ICD–10 (6), which has historically been used to classify the drugs involved in the deaths in NVSS–M, is limited in that the vast majority of drugs are classified into broad categories. For example, oxycodone, hydrocodone, and morphine are all classified to T40.2 (Poisoning: Other opioids) (7). There are a few notable exceptions, such as heroin (T40.1), methadone (T40.3), and cocaine (T40.5), which are separately coded in the case of a drug overdose death. In contrast, the methods described in this report allow for the identification of drugs that are not uniquely identified in ICD–10.

The identification of specific drugs provides flexibility in analyses. Specific drugs can be categorized according to classification schemes different than those of the ICD–10 categories. Identifying specific drugs also allows comparisons between drugs within a particular class. In addition, identifying specific drugs allows for more detailed analysis on deaths involving multiple drugs that are classified to the same or even different categories.

The literal text analysis methodology was developed to extract information on the specific drugs involved in deaths from the nonstructured literal text data obtained from death certificates. Utility of the methodology depends on the quality and quantity of information in literal text. The methodology will not identify a drug mention among deaths whose literal text only states "POISONING" or "OVERDOSE," but does not have any reference to drugs. Many issues were considered when designing this methodology, including the unstructured nature of the data, the number of drugs mentioned, the contextual information describing the drug involvement, and the efficiency of the programs to extract information on the drugs involved. Ultimately, the methods that were developed imitate, to some

degree, the current processes used to identify statements in death certificates for eventual translation into ICD–10 codes. With the search terms, descriptors, and contextual phrases identified, it is possible to approximately construct literal text statements related to drug-involved mortality.

Development and maintenance of the DMI lists and programs

The importance of creating comprehensive lists to be used by the DMI programs cannot be overstated. The DMI programs are a series of steps that identify drug mentions, descriptors of those drug mentions, and other contextual information. Each step of the processing of literal text requires lists: search terms, descriptors, joining phrases, or contextual phrases. Incomplete lists used by the DMI programs may result in failure in any of the processing steps, which would result in a failure to associate drug mentions with the appropriate contextual information.

The development of the lists used by the DMI programs requires an understanding of the drugs of interest as well as iterative manual reviews of literal text, and this development process is time intensive. The high percentage of agreement between the DMI programs and the manual review suggests that the lists used by the DMI programs were generally comprehensive. However, even with the careful development of these lists, the DMI programs found a few mentions that did not refer to drugs or failed to identify DMIs. For example, the DMI programs found false-positive mentions of "PCP" that referred to pneumocystis pneumonia, but the programs failed to find mentions of "CONTRAST DYE," which was a drug class that was not a search term. Incompleteness in the list of contextual phrases also yielded false-positive DMIs (e.g., identification of "FOLIC ACID" in the text "FOLIC ACID DEFICIENCY"). These false-positive and false-negative mentions demonstrate the importance of careful development of these lists. Updating and refining the lists used by the DMI programs will help resolve these issues for future investigations.

This report found that a little over one-third of deaths involving drugs did not include information on the death certificate about the specific drug(s) involved. This finding from the literal text analysis is consistent with other analyses of the ICD–10 coded data (23). Efforts are underway in many states to improve the specificity of drugs listed on death certificates (24,25). It is possible that search terms for certain drugs rarely seen in drug overdose deaths were not included despite the multiple avenues taken to develop the list of search terms.

Future directions

Data from the literal text could potentially be used to detect emerging trends in drug-involved mortality. For instance, the methods used in this report could be modified to identify deaths involving newly approved prescription drugs, new illicit drugs, and other health threats. Furthermore, the software programs used to mine the literal text could be modified to help identify emergent trends in drug-involved mortality, even before the annual mortality statistical files are finalized. With the rise of synthetic drugs, such as the fentanyl analogs (26), this may be necessary in the future. In order to detect emerging trends, periodically updating the text search capabilities is critical to surveillance of drug overdose deaths.

The amount of information that can be extracted from the literal text is a function of the level of detail that certifiers provide. There are general references that provide guidance on filling out death certificates that describe the importance of details (27,28). In addition to these general references, there is guidance for certifying drug overdose deaths, which stresses the importance of including the specific drugs involved (24). Because of the importance of including the specific drugs on death certificates for public health purposes, there are recommendations to help epidemiologists develop partnerships to help improve specificity of drugs on the death certificates (25).

Currently, the literal text analysis methodology focuses on using the contextual information to identify the mentions of drugs involved in the death. In the future, additional analysis of the contextual information may be informative. For instance, the method could be used to explore the route of administration (e.g., inhalation, injection, or transdermal), specific drug effects (e.g., anaphylaxis), and antibiotic resistance.

Conclusion

This report details a new method that was developed to extract information from the National Vital Statistics System death certificate literal text to improve national monitoring of drug-involved mortality. The literal text analysis method described in this report leverages existing information on the death certificates for statistical monitoring of drug-involved mortality deaths. Assessments conducted during the methods development process demonstrate that these methods have high accuracy in identifying the drugs mentioned and involved in mortality as well as the corresponding deaths. These methods could be applied to analyze mortality data for causes of death classified to broad ICD categories or for emerging causes of death with no ICD code assigned. Although the methods are limited by the level of drug-specific detail provided in the death certificate literal text, these methods are an enhancement to current ICD–10-coded mortality data.

References

1. Rudd RA, Aleshire N, Zibbell JE, Gladden RM. Increases in drug and opioid overdose deaths—United States, 2000–2014. MMWR Morb Mortal Wkly Rep 64(50–51):1378–82. 2016. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6450a3.htm.
2. Hedegaard H, Chen LH, Warner M. Drug-poisoning deaths involving heroin: United States, 2000–2013. NCHS data brief, no 190. Hyattsville, MD: National Center for Health Statistics. 2015. Available from: http://www.cdc.gov/nchs/data/databriefs/db190.pdf.
3. Chen LH, Hedegaard H, Warner M. Drug-poisoning deaths involving opioid analgesics: United States, 1999–2011. NCHS data brief, no 166. Hyattsville, MD: National Center for Health Statistics. 2014. Available from: http://www.cdc.gov/nchs/data/databriefs/db166.pdf.
4. Centers for Disease Control and Prevention. Vital signs: Overdoses of prescription opioid pain relievers and other drugs among women—United States, 1999–2010. MMWR Morb Mortal Wkly Rep 62(26):537–42. 2013. Available from: http://www.cdc.gov/mmWr/preview/mmwrhtml/mm6226a3.htm.
5. Kochanek KD, Murphy SL, Xu JQ, Tejada-Vera B. Deaths: Final data for 2014. National vital statistics reports; vol 65 no 4. Hyattsville, MD: National Center for Health Statistics. 2016. Available from: http://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_04.pdf.
6. World Health Organization. International statistical classification of diseases and related health problems, tenth revision (ICD–10). Volume 1. 1992.
7. World Health Organization. International statistical classification of diseases and related health problems, tenth revision (ICD–10). Volume 3, Section III: Table of drugs and chemicals. 2011.
8. Davis J, Sabel J, Wright D, Slavova S. Epi tool to analyze overdose death data. Council of State and Territorial Epidemiologists blog post. 2015. Available from: http://www.cste.org/blogpost/1084057/211072/Epi-Tool-to-Analyze-Overdose-Death-Data.
9. Kim SY, Shapiro-Mendoza CK, Chu SY, Camperlengo LT, Anderson RN. Differentiating cause-of-death terminology for deaths coded as sudden infant death syndrome, accidental suffocation, and unknown cause: An investigation using US death certificates, 2003–2004. J Forensic Sci 57(2):364–9. 2012.
10. Koopman B, Zuccon G, Nguyen A, Bergheim A, Grayson N. Automatic ICD–10 classification of cancers from free-text death certificates. Int J Med Inform 84(11):956–65. 2015.
11. Ossiander EM. Using textual cause-of-death data to study drug poisoning deaths. Am J Epidemiol 179(7):884–94. 2014.
12. Shapiro-Mendoza CK, Kim SY, Chu SY, Kahn E, Anderson RN. Using death certificates to characterize sudden infant death syndrome (SIDS): Opportunities and limitations. J Pediatr 156(1):38–43. 2010.
13. Holman RC, Belay ED, Christensen KY, Maddox RA, Miniño AM, Folkema AM, et al. Human prion diseases in the United States. PLoS One 5(1):e8521. 2010.
14. Warner M, Trinidad JP, Bastian BA, et al. Drugs most frequently involved in drug overdose deaths: United States, 2010–2014. National vital statistics reports; vol 65 no 10. Hyattsville, MD: National Center for Health Statistics. 2016.
15. National Center for Health Statistics. Report of the panel to evaluate the U.S. standard certificates. 2000. Available from: https://www.cdc.gov/nchs/data/dvs/panelreport_acc.pdf.
16. National Center for Health Statistics. U.S. Standard Certificate of Death. 2003 revision. Available from: http://www.cdc.gov/nchs/data/dvs/death11-03final-acc.pdf.
17. National Center for Health Statistics. ICD–10 mortality manual parts 2a, 2b, and 2s. Available from: http://www.cdc.gov/nchs/nvss/instruction_manuals.htm.
18. Rowe M. Personal correspondence. 2016.
19. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Drug Abuse Warning Network methodology report, 2010 update. 2010.
20. Food and Drug Administration. FDA Adverse Event Reporting System. Available from: http://www.fda.gov/Drugs/InformationOnDrugs/ucm135151.htm.
21. SAS Institute Inc. SAS (Release 9.3) [computer software]. 2012.
22. World Health Organization. International statistical classification of diseases and related health problems, tenth revision (ICD–10). 2nd ed. Geneva, Switzerland. 2004.
23. Warner M, Paulozzi LJ, Nolte KB, Davis GG, Nelson LS. State variation in certifying manner of death and drugs involved in drug intoxication deaths. Acad Forensic Pathol 3(2)231–7. 2013.
24. Davis GG, National Association of Medical Examiners and American College of Medical Toxicology Expert Panel on Evaluating and Reporting Opiod Deaths. Complete republication: National Association of Medical Examiners position paper: Recommendations for the investigation, diagnosis, and certification of deaths related to opioid drugs. J Med Toxicol (10)1:100–6. 2014.
25. Sabel J, Poel A, Tuazon E, Paone D, Slavova S, Bunn T, et al. Recommendations and lessons learned for improved reporting of drug overdose deaths on death certificates. Council of State and Territorial Epidemiologists. 2016. Available from: http://c.ymcdn.com/sites/www.cste.org/resource/resmgr/PDFs/PDFs2/4_25_2016_FINAL-Drug_Overdos.pdf.
26. Gladden RM, Martinez P, Seth P. Fentanyl law enforcement submissions and increases in synthetic opioid-involved overdose deaths—27 states, 2013–2014. MMWR Morb Mortal Wkly Rep 65(33)837–43. 2016. Available from: http://www.cdc.gov/mmwr/volumes/65/wr/mm6533a2.htm.
27. National Center for Health Statistics. Medical examiners' and coroners' handbook on death registration and fetal death reporting. 2003. Available from: http://www.cdc.gov/nchs/data/misc/hb_me.pdf.
28. Hanzlick R. Cause of death and the death certificate: Important information for physicians, coroners, medical examiners, and the public. North Field, IL: College of American Pathologists. 2006.

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 4551
Hyattsville, MD 20782–2064

For more NCHS NVSRs, visit:
http://www.cdc.gov/nchs/products/nvsr.htm.

National Vital Statistics Reports, Vol. 65, No. 9, December 20, 2016

Contents

Acknowledgments

This joint project between the National Center for Health Statistics and the the Office of Surveillance and Epidemiology within the Center for Drug Evaluation and Research of the U.S. Food and Drug Administration (FDA) utilized the information on deaths involving drugs captured in the National Vital Statistics System–Mortality linked with death certificate literal text. The authors would like to acknowledge the contributions of Dr. Louis An, a pharmacist with the Office of Surveillance and Epidemiology who assisted in the verification of mappings from the individual drug search terms to the principal variants.

Suggested citation

Trinidad JP, Warner M, Bastian BA, et al. Using literal text from the death certificate to enhance mortality statistics: Characterizing drug involvement in deaths. National vital statistics reports; vol 65 no 9. Hyattsville, MD: National Center for Health Statistics. 2016.

Copyright information

National Center for Health Statistics

For e-mail updates on NCHS publication releases, subscribe online at: http://www.cdc.gov/nchs/govdelivery.htm.
For questions or general information about NCHS: Tel: 1–800–CDC–INFO (1–800–232–4636) • TTY: 1–888–232–6348
Internet: http://www.cdc.gov/nchs • Online request form: http://www.cdc.gov/info
DHHS Publication No. 2017–1120 • CS272542

National Vital Statistics Reports

Volume 67, Number 9 December 12, 2018

Drugs Most Frequently Involved in Drug Overdose Deaths: United States, 2011–2016

by Holly Hedegaard, M.D., M.S.P.H., and Brigham A. Bastian, B.S., National Center for Health Statistics; James P. Trinidad, M.P.H., M.S., U.S. Food and Drug Administration; Merianne Spencer, M.P.H., and Margaret Warner, Ph.D., National Center for Health Statistics

Abstract

Objective—This report identifies the specific drugs involved most frequently in drug overdose deaths in the United States from 2011 through 2016.

Methods—Record-level data from the 2011–2016 National Vital Statistics System–Mortality files were linked to electronic files containing literal text information from death certificates. Drug overdose deaths were identified using the *International Classification of Diseases, Tenth Revision* underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Drug mentions were identified by searching the literal text in three fields of the death certificate: the causes of death from Part I, significant conditions contributing to death from Part II, and a description of how the injury occurred. Contextual information was used to determine drug involvement in the death. Descriptive statistics were calculated for drug overdose deaths involving the 10 most frequently mentioned drugs. Deaths involving more than one drug (e.g., a death involving both heroin and cocaine) were counted in all relevant drug categories (e.g., the same death was included in counts of heroin deaths and in counts of cocaine deaths).

Results—Among drug overdose deaths that mentioned at least one specific drug, the 10 most frequently mentioned drugs during 2011–2016 included fentanyl, heroin, hydrocodone, methadone, morphine, oxycodone, alprazolam, diazepam, cocaine, and methamphetamine. Oxycodone ranked first in 2011, heroin during 2012–2015, and fentanyl in 2016. During the study period, cocaine consistently ranked second or third. From 2011 through 2016, the age-adjusted rate of drug overdose deaths involving heroin more than tripled, as did the rate of drug overdose deaths involving methamphetamine. The rate of drug overdose deaths involving fentanyl and fentanyl analogs doubled each year from 2013 through 2016, from 0.6 per 100,000 in 2013 to 1.3 in 2014, 2.6 in 2015, and 5.9 in 2016. The rate of overdose deaths involving methadone decreased from 1.4 per 100,000 in 2011 to 1.1 in 2016. The 10 most frequently mentioned drugs often were found in combination with each other. The drugs most frequently mentioned varied by the intent of the drug overdose death. In 2016, the drugs most frequently mentioned in unintentional drug overdose deaths were fentanyl, heroin, and cocaine, while the drugs most frequently mentioned in suicides by drug overdose were oxycodone, diphenhydramine, hydrocodone, and alprazolam.

Conclusions—This report identifies patterns in the specific drugs most frequently involved in drug overdose deaths from 2011 through 2016 and highlights the importance of complete and accurate reporting in the literal text on death certificates.

Keywords: opioid • fentanyl • heroin • cocaine • National Vital Statistics System

Introduction

From 1999 through 2016, the age-adjusted rate of drug overdose deaths in the United States more than tripled from 6.1 per 100,000 to 19.8 per 100,000 (1). Multiple studies have used National Vital Statistics System–Mortality (NVSS–M) data, coded using the *International Classification of Diseases, Tenth Revision* (ICD–10), to examine patterns of drug involvement in overdose deaths (1–5). ICD–10 is the classification system used in the United States to categorize the underlying and multiple causes of death (6). One limitation of this classification system is that, with a few exceptions, ICD–10 codes reflect broad categories of drugs rather than unique specific drugs. For example, oxycodone and hydrocodone are both classified in the same category of natural and semisynthetic opioid analgesics (ICD–10 code T40.2). The broad drug categorizations used in ICD–10 make it difficult to use ICD–10-coded data to monitor trends in deaths involving specific drugs (e.g., deaths involving oxycodone specifically).

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics
National Vital Statistics System

To address this limitation, the National Center for Health Statistics (NCHS) and the U.S. Food and Drug Administration (FDA) collaboratively developed methods to search the literal text from death certificates to identify mentions of specific drugs and other substances, and to search contextual terms to identify involvement of the drug(s) or substance(s) in the death (7). The literal text is the written information provided by the medical certifier, usually a medical examiner or coroner in the case of drug overdose deaths (8,9), and describes the cause of death and other factors or circumstances that contributed to the death. The methods developed by NCHS and FDA search three literal text fields from the U.S. standard death certificate: the causes of death from Part I, significant conditions contributing to death from Part II, and a description of how the injury occurred (7,10).

A previous study presented the findings from use of literal text analysis to identify the specific drugs most frequently involved in drug overdose deaths from 2010 through 2014 (11). This report uses the same methodology and an enhanced search term list to provide results for drug overdose deaths from 2011 through 2016.

Methods

Data source and study population

NVSS–M data from 2011 through 2016 were used in this descriptive analysis. NVSS–M data contain cause-of-death, demographic, and geographic information extracted from death certificates (12). The study population was limited to decedents who were U.S. residents with an ICD–10 underlying cause-of-death of drug overdose: X40–X44 (unintentional), X60–X64 (suicide or intentional self-harm), X85 (assault), and Y10–Y14 (undetermined intent). During the study period, the manner of death was unintentional for 80%–86% of drug overdose deaths, suicide for 8%–13%, homicide for 0.2%, and undetermined intent for 6%–7% (1,13). The underlying cause-of-death codes reflect deaths resulting from acute intoxication from drugs (i.e., drug overdose). Deaths from chronic exposure to drugs (e.g., liver toxicity) or adverse effects experienced from therapeutic or prophylactic dosages of drug were not included. Use of this code set (X40–X44, X60–X64, X85, and Y10–Y14) is consistent with other NCHS publications on drug overdose deaths and facilitates comparison with other analyses using the ICD–10-coded data (1).

NVSS–M files were linked to electronic files containing literal text data, also extracted from death certificates (7). Mentions of drugs or other substances (described below) were identified using the literal text data from three fields of the death certificate: the causes of death from Part I, significant conditions contributing to death from Part II, and a description of how the injury occurred.

Identifying drug mentions and involvement of the drug in the death

The method for searching literal text information to characterize the drugs involved in deaths has been described elsewhere (7). Briefly, the method involves searching the literal text for mentions of drugs and other substances, as well as terms that provide context about the involvement of the drug in the death (i.e., whether the drug contributed to the death). For example, the phrase "METHICILLIN RESISTANT STAPHYLOCOCCUS AUREUS INFECTION" does not suggest drug involvement in mortality, but rather a type of bacterial infection. Similarly, the phrase "NOT DRUG RELATED" clearly indicates that the death did not involve a drug, even though "DRUG" is mentioned in the phrase. The drug or substance mentioned in a literal text field is assumed to be involved in the death unless the contextual information indicates otherwise. Software programs, referred to as the Drug Mention with Involvement (DMI) programs, have been developed using SAS version 9.4 to automate the process (7).

DMI programs identify mentions of drugs and other substances using various search terms. Search terms include generic drug names, brand names, common usage or street names, abbreviations, metabolites, misspellings, and other variations. The list of search terms used in this report is broader than that used in a previous report (7), and was developed to maintain as much substance specificity as possible. The new search term list was applied to the literal text for all years of the study (2011–2016). Because a new search term list was used in this analysis, the results for 2011–2014 may differ slightly from those reported previously (from 0 to 36 additional deaths depending on the drug and the year) (11).

Each search term was mapped to a "principal variant," the overarching label assigned to a drug, a drug class, or exposure not otherwise specified. For example, terms such as "COCAIEN", "COCAINE CRACK", "COCAINE HYDROCHLORIDE", and "COCAINETOXICITY" were all mapped to the principal variant "COCAINE". In general, the principal variant was the generic drug name. Some search terms—mostly for combination drug products—were mapped to two or more principal variants. Use of principal variants makes it possible to generate aggregate counts for all search terms that refer to the same drug or substance. Principal variants also were categorized according to whether they referred to specific drugs or substances (e.g., methadone), classes of drugs or substances (e.g., opioids), or nonspecific references to exposures to drugs (e.g., words such as "DRUG", "MULTIDRUG", or "POLYPHARMACY"). The DMI Search Terms and principal variants table is provided in an accompanying CSV file (https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Program_Code/oae/).

A frequency distribution of the principal variants identified the top 20 drugs for each year from 2011 through 2016. A referent drug category was created for each of the top 20 drugs. The term "referent drug" in the tables and figures in this report generally refers to a single principal variant for the drug of interest. However, due to the greater detail in the updated principal variant list, some of the referent drug categories are comprised of two or more principal variants, generally reflecting

a drug and its metabolites. For example, the principal variant HYDROCODONE and the principal variant NORHYDROCODONE (a metabolite of hydrocodone) were grouped together to create the referent drug category of HYDROCODONE. The referent group FENTANYL included fentanyl as well as fentanyl metabolites, precursors, and analogs. The grouping of principal variants into referent groups was based on expertise from FDA and NCHS. The referent groups table, which contains a list of search terms and the principal variants included in each referent drug category, is provided in the CSV file (https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Program_Code/oae/).

Analysis

Results are reported as numbers, percentages, or rates for the deaths involving the referent drug. Deaths involving more than one referent drug (e.g., a death involving both heroin and cocaine) are counted in all relevant drug categories (e.g., the same death is included in counts of heroin deaths and in counts of cocaine deaths); therefore counts are not mutually exclusive. Age-adjusted death rates are calculated using the direct method and the 2000 standard U.S. population (12). Trends in age-adjusted death rates are evaluated using the National Cancer Institute's Joinpoint Regression Program (version 4.6.0.0) (14). Joinpoint software fitted weighted least-squares regression models to the rates on the log transform scale. Allowing one observed time point at each end and two for the middle line segments, the Grid Search Algorithm searched for a maximum of two joinpoints at an overall alpha level of 0.05 (15). Any mention of an annual percent change in this report indicates a statistically significant trend. Comparisons of rates between years were tested for statistical significance using methods described elsewhere (12).

Analyses of mentions of other drugs reported in addition to the referent drug (concomitant drugs) were also conducted. Only deaths with mention of at least two specific drugs (the referent drug and at least one concomitant drug) are included in this analysis. Alcohol, nicotine, and nondrug substances are not included in the analysis.

The numbers and rates of drug-specific overdose deaths shown in the tables and figures should be considered the minimum number or rate for that referent drug category because there could be additional deaths in which the drug was involved, but the drug was not reported in the literal text on the death certificate.

Assessing improvement in reporting on death certificates

The ICD–10 multiple-cause codes T36–T50.8 provide information on the types of drugs or drug classes involved in the death. The percentage of deaths with an underlying-cause code of X40–X44, X60–X64, X85, and Y10–Y14 that have a multiple-cause code of T36–T50.8 is a measure of the specificity of reporting of drugs or drug classes in drug overdose deaths. This measure was used to assess possible changes in reporting through the years of the study. The percentage of drug overdose deaths with codes T36–T50.8 increased each year (75% in 2011, 76% in 2012, 78% in 2013, 81% in 2014, 83% in 2015, and 85% in 2016).

This improvement in reporting of specific drugs and drug classes during the study period could potentially influence the observed trends in drug overdose deaths for specific drugs (Figures 1–3). To assess the possible influence of improved reporting, an adjustment analysis was conducted. In this analysis, an adjustment factor was applied to each number and age-adjusted rate for drug overdose deaths involving specific drugs. The adjustment factor assumed that the specificity of drug reporting remained constant from 2011 through 2016 at the 2016 rate (i.e., 85.4% of drug overdose deaths with an ICD–10 multiple-cause code of T36–T50.8). A description of the methodology and the results from the adjustment analysis are provided in the Technical Notes.

Results

The number of drug overdose deaths per year increased 54%, from 41,340 deaths in 2011 to 63,632 deaths in 2016 (Table A). From the literal text analysis, the percentage of drug overdose deaths mentioning at least one specific drug or substance increased from 73% of the deaths in 2011 to 85% of the deaths in 2016. The percentage of drug overdose deaths that mentioned only a drug class but not a specific drug or substance declined from 5.1% of deaths in 2011 to 2.5% in 2016. Review of the literal text for these deaths indicated that the deaths that mentioned only a drug class frequently involved either an opioid or an opiate (ranging from 54% in 2015 to 60% in 2016). The percentage of deaths that did not mention a specific drug or substance or a drug class declined from 22% of drug overdose deaths in 2011 to 13% in 2016.

Most frequently mentioned drugs

Table B shows the relative ranking of the top 15 drugs involved in drug overdose deaths for each year from 2011 through 2016 among deaths that mentioned at least one specific drug. The number of deaths for each drug should be interpreted in light of the improvements in reporting as described in Table A, and should be considered the minimum number for that drug because there could be additional deaths in which the drug was involved, but the drug was not reported in the literal text.

The top 15 drugs were identified based on the number of drug overdose deaths per referent drug category. While the ranking changed from year to year, the top 10 drugs involved in overdose deaths remained consistent throughout the 6-year period. The top 10 drugs belonged to three drug classes:

- Opioids: fentanyl, heroin, hydrocodone, methadone, morphine, and oxycodone
- Benzodiazepines: alprazolam and diazepam
- Stimulants: cocaine and methamphetamine

The drugs that ranked 11–15 varied from year to year and included such drugs as diphenhydramine, citalopram, acetaminophen, carisoprodol, tramadol, oxymorphone, amitriptyline, clonazepam, gabapentin, and amphetamine.

Table A. Number and percentage of drug overdose deaths with mention of a specific drug, with mention of only a drug class, and with no mention of a drug class or specific drug: United States, 2011–2016

Drug overdose deaths	2011		2012		2013		2014		2015		2016	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
All drug overdose deaths	41,340	100.0	41,502	100.0	43,982	100.0	47,055	100.0	52,404	100.0	63,632	100.0
Drug overdose deaths with mention of at least one specific drug or other substance	30,103	72.8	30,923	74.5	33,640	76.5	37,631	80.0	43,141	82.3	54,137	85.1
Drug overdose deaths with mention of a class only (no specific drug or other substance)	2,122	5.1	2,093	5.0	1,918	4.4	1,653	3.5	1,650	3.1	1,566	2.5
Drug overdose deaths without mention of a specific drug, other substance, or class[1]	9,115	22.0	8,486	20.4	8,424	19.2	7,771	16.5	7,613	14.5	7,929	12.5

[1]Category includes drug overdose deaths with mentions of substances not otherwise specified (NOS) (e.g., mention of "POLYPHARMACY" or "DRUG"), uninformative text, and drug overdose deaths with no mentions identified (e.g., text stating "OVERDOSE" with no mention of a drug).

NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* (ICD–10) underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Percentages may not add to 100 due to rounding. The reporting of at least one specific drug or drug class in the literal text, as identified using ICD–10 multiple cause-of-death codes T36-T50.8, improved from 75% of drug overdose deaths in 2011 to 85% of drug overdose deaths in 2016.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text, 2011–2016.

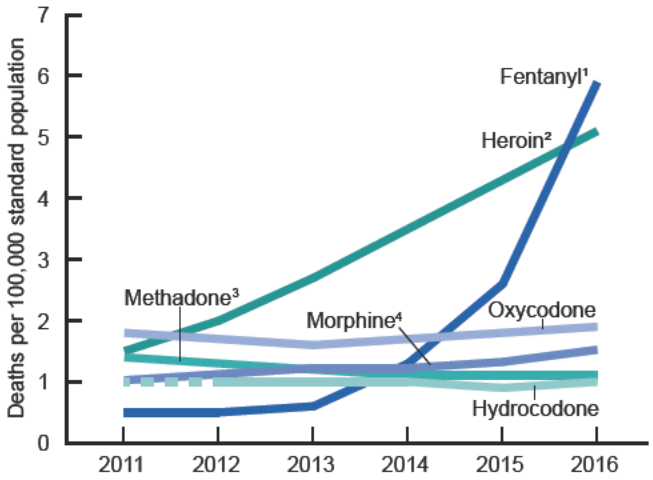
For the top 15 drugs:

- Among drug overdose deaths that mentioned at least one specific drug, oxycodone ranked first in 2011, heroin from 2012 through 2015, and fentanyl in 2016.
- In 2011 and 2012, fentanyl was mentioned in approximately 1,600 drug overdose deaths each year, but mentions increased in 2013 (1,919 deaths), 2014 (4,223 deaths), 2015 (8,251 deaths), and 2016 (18,335 deaths). In 2016, 29% of all drug overdose deaths mentioned involvement of fentanyl.
- The number of drug overdose deaths involving heroin increased threefold, from 4,571 deaths or 11% of all drug overdose deaths in 2011 to 15,961 deaths or 25% of all drug overdose deaths in 2016.
- Throughout the study period, cocaine ranked second or third among the top 15 drugs. From 2014 through 2016, the number of drug overdose deaths involving cocaine nearly doubled from 5,892 to 11,316.
- The number of drug overdose deaths involving methamphetamine increased 3.6-fold, from 1,887 deaths in 2011 to 6,762 deaths in 2016.
- The number of drug overdose deaths involving methadone decreased from 4,545 deaths in 2011 to 3,493 deaths in 2016.

Age-adjusted rates for drug overdose deaths involving the most frequently mentioned drugs, 2011–2016

Trends from 2011 through 2016 in the age-adjusted rates of drug overdose deaths involving the 10 most frequently mentioned drugs are shown in Figures 1–3. Improvements in reporting should be considered when interpreting these trends (see Technical Notes). As a reference, from 2011 through 2016, the age-adjusted rate of all drug overdose deaths, whether or not a specific drug was mentioned, increased from 13.2 per 100,000 to 19.8, an average increase of 9% per year.

- From 2011 through 2016, the age-adjusted rate of drug overdose deaths involving heroin more than tripled from 1.5 per 100,000 population to 5.1. The rate increased on average by about 34% per year from 2011 through 2014, and by about 20% per year from 2014 through 2016 (Figure 1, Table).



[1]Significant increasing trend for 2013–2016, $p < 0.05$.
[2]Significant increasing trend for 2011–2016 with different rates of change over time, $p < 0.05$.
[3]Significant decreasing trend for 2011–2014, $p < 0.05$.
[4]Significant increasing trend for 2011–2015, $p < 0.05$.
NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Deaths may involve other drugs in addition to the referent drug (i.e., the one listed). Deaths involving more than one drug (e.g., a death involving both heroin and cocaine) are counted in both totals. Caution should be used when comparing numbers across years. The reporting of at least one specific drug in the literal text improved from 73% of drug overdose deaths in 2011 to 85% of drug overdose deaths in 2016.
SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text, 2011–2016.

Figure 1. Age-adjusted rates for drug overdose deaths involving selected opioids, 2011–2016

Table B. Top 15 drugs involved in drug overdose deaths: United States, 2011–2016

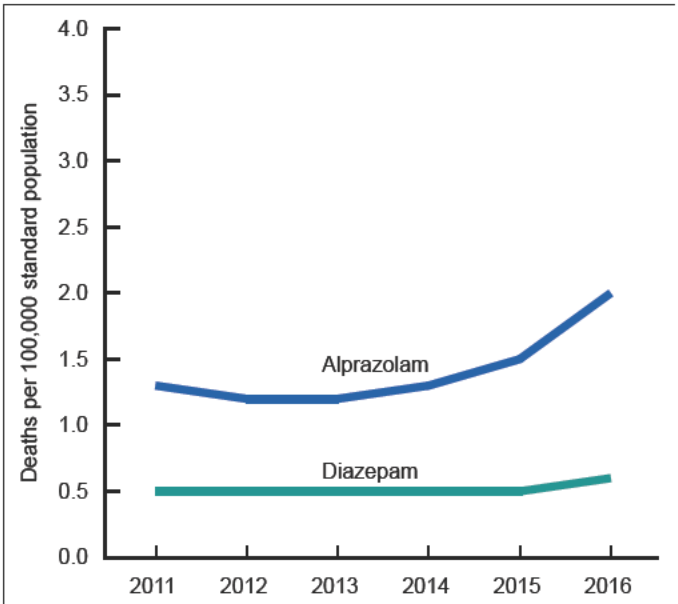
	2011 (*n* = 41,340)			2012 (*n* = 41,502)			2013 (*n* = 43,982)		
Rank[1]	Referent drug	Number of deaths[2]	Percent of deaths[3]	Referent drug	Number of deaths[2]	Percent of deaths[3]	Referent drug	Number of deaths[2]	Percent of deaths[3]
1	Oxycodone	5,587	13.5	Heroin	6,155	14.8	Heroin	8,418	19.1
2	Cocaine	5,070	12.3	Oxycodone	5,178	12.5	Cocaine	5,319	12.1
3	Heroin	4,571	11.1	Cocaine	4,780	11.5	Oxycodone	4,967	11.3
4	Methadone	4,545	11.0	Methadone	4,087	9.8	Morphine	3,772	8.6
5	Alprazolam	4,066	9.8	Alprazolam	3,803	9.2	Alprazolam	3,724	8.5
6	Morphine	3,290	8.0	Morphine	3,513	8.5	Methadone	3,700	8.4
7	Hydrocodone	3,206	7.8	Hydrocodone	3,037	7.3	Methamphetamine	3,194	7.3
8	Methamphetamine	1,887	4.6	Methamphetamine	2,267	5.5	Hydrocodone	3,113	7.1
9	Diazepam	1,698	4.1	Fentanyl	1,615	3.9	Fentanyl	1,919	4.4
10	Fentanyl	1,662	4.0	Diazepam	1,577	3.8	Diazepam	1,618	3.7
11	Diphenhydramine	1,226	3.0	Diphenhydramine	1,300	3.1	Diphenhydramine	1,360	3.1
12	Oxymorphone	1,190	2.9	Citalopram	1,042	2.5	Tramadol	1,009	2.3
13	Citalopram	1,043	2.5	Tramadol	935	2.3	Clonazepam	946	2.2
14	Acetaminophen	879	2.1	Oxymorphone	866	2.1	Citalopram	914	2.1
15	Tramadol	849	2.1	Amitriptyline	835	2.0	Amitriptyline	815	1.9

	2014 (*n* = 47,055)			2015 (*n* = 52,404)			2016 (*n* = 63,632)		
Rank[1]	Referent drug	Number of deaths[2]	Percent of deaths[3]	Referent drug	Number of deaths[2]	Percent of deaths[3]	Referent drug	Number of deaths[2]	Percent of deaths[3]
1	Heroin	10,882	23.1	Heroin	13,318	25.4	Fentanyl	18,335	28.8
2	Cocaine	5,892	12.5	Fentanyl	8,251	15.7	Heroin	15,961	25.1
3	Oxycodone	5,431	11.5	Cocaine	7,324	14.0	Cocaine	11,316	17.8
4	Alprazolam	4,237	9.0	Oxycodone	5,792	11.1	Methamphetamine	6,762	10.6
5	Fentanyl	4,223	9.0	Methamphetamine	5,092	9.7	Alprazolam	6,209	9.8
6	Morphine	4,024	8.6	Alprazolam	4,801	9.2	Oxycodone	6,199	9.7
7	Methamphetamine	3,747	8.0	Morphine	4,226	8.1	Morphine	5,014	7.9
8	Methadone	3,498	7.4	Methadone	3,376	6.4	Methadone	3,493	5.5
9	Hydrocodone	3,299	7.0	Hydrocodone	3,051	5.8	Hydrocodone	3,199	5.0
10	Diazepam	1,748	3.7	Diphenhydramine	1,798	3.4	Diazepam	2,022	3.2
11	Diphenhydramine	1,614	3.4	Diazepam	1,796	3.4	Diphenhydramine	2,008	3.2
12	Tramadol	1,175	2.5	Clonazepam	1,328	2.5	Clonazepam	1,656	2.6
13	Clonazepam	1,139	2.4	Gabapentin	1,222	2.3	Gabapentin	1,546	2.4
14	Citalopram	1,014	2.2	Tramadol	1,177	2.2	Tramadol	1,250	2.0
15	Oxymorphone	909	1.9	Oxymorphone	1,006	1.9	Amphetamine	1,193	1.9

[1]Ranks were not tested for statistical significance.
[2]Number of drug overdose deaths involving the referent drug.
[3]Percentage of drug overdose deaths involving the referent drug.

NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* (ICD–10) underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Deaths may involve other drugs in addition to the referent drug (i.e., the one listed). Deaths involving more than one drug (e.g., a death involving both heroin and cocaine) are counted in both totals. Caution should be used when comparing numbers across years. The reporting of at least one specific drug or drug class in the literal text, as identified using ICD–10 multiple cause-of-death codes T36–T50.8, improved from 75% of drug overdose deaths in 2011 to 85% of drug overdose deaths in 2016.

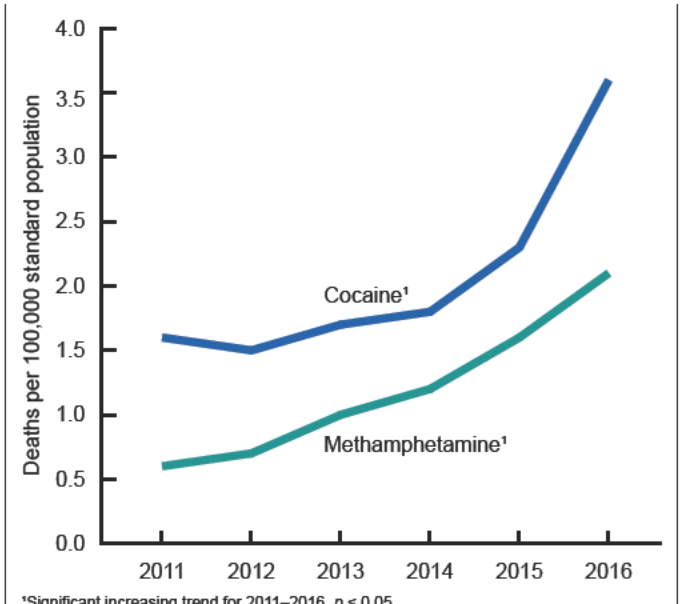
SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text, 2011–2016.



NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Deaths may involve other drugs in addition to the referent drug (i.e., the one listed). Deaths involving more than one drug (e.g., a death involving both heroin and cocaine) are counted in both totals. Caution should be used when comparing numbers across years. The reporting of at least one specific drug in the literal text improved from 73% of drug overdose deaths in 2011 to 85% of drug overdose deaths in 2016.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text, 2011–2016.

Figure 2. Age-adjusted rates for drug overdose deaths involving selected benzodiazepines, 2011–2016



¹Significant increasing trend for 2011–2016, $p < 0.05$.

NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Deaths may involve other drugs in addition to the referent drug (i.e., the one listed). Deaths involving more than one drug (e.g., a death involving both heroin and cocaine) are counted in both totals. Caution should be used when comparing numbers across years. The reporting of at least one specific drug in the literal text improved from 73% of drug overdose deaths in 2011 to 85% of drug overdose deaths in 2016.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text, 2011–2016.

Figure 3. Age-adjusted rates for drug overdose deaths involving selected stimulants, 2011–2016

- From 2011 through 2013, there was no statistical change in the age-adjusted rate of drug overdose deaths involving fentanyl. From 2013 through 2016, the rate increased on average by about 113% per year, from 0.6 per 100,000 population in 2013, to 1.3 in 2014, 2.6 in 2015, and 5.9 in 2016.
- The age-adjusted rate of drug overdose deaths involving morphine increased from 1.0 per 100,000 population in 2011 to 1.5 in 2016. The rate increased on average by about 6% per year from 2011 through 2015. Between 2015 and 2016, the rate changed 18%, however, this trend was not statistically significant.
- The age-adjusted rate of drug overdose deaths involving methadone decreased from 1.4 per 100,000 population in 2011 to 1.1 in 2016. The rate decreased on average by about 10% per year from 2011 through 2014. From 2014 through 2016, there was no significant change in the rate.
- From 2011 through 2016, there was no significant change in the age-adjusted death rate for drug overdose deaths involving hydrocodone.
- The age-adjusted rate of drug overdose deaths involving oxycodone decreased from 1.8 per 100,000 population in 2011 to 1.6 in 2013, then increased to 1.9 in 2016; however, these decreasing and increasing trends were not statistically significant.
- The age-adjusted rate of drug overdose deaths involving alprazolam decreased from 1.3 per 100,000 population in 2011 to 1.2 in 2013, then increased to 2.0 in 2016; however, these decreasing and increasing trends were not statistically significant (Figure 2, Table).
- From 2011 through 2016, there was no significant change in the age-adjusted rate for drug overdose deaths involving diazepam.
- The age-adjusted rate of drug overdose deaths involving cocaine increased from 1.6 per 100,000 population in 2011 to 3.6 in 2016. The rate increased on average by about 18% per year (Figure 3, Table).
- The age-adjusted rate of drug overdose deaths involving methamphetamine more than tripled from 0.6 per 100,000 population in 2011 to 2.1 in 2016. The rate increased on average by about 29% per year.
- In the adjustment analysis, the findings for the trends in rates based on observed and adjusted numbers were, in general, the same for fentanyl, oxycodone, diazepam, cocaine, and methamphetamine (see Technical Notes). For heroin, the inflection point in 2014 was no longer found, resulting in a percent change in the rate of about 24% per year from 2011 through 2016. For morphine, the inflection point in 2015 was no longer found, resulting in a percent change in the rate of about 5% per year from 2011 through 2016. For methadone, rates decreased by about 12% per year from 2011 through 2014, and by about 3% from 2014 through 2016. For hydrocodone, there was a significant decline in the age-adjusted rates of about 4% per year from 2011 through 2016. For alprazolam, the inflection point in 2013 was no longer found, and as with the observed values, the increasing trend from 2011 through 2016 was not statistically significant.

Drug overdose deaths in 2016 involving multiple drugs

Table C shows the percentage of drug overdose deaths with concomitant involvement of other drugs for the top 10 drugs involved in drug overdose deaths in 2016. The percentage of deaths with concomitant involvement of other drugs varied by drug. For example, almost all drug overdose deaths involving alprazolam or diazepam (96%) mentioned involvement of other drugs. In contrast, 50% of the drug overdose deaths involving methamphetamine, and 69% of the drug overdose deaths involving fentanyl mentioned involvement of one or more other specific drugs.

Table D shows the most frequent concomitant drug mentions for each of the top 10 drugs involved in drug overdose deaths in 2016.

- Two in five overdose deaths involving cocaine also mentioned fentanyl.
- Nearly one-third of drug overdose deaths involving fentanyl also mentioned heroin (32%).
- Alprazolam was mentioned in 26% of the overdose deaths involving hydrocodone, 22% of the deaths involving methadone, and 25% of the deaths involving oxycodone.
- More than one-third of the overdose deaths involving cocaine also mentioned heroin (34%).
- More than 20% of the overdose deaths involving methamphetamine also mentioned heroin.

Most frequently mentioned drugs involved in drug overdose deaths in 2016, by intent of death

Table E shows the top 10 drugs involved in drug overdose deaths in 2016 by intent of death, for deaths in which at least one specific drug was identified. Results are shown for unintentional drug overdose deaths (ICD–10 underlying-cause codes X40–X44), suicides by drug overdose (ICD–10 underlying-cause codes X60–X64), and drug overdose deaths for which the intent could not be determined (undetermined intent; [ICD–10 underlying-cause codes Y10–Y14]). The results for 110 deaths with an intent of homicide (ICD–10 underlying-cause code X85) are not shown due to small numbers.

In 2016, unintentional drug overdose deaths most frequently mentioned fentanyl, heroin, and cocaine, while suicides by drug overdose most frequently mentioned oxycodone, diphenhydramine, hydrocodone, and alprazolam. Methadone ranked in the top 10 for unintentional and undetermined intent deaths, but not among suicides by drug overdose. Quetiapine, tramadol, bupropion, and zolpidem ranked in the top 10 for suicides by drug overdose, but not for unintentional drug overdose deaths and overdose deaths of undetermined intent.

Discussion

Findings for specific drugs

From 2011 through 2016, the number of drug overdose deaths increased by 54%, from 41,340 deaths in 2011 to 63,632 deaths in 2016. The most frequently mentioned drugs involved in these deaths were the opioids heroin, oxycodone, methadone, morphine, hydrocodone, and fentanyl; the benzodiazepines alprazolam and diazepam; and the stimulants cocaine and methamphetamine.

Among drug overdose deaths that mentioned at least one specific drug, oxycodone ranked first in 2011, heroin ranked first from 2012 through 2015, and fentanyl ranked first in 2016. Cocaine ranked second or third throughout the study period.

An analysis of trends among the most frequently mentioned drugs showed that, for several drugs, the age-adjusted rate of

Table C. Number and percentage of deaths with concomitant drug involvement for drug overdose deaths involving the top 10 drugs: United States, 2016

Referent drug	Number of drug overdose deaths involving the referent drug	Number of drug overdose deaths involving the referent drug and one or more concomitant drugs	Percentage of drug overdose deaths involving the referent drug and one or more concomitant drugs
Opioids			
Fentanyl	18,335	12,694	69.2
Heroin	15,961	11,248	70.5
Hydrocodone	3,199	2,743	85.7
Methadone	3,493	2,551	73.0
Morphine	5,014	4,175	83.3
Oxycodone	6,199	5,027	81.1
Benzodiazepines			
Alprazolam	6,209	5,970	96.2
Diazepam	2,022	1,951	96.5
Stimulants			
Cocaine	11,316	8,363	73.9
Methamphetamine	6,762	3,370	49.8

NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Only deaths with at least one specific drug identified are included in the analysis.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text, 2016.

Table D. Most frequent concomitant drugs for drug overdose deaths involving the top 10 drugs: United States, 2016

Referent drug	Number of drug overdose deaths involving the referent drug	Most frequent concomitant drug: Concomitant drug	Most frequent concomitant drug: Number and percentage[1] of deaths involving both drugs	Second most frequent concomitant drug: Concomitant drug	Second most frequent concomitant drug: Number and percentage[1] of deaths involving both drugs	Third most frequent concomitant drug: Concomitant drug	Third most frequent concomitant drug: Number and percentage[1] of deaths involving both drugs
Opioids							
Fentanyl	18,335	Heroin	5,915 (32.3)	Cocaine	4,598 (25.1)	Alprazolam	1,760 (9.6)
Heroin	15,961	Fentanyl	5,915 (37.1)	Cocaine	3,804 (23.8)	Alprazolam	1,668 (10.5)
Hydrocodone	3,199	Alprazolam	822 (25.7)	Oxycodone	551 (17.2)	Fentanyl	478 (14.9)
Methadone	3,493	Alprazolam	751 (21.5)	Fentanyl	528 (15.1)	Heroin	483 (13.8)
Morphine	5,014	Fentanyl	1,612 (32.1)	Cocaine	846 (16.9)	Heroin	687 (13.7)
Oxycodone	6,199	Alprazolam	1,571 (25.3)	Fentanyl	1,150 (18.6)	Morphine	668 (10.8)
Benzodiazepines							
Alprazolam	6,209	Fentanyl	1,760 (28.3)	Heroin	1,668 (26.9)	Oxycodone	1,571 (25.3)
Diazepam	2,022	Oxycodone	576 (28.5)	Fentanyl	502 (24.8)	Heroin	404 (20.0)
Stimulants							
Cocaine	11,316	Fentanyl	4,598 (40.6)	Heroin	3,804 (33.6)	Alprazolam	1,031 (9.1)
Methamphetamine	6,762	Heroin	1,477 (21.8)	Fentanyl	753 (11.1)	Cocaine	562 (8.3)

[1]Percentage of drug overdose deaths involving the referent drug that also involved the concomitant drug. Deaths may involve more than one concomitant drug in addition to the referent drug.

NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Only deaths with at least one specific drug identified are included in the analysis.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text, 2016.

Table E. Top 10 drugs involved in drug overdose deaths, by intent of death: United States, 2016

Rank[1]	Unintentional (*n* = 54,793): Referent drug	Unintentional (*n* = 54,793): Number of deaths[2]	Unintentional (*n* = 54,793): Percent of deaths[3]	Suicide (*n* = 5,086): Referent drug	Suicide (*n* = 5,086): Number of deaths[2]	Suicide (*n* = 5,086): Percent of deaths[3]	Undetermined (*n* = 3,643): Referent drug	Undetermined (*n* = 3,643): Number of deaths[2]	Undetermined (*n* = 3,643): Percent of deaths[3]
1	Fentanyl	16,981	31.0	Oxycodone	648	12.7	Fentanyl	1,185	32.5
2	Heroin	15,075	27.5	Diphenhydramine	576	11.3	Heroin	766	21.0
3	Cocaine	10,618	19.4	Hydrocodone	472	9.3	Morphine	619	17.0
4	Methamphetamine	6,448	11.8	Alprazolam	468	9.2	Cocaine	579	15.9
5	Alprazolam	5,510	10.1	Acetaminophen	343	6.7	Oxycodone	322	8.8
6	Oxycodone	5,225	9.5	Quetiapine	297	5.8	Methadone	264	7.2
7	Morphine	4,122	7.5	Morphine	268	5.3	Alprazolam	225	6.2
8	Methadone	3,110	5.7	Tramadol	266	5.2	Methamphetamine	195	5.4
9	Hydrocodone	2,556	4.7	Bupropion	264	5.2	Hydrocodone	169	4.6
10	Diazepam	1,723	3.1	Zolpidem	251	4.9	Diphenhydramine	152	4.2

[1]Ranks were not tested for statistical significance.
[2]Number of drug overdose deaths involving the referent drug.
[3]Percentage of drug overdose deaths involving the referent drug.

NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* underlying cause-of-death codes X40–X44 (unintentional), X60–X64 (suicide), and Y10–Y14 (undetermined). Only deaths with at least one specific drug identified are included in the analysis. The results for 110 deaths with an intent of homicide (X85) are not shown due to small numbers. Deaths may involve other drugs in addition to the referent drug (i.e., the one listed). Deaths involving more than one drug (e.g., a death involving both heroin and cocaine) are counted in both totals.

SOURCE: NCHS National Vital Statistics System, Mortality files linked with death certificate literal text, 2016.

drug overdose deaths increased considerably within a relatively short period. From 2011 through 2016, the rate of drug overdose deaths involving heroin more than tripled, as did the rate of drug overdose deaths involving methamphetamine. The rate of drug overdose deaths involving fentanyl and fentanyl analogs doubled each year from 2013 through 2016, from 0.6 per 100,000 in 2013 to 1.3 in 2014, 2.6 in 2015, and 5.9 in 2016. Among the drugs discussed in this report, only methadone showed a decreasing drug overdose death rate, from 1.4 per 100,000 in 2011 to 1.1 in 2016.

Results from the literal text analysis highlight the concomitant occurrence of more than one drug in many drug overdose deaths. For the top 10 drugs involved in drug overdose deaths in 2016, the proportion of deaths involving both the referent drug and at least one other concomitant drug ranged from 50% for methamphetamine to 96% for alprazolam or diazepam. Approximately 70% of drug overdose deaths involving fentanyl or heroin—the top two drugs involved in drug overdose deaths in 2016—involved at least one other specific drug.

The 10 most frequently mentioned drugs were often found in combination with each other. Drug combinations often involved drugs of different drug classes. For example, the opioid fentanyl and the stimulant cocaine were mentioned concomitantly in nearly 4,600 deaths. The opioid oxycodone and the benzodiazepine alprazolam were mentioned concomitantly in more than 1,500 deaths. In some instances, the most frequently mentioned concomitant drug was in the same drug class as the referent drug. For example, the opioids fentanyl and heroin were both mentioned in approximately 5,900 deaths. While the literal text can be used to identify the mention of the two drugs (fentanyl and heroin), the details to distinguish whether the heroin and fentanyl were taken as one (i.e., heroin laced with fentanyl) or as two separate drugs are often not available.

The drugs most frequently mentioned in the literal text varied by the intent of the drug overdose death. In 2016, unintentional drug overdose deaths most frequently mentioned fentanyl, heroin, and cocaine, while suicides by drug overdose more frequently mentioned oxycodone, diphenhydramine, hydrocodone, and alprazolam.

Data considerations and study limitations

This report used analysis of the literal text on death certificates to identify the drugs involved in overdose deaths (7). Software programs search the literal text for mentions of drugs and for terms that provide context about the involvement of the drug in the death. As shown in Table C, drug overdose deaths frequently involve multiple drugs. For deaths in which multiple drugs are involved, whether the death was caused by just one of the drugs present or was caused by a combination of some or all of the drugs present cannot be determined from the literal text analysis. This limitation in identifying the specific contribution of any given drug to the death should be considered when reviewing the findings in this report.

Reporting of deaths with at least one specific drug in the death certificate literal text improved from 73% of drug overdose deaths in 2011 to 85% in 2016. While improved reporting enhances the quality of the data, it also creates complexity in interpreting the trends and rankings observed. The findings in this report should be considered in light of the improvements in reporting. For example, some of the observed increase from 2011 through 2016 in drug overdose deaths involving the top 10 drugs is likely attributable to improvements in reporting. However, it is unlikely that the large increases seen for some drugs such as fentanyl, heroin, cocaine, and methamphetamine (i.e., drugs with an annual percent increase in mortality rates of 18% or greater) are due solely to improvements in reporting. True increases in the number of deaths involving these drugs are likely to have occurred. Similarly, decreases in rates such as those seen for drug overdose deaths involving methadone are likely to be, at least in part, due to a true decrease. It is also possible that the improvements in reporting could obscure real decreases. For example, using observed values, there was no statistically significant change in the age-adjusted rate of drug overdose deaths involving hydrocodone from 2011 through 2016. However, after adjustment for improved reporting (see Technical Notes), the age-adjusted rate of drug overdose deaths involving hydrocodone showed a significant decline of about 4% per year from 2011 through 2016.

Methods based on literal text analysis are dependent on the quality and completeness of the literal text, which may vary from jurisdiction to jurisdiction due to variation in death investigation and reporting practices or other differences in the medicolegal death investigation systems across the United States (16,17). Issues that contribute to variation in literal text information on drug overdose deaths have been discussed in detail elsewhere (11,18), and briefly, include certain factors.

Variation in death investigation practice and reporting—This includes whether or not toxicological laboratory testing is performed to determine the type(s) of drugs present. The substances tested for and the circumstances under which the tests are performed may vary by jurisdiction, decedent, and over time.

Interpretation of toxicology results—Interpretation of findings depends on which tests are ordered, the characteristics of the causative agent(s), the characteristics of the metabolites, and other evidence gathered during the investigation and examination.

Attribution to a specific drug—Some drugs have the same metabolites or are metabolites of other drugs, potentially resulting in misattribution of the specific drugs involved in the death. For example, mentions of morphine may actually refer to involvement of heroin because morphine is a metabolite of heroin (9). This could potentially result in underestimation of the number of deaths involving heroin and overestimation of the number of deaths involving morphine.

Determination of which drugs to report on the death certificate—Some medical certifiers focus on a single lethal drug rather than listing multiple drugs involved in the death, while others may list multiple drugs because they believe the drugs to be of equal lethality or that the interaction of all drugs mentioned is important. Some certifiers may not want to impose an order when listing the drugs that were present. Others have noted that space limitations in the software programs they use to complete electronic death registration limit their ability to include all drugs that contributed to the death.

These and other factors may contribute to the variation in the completeness and accuracy of the information on the death certificate about the specific drugs involved in the death. The literal text analysis is dependent on the quality of the information available. Therefore, the results presented in this report should be considered the minimum number or rate for that specific drug because there could be additional deaths in which the drug was involved, but the drug was not reported in the literal text.

Finally, it is possible that drugs rarely seen in drug overdose deaths were not included in the search term list used in this study, despite the multiple avenues taken to develop the list of search terms (7). This also could result in underestimation of the number of deaths involving a specific drug.

Conclusions

Literal text analysis can be used to extract key information from death certificates to improve national monitoring of drug overdose deaths. This report identifies the specific drugs most frequently mentioned in drug overdose deaths from 2011 through 2016, and shows that the most frequent drugs mentioned varied over time and by intent of death (i.e., unintentional drug overdose, suicide by drug overdose, and overdose death of undetermined intent). Results from the literal text analysis also confirm that many drug overdose deaths involve multiple drugs.

With slight modification, the methods used in this report can be used to identify deaths involving newly approved prescription drugs and new substances of abuse. Periodic updating of search terms and text search capabilities is essential for the ongoing surveillance and monitoring of emerging trends in drug overdose deaths using literal text analysis. In addition, this report highlights the critical importance of reporting the specific drugs involved in drug overdose deaths in the literal text on death certificates.

References

1. Hedegaard H, Warner M, Miniño AM. Drug overdose deaths in the United States, 1999–2016. NCHS Data Brief, no 294. Hyattsville, MD: National Center for Health Statistics. 2017. Available from: https://www.cdc.gov/nchs/data/databriefs/db294.pdf.
2. Seth P, Scholl L, Rudd RA, Bacon S. Overdose deaths involving opioids, cocaine, and psychostimulants—United States, 2015–2016. MMWR Morb Mortal Wkly Rep 67(12):349–58. 2018. Available from: https://www.cdc.gov/mmwr/volumes/67/wr/mm6712a1.htm.
3. Jones CM, Einstein EB, Compton WM. Changes in synthetic opioid involvement in drug overdose deaths in the United States, 2010–2016. JAMA 319(17):1819–21. 2018. Available from: https://jamanetwork.com/journals/jama/fullarticle/2679931?resultClick=1.
4. Rudd RA, Seth P, David F, Scholl L. Increases in drug and opioid-involved overdose deaths—United States, 2010–2015. MMWR Morb Mortal Wkly Rep 65(50–51):1445–52. 2016. Available from: https://www.cdc.gov/mmwr/volumes/65/wr/mm655051e1.htm.
5. Jones CM, Baldwin GT, Compton WM. Recent increases in cocaine-related overdose deaths and the role of opioids. Am J Public Health 107(3):430–2. 2017. Available from: http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.2016.303627.
6. World Health Organization. International statistical classification of diseases and related health problems, Tenth revision (ICD–10). 1st ed. Geneva, Switzerland. 1992.
7. Trinidad JP, Warner M, Bastian BA, Miniño AM, Hedegaard H. Using literal text from the death certificate to enhance mortality statistics: Characterizing drug involvement in deaths. National Vital Statistics Reports; vol 65 no 9. Hyattsville, MD: National Center for Health Statistics. 2016. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_09.pdf.
8. National Center for Health Statistics. Medical examiners' and coroners' handbook on death registration and fetal death reporting. Hyattsville, MD: National Center for Health Statistics. 2003. Available from: https://www.cdc.gov/nchs/data/misc/hb_me.pdf.
9. Davis GG. National Association of Medical Examiners position paper: Recommendations for the investigation, diagnosis, and certification of deaths related to opioid drugs. Acad Forensic Pathol 3(1):77–83. 2013.
10. National Center for Health Statistics. 2003 revision of the U.S. Standard Certificate of Death. Available from: https://www.cdc.gov/nchs/data/dvs/death11-03final-acc.pdf.
11. Warner M, Trinidad JP, Bastian BA, Miniño AM, Hedegaard H. Drugs most frequently involved in drug overdose deaths: United States, 2010–2014. National Vital Statistics Reports; vol 65 no 10. Hyattsville, MD: National Center for Health Statistics. 2016. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_10.pdf.
12. Murphy SL, Xu JQ, Kochanek KD, Curtin SC, Arias E. Deaths: Final data for 2015. National Vital Statistics Reports; vol 66 no 6. Hyattsville, MD: National Center for Health Statistics. 2017. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_06.pdf.
13. Chen LH, Hedegaard H, Warner M. Drug-poisoning deaths involving opioid analgesics: United States, 1999–2011. NCHS Data Brief, no 166. Hyattsville, MD: National Center for Health Statistics. 2014. Available from: https://www.cdc.gov/nchs/data/databriefs/db166.pdf.
14. National Cancer Institute. Joinpoint Regression Program (Version 4.6.0.0) [computer software]. 2018.
15. Ingram DD, Malec DJ, Makuc DM, Kruszon-Moran D, Gindi RM, Albert M, et al. National Center for Health Statistics Guidelines for Analysis of Trends. National Center for Health Statistics. Vital Health Stat 2(179). 2018. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_179.pdf.
16. National Research Council. Strengthening forensic science in the United States: A path forward. Washington, DC: National Academies Press. 2009.
17. Harruff RC, Couper FJ, Banta-Green CJ. Tracking the opioid drug overdose epidemic in King County, Washington using an improved methodology for certifying heroin related deaths. Acad Forensic Pathol 5(3):499–506. 2015.
18. Council of State and Territorial Epidemiologists. Recommendations and lessons learned for improved reporting of drug overdose deaths on death certificates. 2016. Available from: https://c.ymcdn.com/sites/www.cste.org/resource/resmgr/PDFs/PDFs2/4_25_2016_FINAL-Drug_Overdos.pdf.

Detailed Table

Table. Age-adjusted rates for drug overdose deaths involving selected opioids, benzodiazepines, and stimulants: United States, 2011–2016

Referent drug	2011	2012	2013	2014	2015	2016
Opioids						
Fentanyl	0.5	0.5	0.6	1.3	2.6	5.9
Heroin	1.5	2.0	2.7	3.5	4.3	5.1
Hydrocodone	1.0	1.0	1.0	1.0	0.9	1.0
Methadone	1.4	1.3	1.2	1.1	1.1	1.1
Morphine	1.0	1.1	1.2	1.2	1.3	1.5
Oxycodone	1.8	1.7	1.6	1.7	1.8	1.9
Benzodiazepines						
Alprazolam	1.3	1.2	1.2	1.3	1.5	2.0
Diazepam	0.5	0.5	0.5	0.5	0.5	0.6
Stimulants						
Cocaine	1.6	1.5	1.7	1.8	2.3	3.6
Methamphetamine	0.6	0.7	1.0	1.2	1.6	2.1

NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* (ICD–10) underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Deaths may involve other drugs in addition to the referent drug (i.e., the one listed). Deaths involving more than one drug (e.g., a death involving both heroin and cocaine) are counted in both totals. Caution should be used when comparing numbers across years. The reporting of at least one specific drug or drug class in the literal text, as identified using ICD–10 multiple cause-of-death codes T36–T50.8, improved from 75% of drug overdose deaths in 2011 to 85% of drug overdose deaths in 2016.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text, 2011–2016.

Technical Notes

Assessment of trends in drug-specific rates using an adjustment factor to account for improvements in reporting of specific drugs

The percentage of drug overdose deaths with an *International Classification of Diseases, Tenth Revision* multiple cause-of-death code T36–T50.8 indicates the reporting of specific drugs and drug classes in mortality data. During the study period 2011–2016, the percentage of drug overdose deaths with a multiple cause-of-death code T36–T50.8 increased from 74.9% of deaths in 2011 to 85.4% in 2016.

The improvement in reporting of specific drugs and drug classes during the study period could potentially influence the observed trends in drug overdose deaths for specific drugs (Figures 1–3). To assess the possible influence of improved reporting, an adjustment analysis was conducted. In this analysis, an adjustment factor was applied to each number and age-adjusted rate for drug overdose deaths involving the top 10 drugs involved in drug overdose deaths during 2011–2016. The adjustment factor was based on two assumptions: (1) the percentage of deaths with one or more drugs or drug classes specified in each year from 2011 through 2016 was the same and equal to the percentage in 2016 (85.4%), and (2) in each year, the distribution of deaths by specific drug was the same for deaths that identified one or more specific drugs, as for deaths that did not identify a specific drug. The adjustment factor was used to estimate the rate if the percentage of deaths with one or more drugs or drug classes specified had been uniform from 2011 through 2016.

The Technical Notes Table shows the crude rate, age-adjusted rate, and age-adjusted rate after application of the factor to adjust for improved reporting. The findings for the 2011–2016 trends in rates based on observed and adjusted numbers were, in general, the same (i.e., a statistically significant increase, decrease, or no change in the rate) for fentanyl, oxycodone, diazepam, cocaine, and methamphetamine. For heroin, the inflection point in 2014 was no longer found, resulting in a percent change in the rate of about 24% per year from 2011 through 2016. For morphine, the inflection point in 2015 was no longer found, resulting in a percent change in the rate of about 5% per year from 2011 through 2016. For methadone, rates decreased by about 12% per year from 2011 through 2014, and by about 3% from 2014 through 2016. For hydrocodone, there was a significant decline in the age-adjusted rates of about 4% per year from 2011 through 2016. For alprazolam, the inflection point in 2013 was no longer found, and as with the observed values, the increasing trend from 2011 through 2016 was not statistically significant.

Table. Crude rates, age-adjusted rates, and adjusted age-adjusted rates for drug overdose deaths involving selected drugs, 2011–2016

Drug	Crude rate						Age-adjusted rate						Trend 1		Trend 2		Adjusted age-adjusted rate						Trend 1		Trend 2	
	2011	2012	2013	2014	2015	2016	2011	2012	2013	2014	2015	2016	Years	Annual percent change	Years	Annual percent change	2011	2012	2013	2014	2015	2016	Years	Annual percent change	Years	Annual percent change
Opioids																										
Fentanyl	0.53	0.51	0.61	1.32	2.57	5.67	0.53	0.52	0.61	1.34	2.64	5.89	2011–2013	8.4	2013–2016	[1]113.3	0.60	0.59	0.67	1.42	2.71	5.89	2011–2013	6.1	2013–2016	[1]107.0
Heroin	1.47	1.96	2.66	3.41	4.14	4.94	1.47	1.99	2.70	3.50	4.27	5.10	2011–2014	[1]33.7	2014–2016	[1]20.0	1.67	2.23	2.96	3.71	4.39	5.10	2011–2016	[1]23.7	…	…
Hydrocodone	1.03	0.97	0.98	1.03	0.95	0.99	1.01	0.96	0.96	1.02	0.92	0.95	2011–2016	–1.0	…	…	1.15	1.08	1.06	1.08	0.94	0.95	2011–2016	[1]–3.7	…	…
Methadone	1.46	1.30	1.17	1.10	1.05	1.08	1.44	1.29	1.17	1.07	1.05	1.08	2011–2014	[1]–9.6	2014–2016	0.4	1.64	1.46	1.29	1.14	1.08	1.08	2011–2014	[1]–11.8	2014–2016	[1]–2.7
Morphine	1.06	1.12	1.19	1.26	1.31	1.55	1.03	1.10	1.16	1.24	1.29	1.53	2011–2015	[1]5.7	2015–2016	17.6	1.18	1.23	1.28	1.31	1.33	1.53	2011–2016	[1]4.6	…	…
Oxycodone	1.79	1.65	1.57	1.70	1.80	1.92	1.79	1.66	1.55	1.69	1.78	1.91	2011–2013	–6.5	2013–2016	7.0	2.04	1.87	1.70	1.79	1.83	1.91	2011–2013	–8.4	2013–2016	3.7
Benzodiazepines																										
Alprazolam	1.30	1.21	1.18	1.33	1.49	1.92	1.32	1.23	1.20	1.33	1.52	1.96	2011–2013	–6.7	2013–2016	18.2	1.51	1.38	1.31	1.41	1.56	1.96	2011–2016	5.8	…	…
Diazepam	0.54	0.50	0.51	0.55	0.56	0.63	0.54	0.49	0.50	0.55	0.55	0.63	2011–2016	3.5	…	…	0.61	0.55	0.55	0.58	0.56	0.63	2011–2016	0.7	…	…
Stimulants																										
Cocaine	1.63	1.52	1.68	1.85	2.28	3.50	1.62	1.52	1.67	1.85	2.29	3.55	2011–2016	[1]18.4	…	…	1.85	1.71	1.83	1.96	2.36	3.55	2011–2016	[1]15.2	…	…
Methamphetamine	0.61	0.72	1.01	1.18	1.58	2.09	0.62	0.73	1.02	1.18	1.60	2.12	2011–2016	[1]28.6	…	…	0.71	0.82	1.12	1.25	1.64	2.12	2011–2016	[1]25.1	…	…

… Category not applicable.
[1]Significant change in rate, $p < 0.05$.

NOTES: Drug overdose deaths are identified using *International Classification of Diseases, Tenth Revision* underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Deaths may involve other drugs in addition to the referent drug (i.e., the one listed). Deaths involving more than one drug (e.g., a death involving both heroin and cocaine) are counted in both rates. Trends in death rates were evaluated using the Joinpoint Regression Program set to identify a maximum of two joinpoints.

SOURCE: NCHS, National Vital Statistics System, Mortality files linked with death certificate literal text, 2011–2016.

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782–2064

For more NCHS NVSRs, visit:
https://www.cdc.gov/nchs/products/nvsr.htm.

National Vital Statistics Reports, Vol. 67, No. 9, December 12, 2018

Contents

Suggested citation

Hedegaard H, Bastian BA, Trinidad JP, Spencer M, Warner M. Drugs most frequently involved in drug overdose deaths: United States, 2011–2016. National Vital Statistics Reports; vol 67 no 9. Hyattsville, MD: National Center for Health Statistics. 2018.

For e-mail updates on NCHS publication releases, subscribe online at: https://www.cdc.gov/nchs/govdelivery.htm.
For questions or general information about NCHS: Tel: 1–800–CDC–INFO (1–800–232–4636) • TTY: 1–888–232–6348
Internet: https://www.cdc.gov/nchs • Online request form: https://www.cdc.gov/info
DHHS Publication No. 2019–1120 • CS298465

From: "Shimabukuro, Tom (CDC/DDID/NCEZID/DHQP)" <[redacted]>
To: "Sharan, Martha (CDC/DDID/NCEZID/DHQP)" <[redacted]>, "Su, John (CDC/DDID/NCEZID/DHQP)" <[redacted]>
Cc: "Nordlund, Kristen (CDC/OD/OADC)" <[redacted]>, "Thompson, PerStephanie (CDC/DDID/NCEZID/DHQP)" <[redacted]>, "Marquez, Paige L. (CDC/DDID/NCEZID/DHQP)" <[redacted]>, "Vaccine Safety (CDC)" <[redacted]>
Subject: RE: Reporter inquiry
Date: Thu, 23 Jun 2022 12:28:31 +0000
Importance: Normal
Inline-Images: image001.png

Okay, but the main reason we didn't do PRR is b/c FDA EB data mining is the 'gold standard' for disproportionality analysis so we had a better and more efficient way of doing disproportionality analysis at a time when we were occupied trying to monitor an onslaught of reports. There really isn't a reason for CDC to do PRR if FDA is conducting EB data mining b/c it's basically redundant. Now that we are further along in the pandemic and we better understand some of the limitations of FDA's EB data mining for COVID-19 vaccines we are doing some exploratory work with PRR, but it's still not a major component of our monitoring.

I find some of the statements in the response a bit problematic.

I disagree with this statement: Various technical limitations, including insufficient data, precluded PRR analyses during that time in the vaccination campaign. PRR is a simple (maybe overly simplistic) mathematical calculation. There are no technical limitations to doing PRR, it's easy, and there are plenty of data in VAERS to do PRR whenever we want to and on whichever vaccines we choose. The issue is whether it's a good idea for CDC to do PRR when FDA is doing EB data mining (see below).

I somewhat disagree with this statement: PRR analyses of COVID-19 vaccines early in the vaccination campaign were inappropriate and thus not conducted. It's only inappropriate in the sense that EB data mining is a better test of disproportionality b/c PRR tends to generate all kinds of spurious findings. FYSA, the Uppsala monitoring center in Europe, which is affiliated with WHO, uses PRR and ROR as its primary disproportionality analysis. Also, the statement seems to indirectly imply that it might be appropriate to use PRR now (vs. early), but I would questions whether PRR is appropriate even now. The test still has all its original limitations.

I think the main message should be that FDA's EB data mining supersedes PRR in importance and from the perspective of generating informative data. CDC surveillance focus early on was descriptive analysis of large volumes of data and focusing on adverse events of special interest (e.g., anaphylaxis).

From: Sharan, Martha (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Thursday, June 23, 2022 8:01 AM
To: Shimabukuro, Tom (CDC/DDID/NCEZID/DHQP) <[redacted]>; Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>
Cc: Nordlund, Kristen (CDC/OD/OADC) <[redacted]>; Thompson, PerStephanie (CDC/DDID/NCEZID/DHQP) <[redacted]>; Marquez, Paige L. (CDC/DDID/NCEZID/DHQP) <[redacted]>; Vaccine Safety (CDC) <[redacted]>
Subject: RE: Reporter inquiry

Hi Tom and John:
I think this reporter is going to need adequate information from CDC to write her piece countering the CHD.
We can point her to ACIP presentations ands studies, but I don't think she's going to be able to build a piece on her own from all that material, especially if she has not been following it.

So, I think the information John provided will be much more helpful. We may need to edit it down just a bit, but I can work on that. I can also reach out to the reporter to get a reading on how much detail she needs.

There have been 2 inquiries about this: AP and Washington Examiner.

Thanks,
Martha

Martha Sharan
Public Affairs
CDC/Division of Healthcare Quality Promotion
[redacted]

From: Shimabukuro, Tom (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Wednesday, June 22, 2022 5:42 PM
To: Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>; Sharan, Martha (CDC/DDID/NCEZID/DHQP) <[redacted]>
Cc: Nordlund, Kristen (CDC/OD/OADC) <[redacted]>; Thompson, PerStephanie (CDC/DDID/NCEZID/DHQP) <[redacted]>; Marquez, Paige L. (CDC/DDID/NCEZID/DHQP) <[redacted]>; Vaccine Safety (CDC) <[redacted]>
Subject: RE: Reporter inquiry

Can we just simply point the reporter to the publications website and the ACIP presentations website to demonstrate the monitoring and signal detection/signal assessment activities that have been happening since December 2020. We could also say that PRR is a form of disproportionality analysis and FDA empirical Bayesian data mining is the primary disproportionality analysis used in VAERS. CDC selectively uses PRR as a supplement or complement to FDA's EB data mining.

From: Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Wednesday, June 22, 2022 5:20 PM
To: Sharan, Martha (CDC/DDID/NCEZID/DHQP) <[redacted]>
Cc: Nordlund, Kristen (CDC/OD/OADC) <[redacted]>; Thompson, PerStephanie (CDC/DDID/NCEZID/DHQP) <[redacted]>; Marquez, Paige L. (CDC/DDID/NCEZID/DHQP) <[redacted]>; Vaccine Safety (CDC) <[redacted]>; Shimabukuro, Tom (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: RE: Reporter inquiry

Hi folks,

Adding PerStephanie to this email, as she's more versed in FOIA matters, and can advise if any details I provide are privileged (ie, shouldn't be released in response to this inquiry). Also adding Paige, who can provide more technical details if I get my facts crossed, and the Vaccine Safety mailbox for tracking purposes. And Tom, for his awareness (it's AP, after all).

The FOIA below mentioned specified analyses of proportionality reporting ratios (PRRs) during February 1, 2021, through Sept. 30, 2021. You might recall that emergency use authorization (EUA) for the mRNA vaccines was granted in December 2020, and EUA for Janssen's vaccine was in March 2021. While there was a considerable reporting volume to VAERS during this time period, preliminary reports of adverse events of special interest (AESIs) were more limited. Thus, during the time period specified, we were early in the vaccination campaign, and insufficient data had accrued during that time for PRR analysis.

Also, selection of appropriate comparator vaccines was a challenge: ideally, PRRs are conducted between vaccines of similar type or technology (e.g., comparing live virus vaccines like MMR and varicella (Varivax), or conjugated vaccines like Prevnar (pneumococcal conjugate vaccine) and the meningococcal conjugate vaccines). No such comparator existed for the mRNA vaccines. We've performed some draft analyses against the influenza vaccines (with no surprising

results – a lot of the tagged adverse events (AEs) were typical of the mRNA vaccines); these draft analyses were performed after the period specified in the FOIA – and again, they were draft, so we wouldn't want to consider them formal analyses.

Lastly, interpretation of PRRs is tricky – the usual threshold we use is 2.0, indicating that a given AE was reported after one vaccine twice as often as after the comparator vaccine(s). As you can imagine, quite a few AEs yield a PRR of 2.0 or greater. This threshold indicates no statistical significance; it's an arbitrary cut point, like selecting a p value of 0.05 (which corresponds to a given outcome in 5 of 100 occurrences, such as 5 heads out of 100 coin flips). In sum, PRRs are noisy and challenging to interpret. To call a PRR of 2 or greater a "safety signal" (as the author of the CHD article does) is a gross overstatement.

CDC (and VAERS specifically) *was* engaged in safety signal analysis during the period specified in the FOIA – but what the CHD author fails to grasp is that "signal detection" is a sum of both quantitative and qualitative analysis. Case in point: thrombosis with thrombocytopenia syndrome (TTS) after Janssen's vaccine was identified during the FOIA period. VAERS contacted the Advisory Committee on Immunization Practices (ACIP) within 3 weeks of initiating use of the vaccine when 6 cases of cerebral venous sinus thrombosis (CVST) with low platelets had been identified. With only 6 cases of CVST, no thromboembolic symptom would flag via PRR, or even Empirical Bayesian data mining, analyses. However, a similar syndrome had been observed after AstraZeneca's vaccine in Europe – and, like Janssen's vaccine, AstraZeneca's vaccine was based on an adenoviral vector. In this case, VAERS identified this safety signal not through quantitative techniques per se, but via pattern recognition.

With the above background, I might suggest the following response:

"The author of the Children's Health Defense article mischaracterized safety signal analysis. In brief, Proportionality Reporting Ratios (PRRs) can be helpful in identifying potential vaccine safety concerns, or "safety signals", but PRRs are a single tool and do not by themselves indicate such safety signals.

PRRs compare the counts of reports of a given adverse event (AE) after one vaccine to after another vaccine (or vaccines). For example, a PRR of 2.0 indicates that a given AE was reported twice as often after one vaccine as after another vaccine(s). A known limitation of the Vaccine Adverse Event Reporting System (VAERS) is that reporting to VAERS can be influenced by numerous factors, including increased public attention or awareness of a given AE. Thus, a PRR by itself does not constitute a safety signal: there can be numerous explanations for why a PRR might be elevated for a given vaccine. PRRs can be helpful tools, but they do not indicate potential safety concerns with a vaccine on their own.

Further, the FOIA requested PRR analyses from early in the COVID-19 vaccination campaign. Various technical limitations, including insufficient data, precluded PRR analyses during that time in the vaccination campaign.

More importantly, CDC has been engaged in safety signal surveillance since COVID-19 vaccines have been in use. During the first month of their availability, data on anaphylaxis after mRNA COVID-19 vaccines were published (including in highly visible journals, like the Journal of the American Medical Association (JAMA)), indicating an observed incidence comparable to after other vaccines. VAERS detected what would become known as thrombosis with thrombocytopenia syndrome (TTS) after Janssen's vaccine, leading to a pause in the use of the vaccine mere weeks after its use was initiated. VAERS reviewed reports of myocarditis after mRNA COVID-19 vaccines during Summer 2021, providing a highly thorough characterization of such reports. These examples indicate that the vaccine safety surveillance systems in use by CDC and FDA identify potential vaccine safety concerns in a timely and effective manner.

In sum, PRR analyses of COVID-19 vaccines early in the vaccination campaign were inappropriate and thus not conducted. However, CDC and FDA have been actively engaged in vaccine safety surveillance ever since COVID-19 vaccines have been in use."

Please let me know what you think, and if you have any comments, feedback, or any questions. Thanks!

- John

From: Sharan, Martha (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Wednesday, June 22, 2022 12:51 PM
To: Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>

Cc: Nordlund, Kristen (CDC/OD/OADC) <[redacted]>
Subject: FW: Reporter inquiry

Hi John… wanted to check with you on this request from Associated Press. Can you offer a response to this one? I'm cc'ing Kristen, since the request came through her and we don't have a standard response to what the reporter is asking.
Thanks,
Martha

Martha Sharan
Public Affairs
CDC/Division of Healthcare Quality Promotion
[redacted]

Hi,

I'm a fact-checking reporter at The Associated Press. I'm looking into a new post by Children's Health Defense that is being shared on social media, alleging that a FOIA response from CDC shows the agency "admitted it never analyzed the Vaccine Adverse Event Reporting System for safety signals for COVID-19 vaccines":
https://childrenshealthdefense.org/defender/cdc-vaers-covid-vaccine-safety/

The claim appears to center on a request for "proportional reporting ratio" calculations – and a CDC response saying "no PRRs were conducted by CDC."

Could CDC explain what, exactly, was requested in this FOIA request (#22-01479-FOIA)? Has the agency performed any PRRs in relation to the COVID-19 vaccines? Has the agency analyzed VAERS data pertaining to the COVID-19 vaccines in other ways?

Thanks,
Angelo

--

Angelo Fichera
Reporter, News Verification
The Associated Press
[redacted]

From: "Menschik, David" <[redacted]>
To: "Su, John (CDC)" <[redacted]>
Subject: data mining limitations
Date: Wed, 22 Sep 2021 16:33:23 +0000
Importance: Normal
Attachments: mRNA_6mo_safety_review-update98forOS_091521.docx

---

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…)

Reactogenicity and Adverse Events during the First Six Months of mRNA COVID-19 Vaccination in the United States: A Prospective Observational Study of Reports to Vaccine Adverse Events Reporting System (VAERS) and v-safe

Hannah G. Rosenblum, MD[1,2]; Julianne M. Gee, MPH[1]; Ruiling Liu, PhD[1]; Paige L. Marquez, MSPH[1]; Bicheng Zhang, MS[1]; Penelope Strid, MPH[1], Winston E. Abara, MD[1]; Michael M. McNeil, MD, MPH[1]; Lauri E. Markowitz, MD[1]; Tanya R. Myers, PhD[1]; Anne M. Hause, PhD, MSPH[1]; John R. Su, MD, PhD[1]; Bethany Baer, MD[3]; David Menschik, MD, MPH[3]; Tom T. Shimabukuro, MD, MPH, MBA[1], David K. Shay, MD, MPH[1]

[1]CDC COVID-19 Response Team, Centers for Disease Control and Prevention, Atlanta, Georgia

[2]Epidemic Intelligence Service, Centers for Disease Control and Prevention, Atlanta, Georgia

[3]Food and Drug Administration, Silver Spring, Maryland

Corresponding author: Julianne Gee, [redacted]

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 (CDC) or the Food and Drug Administration (FDA)

Acknowledgements

We wish to acknowledge the following contributors: CDC: Amelia Jazwa, Tara Johnson, Charles Licata, Stacey Martin; FDA: Jane Baumblatt, Deborah Thompson, Kerry Welsh, Narayan Nair, Kosal Nguon (Commonwealth Informatics); v-safe participants; Oracle v-safe development team. Mention of a product or company name is for identification purposes only and does not constitute endorsement by the CDC or the FDA.

Target journal: Lancet ID

Manuscript word count: ***/3500

Abstract (word count: 233/250)

Background: In December 2020, two mRNA-based COVID-19 vaccines were authorized for use in the United States. Vaccine safety was monitored using Vaccine Adverse Event Reporting System (VAERS), a national passive surveillance system, and v-safe, an active surveillance system.

Methods: VAERS and v-safe data from December 14, 2020—June 14, 2021 were analyzed. Empirical Bayesian data mining was used to identify disproportional reporting of events by vaccine in VAERS. Proportions of v-safe participants reporting local and systemic reactions or health impacts the week following first and second vaccine doses were determined.

Findings: During the analytic period, 298,792,852 total doses of mRNA vaccines were administered in the United States. VAERS received and processed 340,522 reports; 92·1% were classified as non-serious; 6·6%: serious, non-death; and 1·3% as death. Over half of 7,914,583 v-safe participants self-reported local and systemic reactogenicity, more frequently after dose 2. Injection-site pain, fatigue, and headache were most commonly reported during days 0–7 following vaccination. Reactogenicity was reported most frequently one day after vaccination and rapidly declined; most reported reactions were mild. More reports of being unable to work or do normal activities occurred after dose 2 (32·1%) than dose 1 (11·9%); <1% of participants reported seeking medical care after vaccination.

Interpretation: Safety data from >298 million doses of mRNA COVID-19 vaccine administered in the first 6 months of the U.S. vaccination program show the majority of reported adverse events were mild and short in duration.

Funding: No external sources of funding were used. CDC received nonfinancial technical support to develop and maintain the v-safe infrastructure from Oracle.

Commented [RH(1]: Note all death results/interpretation has been removed from abstract
Rosenblum, Hannah (CDC)
2021-09-08 10:30:00

Introduction

In December 2020, two messenger RNA (mRNA) coronavirus disease 2019 (COVID-19) vaccines (BNT162b2 developed by Pfizer-BioNTech and mRNA-1273 developed by Moderna) were granted Emergency Use Authorization (EUA) by the U.S. Food and Drug Administration (FDA) as 2-dose series and recommended for use by the Advisory Committee on Immunization Practices (ACIP).[1,2] The mRNA vaccine platform uses lipid nanoparticles as a carrier system for the mRNA which encodes the SARS-COV-2-spike protein. In clinical trials, both mRNA COVID-19 vaccines had acceptable safety profiles.[3,4] Reactogenicity (i.e., local and systemic reactions) was observed after receipt of vaccine in clinical trials of both vaccines; the most frequently reported symptoms included injection site pain, fatigue, and headache. Reactogenicity was more frequently reported following dose 2, and more common among participants aged <65 years.[3-5]

Commented [BR(2]: Sounds awkward as written. Actually, the lipid nanoparticles envelop the mRNA, which encodes the genetic sequence information for the viral SARS-CoV-2 spike protein. Another way to put it "The messenger RNA vaccine platform uses lipid nanoparticles as a carrier system for the mRNA which encodes the SARS-CoV-2 spike protein".
Office of Science
2021-08-11 07:50:00

Commented [RH(3R2]: thanks- has been edited
Rosenblum, Hannah (CDC)
2021-08-17 14:32:00

Post-authorization safety monitoring is necessary to better understand the safety profiles of mRNA-based COVID-19 vaccines in larger and more heterogeneous populations.[6] Phased administration of COVID-19 vaccines in the United States began with healthcare workers and residents of long-term care facilities and expanded to the general population by spring 2021; however, implementation plans varied by state.[7] The Vaccine Adverse Event Reporting System (VAERS), a spontaneous reporting (i.e., passive surveillance) system,[8] and v-safe,[9] a new active monitoring system, were the primary safety data sources used in initial reports of adverse events following administration of COVID-19 vaccines in the United States vaccination program.[10-11] Since the inception of the program, regular vaccine safety updates from these systems have been provided through websites, publications, and presentations to advisory committees.[10-14] Here, we review VAERS and v-safe safety data during the first six months of the U.S. vaccination program, when over 298 million doses of mRNA COVID-19 vaccines were administered.

Methods

VAERS

VAERS is an established, national spontaneous reporting system that serves as an early warning system for detecting potential safety problems for vaccines authorized or licensed in the United States.[8] Co-administered by Centers for Disease Control and Prevention (CDC) and FDA, VAERS accepts reports from health care providers, manufacturers, and the public. VAERS reports include information about the vaccinated person, type of vaccine administered, and the adverse event (AE) experienced. For this analysis, VAERS reports submitted and processed by June 14, 2021 were included.[15] Processed reports were those checked for data quality, de-duplicated, and coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology.[8] Each VAERS report may be assigned more than one MedDRA Preferred Term (PT); PTs do not necessarily indicate a medically confirmed diagnosis, and include signs and symptoms of illness and results of diagnostic tests.

Based on the Code of Federal Regulations,[16] VAERS reports were classified as serious if any of the following were documented: hospitalization, prolongation of existing hospitalization, permanent disability, life-threatening illness, congenital anomaly or birth defect, or death. Adverse events of special interest (AESI)[17] were selected for enhanced COVID-19 vaccine safety monitoring based on biologic plausibility, previous vaccine safety experience, and theoretical concerns related to COVID-19.[17] Death certificates and autopsy reports were requested for death reports. CDC physicians reviewed VAERS reports and available death certificates for each decedent to form an impression about cause of death. Causes of death were further categorized into the following groups, using the National Center for Health Statistics the 15 most common major *International Classification of Disease, Tenth Revision (ICD-10*) diagnostic categories reported on U.S. death certificates[18]: COVID-19 disease; other (i.e., diagnosis did not belong in one of the other pre-specified categories); or unknown/unclear if a likely cause could not be determined.

V-safe

V-safe is a voluntary smartphone-based system that uses text messaging and secure web-based surveys to actively monitor vaccine safety, and has been specifically designed to gather information about COVID-19 vaccine AEs, particularly for common local injection site and systemic reactions.[19] V-safe participants receive text messages that link to web-based health check-ins and respond to questions in surveys following vaccination, initially daily (days 0–7), then weekly (days 14–42) and lastly at 3, 6 and 12 months post vaccination. The system resets to the initial survey frequency after receipt of dose 2. We analyzed survey reports from days 0–7 for reactogenicity, severity[9] (mild, moderate, severe), and health impact (i.e., unable to perform normal daily activities, unable to work, and/or received care from a medical professional). Participants who reported receiving medical care were contacted by v-safe staff and VAERS reports were completed if clinically indicated.

Data analyses

We conducted descriptive analyses of available VAERS and v-safe data from December 14, 2020–June 14, 2021 following first and second doses of BNT162b2 and mRNA-1273 vaccines. For VAERS, bivariate analyses included sex, age groups, race/ethnicity, serious AEs, time from vaccination to reported death (i.e., onset interval) for death reports, cause of death for death reports, and vaccine type/manufacturer administered. Unadjusted, crude reporting rates to VAERS were calculated for AEs using the total number of doses of mRNA vaccine administered during the six-month period. COVID-19 vaccine administration data were provided through CDC's COVID-19 Data Tracker.[20]

Empirical Bayesian (EB) data mining was used to detect disproportional reporting of post-vaccine outcomes by vaccine received among all VAERS serious and non-serious reports received by June 14, 2021.[21] This statistical method calculates observed to expected PT pairings by comparing a specific vaccine-PT pair to all vaccine-PT pairs in VAERS, adjusting for age, sex, and year of vaccination.[22]

Commented [RH(4]: I removed crude everywhere else, but I think good to leave here- what do others think?
Rosenblum, Hannah (CDC)
2021-09-08 11:49:00

Commented [BR(5]: Required:
This is unclear to the general reader. Please clarify "disproportionately" to what.
Office of Science
2021-08-11 12:05:00

Commented [RH(6R5]: Thank you- this is a typo! It should have read "disproportionately" but rather "disproportionality" or as I have modified it to "disproportional reporting". Thank you!!
Rosenblum, Hannah (CDC)
2021-08-11 17:03:00

Commented [BR(7]: Required: It is unclear how the rate of expected PT pairings were determined. Please explain.
Office of Science
2021-08-11 09:45:00

Commented [RH(8R7]: Have expanded the phrase and re-checked the references as well as added another reference suggested by FDA. Thank you!

Rosenblum, Hannah (CDC)
2021-08-11 17:04:00

These ratios are ranked by the lower 5% bound of the EB geometric mean confidence interval (EB05) and a standard alert threshold of EB05 >2 was used. An EB05 >2 represents a high degree of confidence that a vaccine-PT pair was reported at least twice as frequently as expected. In addition to overall ratios, ratios were calculated for age group, sex, serious reports, and death reports.

V-safe participants who responded to at least one health check-in survey during day 0–7 after vaccination were included in analyses. Descriptive statistics were calculated for participants characteristics (sex, age, race/ethnicity), reaction (type and severity) and health impact by manufacturer, dose number, and number of days following vaccination.

SAS software, version 9.4 (SAS Institute; Cary, NC, USA) was used for analyses. Both VAERS and v-safe conduct surveillance as a public health function and are exempt from institutional review board review. Activities were reviewed by the CDC and were conducted in accordance with applicable federal law and CDC policy (See: 45 C.F.R. part 46.102(l)(2), 21 C.F.R. part 56; 42 U.S.C. §241(d); 5 U.S.C. §552a; 44 U.S.C. §3501 et seq.).

Results

During December 14, 2020–June 14, 2021, a total of 298,792,852 doses of mRNA COVID-19 vaccines were administered in the United States: 167,177,332 were BNT162b2 and 131,639,515 were mRNA-1273 (Supplemental Table 1). A greater proportion of vaccines were administered to females (53·2%) compared with males (45·8%). The median age at vaccination was 50 years (inter-quartile range [IQR]: 33–65) for BNT162b2 and 56 years (IQR: 39–68) for mRNA-1273, respectively. Non-Hispanic White persons accounted for 38·4% of vaccine recipients; however, race/ethnicity was unknown for 34·9% of all vaccine recipients.

VAERS

During the analytic period, VAERS received and processed a total of 340,522 reports: 164,669 were following BNT162b2 and 175,816 were following mRNA-1273 vaccine administration (Table 1). Of these reports, 92·1% were classified as non-serious, 6·6% were serious, not resulting in a death (non-death), and 1·3% were deaths. Seventy-two percent of reports were among females, and 45·3% of reports were among vaccine recipients aged 18–49 years; median age was 50 years (IQR: 36–64). Fifty percent of those reporting race/ethnicity identified as non-Hispanic White; for 22·1%, race/ethnicity was unknown. The most common MedDRA PTs among non-serious reports were headache (20·4%), fatigue (16·6%), pyrexia (16·3%), chills (15·7%), and pain (15·2%). The most common MedDRA PTs among serious reports were dyspnea (15·4%), death (14·1%), pyrexia (11·0%), fatigue (9·7%), and headache (9·5%). The reporting rate to VAERS was 1,049 non-serious reports per million doses, and 90 serious reports per million doses (Table 2). Among the pre-specified AESIs, reporting rates ranged from 0·1 narcolepsy reports per million doses administered to 32 COVID-19 disease reports per million doses administered.

There were 4,496 reports of death in VAERS (Table 3). After review, 24 reports were excluded because of miscoding of death or duplicate reporting. Of the 4,472 reports of deaths analyzed, 2,087 (46·7%) were reported following BNT162b2 and 2,385 (53·3%) following mRNA-1273. Females accounted for 42·6% of reported deaths; the median age of decedents was 76 years (IQR: 66–86). More than 80% of deaths were reported among individuals aged 60 years or older (reporting rate of death per million doses administered by age group: 60–69 years, 2·6; 70–79 years, 3·7; 80–89 years, 3·8; ≥90 years, 2·1).18·3% of decedents were identified as long-term care facility residents. Death certificates or autopsy reports were available for clinical review for 808 (18·1%) reports of deaths analyzed (Table 4). Among these 808 reports, causes of death were most commonly diseases of the heart (46·5%) and COVID-19 disease (12·6%). Causes of death among reports with death certificate or autopsy available are shown by age in Figure 1 and Supplemental Table 2. Among the 3,664 reports of death without a death certificate or

Commented [RH(9]: These figures/tables are new
Rosenblum, Hannah (CDC)
2021-09-08 11:54:00

autopsy, causes of death were most commonly unknown/unclear (54·1%), diseases of the heart (17·0%), and COVID-19 disease (8·7%). Supplemental Table 3 displays specific impressions within each category of cause of death, for all deaths, and for those with death certificate or autopsy. Time interval to death following vaccination was available for 4,119 reports (92·1%) and the median time interval was 10·0 days (range: 0—161 days) after vaccination. The greatest number of reports of deaths occurred on day 1 (10·5%) and day 2 (7·0%) following vaccination (Supplemental Figure 1).

Commented [RH(10]: Not sure if sentence needed
Rosenblum, Hannah (CDC)
2021-09-08 11:56:00

EB Data Mining

No adverse health outcome alerts were identified in EB data mining. However, five mRNA COVID-19 alerts with disproportionality (EB05>2) were identified during the surveillance period. For BNT162b2 vaccine, 'product preparation issue' alerted among all reports (EB05: 2·09; N=757), and among adults ≥65 years (EB05: 2·10; N=205), females (EB05: 2·03; N=394), and males (EB05: 2·01; N=350). Two terms for BNT162b2 vaccine alerted in adults ≥65 years: 'investigation' (EB05: 2.06; N=163) and 'weight' (EB05: 2·01; N=139). For mRNA-1273 vaccine, two terms alerted among all reports: 'poor quality product administered' (EB05: 2·43; N=1,506), and 'product temperature excursion issue' (EB05: 2·17; N=720).

v-safe

During the analytic period, 7,914,583 mRNA COVID-19 vaccine recipients enrolled in v-safe and completed at least one post-vaccination health survey during days 0–7 after vaccination (Table 5). The median age of v-safe participants was 50 years (IQR: 36–63), 62·9% were female, and 59·4% identified as non-Hispanic White. A total of 6,775,515 participants completed at least one survey during day 0–7 after dose 1 (3,455,778 following BNT162b2 and 3,319,737 following mRNA-1273). Of these participants, 68·6% reported a local injection site reaction and 52·7% reported a systemic reaction. Of the 5,674,420 participants who completed a survey after dose 2, a greater percentage reported an injection site

reaction (71·7%) and/or a systemic reaction (70·8%) (Table 6). Local injection site reactions were reported more frequently after mRNA-1273 (dose 1: 73·3%; dose 2: 78·4%) than after BNT162b2 (dose 1: 64·0%; dose 2: 65·3%). A similar pattern was found for systemic reactions after mRNA-1273 (dose 1: 54·3%; dose 2: 75·8%) versus BNT162b2 (dose 1: 51·3%; dose 2: 66·1%). The most frequently reported events after dose 1 of either mRNA vaccine included injection site pain (66·2%), fatigue (33·9%), and headache (27·0%); these reactions were also more frequent after dose 2: injection site pain (68·6%), fatigue (55·7%), headache (46·2%). Differences in proportions of reactogenicity by dose number were similar after stratifying by age group (<65 vs. ≥65 years) and sex. More reactogenicity was reported among younger participants aged <65 years and by females. (Supplemental Table 4).

Proportions of reported severity of reactions by manufacturer, dose number, and day since vaccination are shown in Figure 2. The majority of reported symptoms were mild. Participants reported moderate and severe reactogenicity most commonly on day 1 after dose 2 of either vaccine. The proportion of participants who reported symptoms was greatest on day 1 and then decreased on subsequent days. The highest proportion of participants reported severe symptoms on day 1 following dose 2 of mRNA-1273 (Supplemental Table 6). On all other days, proportions of participants reporting severe symptoms did not exceed 3.0% for any individual symptom (Supplemental Tables 5 and 6).

Reported health impact was greater following dose 2 of either mRNA vaccine (32·1%) compared with dose 1 (11·9%) and after mRNA-1273 of either dose compared with BNT162b2 (Table 6). After dose 1 of either mRNA vaccine, 9·7% of participants were unable to do normal activities and 4·5% were unable to work. After dose 2 of BNT162b2, 20·5% were unable to do normal activities, and 12·3% were unable to work. After dose 2 of mRNA-1273, 32·8% were unable to do normal activities, and 20·0% were unable to work. Less than 1·0% reported receiving medical care after receiving either dose from either manufacturer. Fewer participants reported an emergency room visit (dose 1: 0·1%; dose 2: 0·2%) or hospitalization (dose 1: 0·03%; dose 2: 0·04%).

When stratified by sex, females reported a health impact more frequently than males, peaking on day 1 after vaccination (Supplemental Figure 2). Following dose 2 of mRNA-1273 vaccine, 41·4% of females reported in the day 1 survey an inability to perform normal activities, and 23·5% an inability to work. Among males receiving dose 2 of mRNA-1273 on the day 1 survey, 25·6% were unable to perform normal activity and 16·9% were unable to work (Supplemental Table 7).

Discussion

During the first six months of the U.S. COVID-19 vaccination program, over 298 million doses of mRNA vaccines were administered. COVID-19 vaccine safety in the United States has been monitored with well-established systems, including the Vaccine Safety Datalink[23] and VAERS, and a system developed specifically for COVID-19 vaccine safety monitoring, known as v-safe. The post-authorization safety profile for mRNA COVID-19 vaccines after six months of use in the United States is largely consistent with data presented in the pre-authorization clinical trials.[3,4] Data from U.S. safety monitoring systems have been presented regularly to ACIP's COVID-19 Vaccine Safety Technical Subgroup (VaST) work group[24] and at public ACIP meetings.[25] Data have been presented concerning cases of clinically serious AEs, including anaphylaxis,[13] thrombosis with thrombocytopenia syndrome (TTS),[26] myocarditis,[27] and Guillain-Barré Syndrome (GBS)[28] reported following receipt of COVID-19 vaccines. ACIP has assessed the benefit-risk balance of each of the currently authorized U.S. COVID-19 vaccines; these evaluations have not prompted any changes in U.S. COVID-19 immunization recommendations.[13,27,28]

Our main findings are similar to those obtained from diary-based reporting in pre-authorization clinical trials and early post-authorization reports – data from all reports demonstrate substantial local and systemic reactogenicity.[3-5,10,11] In both VAERS and v-safe, local injection site and systemic reactions were commonly reported, and in v-safe, transient reactions were reported more frequently following mRNA-1273 compared with BNT162b2, and more frequently following dose 2. Overall, females and individuals

aged <65 years reported AEs and reactions more frequently. These findings are similar to those from a large-scale study about reactogenicity conducted in the United Kingdom.[29] Host characteristics known to influence reactogenicity, including age, sex, and the presence of underlying medical conditions, might be associated with this pattern of findings.[30] Females have more vigorous antibody responses[31] to certain vaccines and also tend to report more severe local and systemic reactions to influenza vaccine.[32] Females may also be more likely than males to respond to surveys[33,34] and we hypothesize that younger individuals may be more comfortable with smartphone-based surveys and more likely to respond to survey questions.[35,36]

The impact of vaccination on daily life activities was most frequently reported on the first day after vaccination. Reports about the health impact measures used in v-safe, while self-assessed and subjective, correlate with reports about reactogenicity patterns: more health impact was reported by females than males, by participants aged <65 years compared with older participants, by persons receiving dose 2 compared with dose 1, and by those who received mRNA-1273 versus BNT162b2. Reports of seeking medical care (including telehealth and urgent care) after receipt of either dose of mRNA vaccine were rare, suggesting that reactogenicity was transient and manageable at home. Among those who did report seeking medical care, only a small proportion visited an emergency department or were hospitalized. Reactogenicity and its associated health effects, even if transient, may deter some persons from seeking vaccination. An April 2021 survey conducted by the Kaiser Family Foundation found that nearly half (48%) of unvaccinated adults aged <50 years expressed concern about missing work due to vaccine side effects; this concern was reported by 55% of unvaccinated Black adults and 64% of unvaccinated Hispanic adults.[37] Employees who are provided time off may be more likely to get vaccinated, even after controlling for other demographic factors that might influence vaccine uptake.[38] These data suggest that employee work policies that accommodate days off for vaccination and recovery from side effects may increase vaccination coverage.[39]

Increased public awareness, widespread promotion of VAERS, and outreach and education to healthcare providers about COVID-19 EUA AE reporting requirements are likely all contributing factors to the high volume of VAERS reports following mRNA COVID-19 vaccines as compared to established adult vaccinations.[8] For mRNA COVID-19 vaccines in this six-month period, VAERS has processed and received more than six times the number of average reports per year (typically 50,000 reports are received per year for all vaccines in all age groups). For example, the number of reports of death in VAERS following mRNA vaccine in this period exceeds the number of deaths reported to VAERS for all other vaccines in a summary report from 1997–2013 by eight times.[40] The concentrated reporting of deaths on days 1 and 2 following vaccination may represent reporting bias, as the likelihood to report a serious AE may increase when it occurs in close temporal proximity to vaccination.

Comparing deaths reported to VAERS following mRNA vaccination by cause to national mortality data[41] is challenging, as more common causes of deaths in younger individuals (for example, accidents or suicide) may be less likely to be reported to VAERS. The overrepresentation of diseases of the heart as cause of death in general may be driven by non-specific causes of death on death certificates such as cardiac arrest or cardiopulmonary arrest, which are terminal events, but might be chosen if no immediate explanation is available. Additional studies are needed to characterize deaths in VAERS …???

During the 6-month period we analyzed, patterns of reports to VAERS are similar to other vaccines that are routinely administered to adults and the majority of reported events were non-serious.[42,43,44,45] None of the EB data mining alerts suggested an unexpected vaccine safety problem. Serious AEs have been detected following receipt of COVID-19 vaccines during U.S. safety monitoring and reviewed in detail.[26-28] Early reports of anaphylaxis prompted recommendations about specific clinical management including screening and recommendation of a post-observation period following vaccination.[46] After myocarditis was observed following mRNA vaccination,[47,48] particularly in males aged <30 years, CDC issued clinical guidance and management recommendations,[49] and presented a benefit-risk assessment to ACIP.[27] The risk of TTS[26] and GBS[28] is elevated following receipt of Janssen COVID-19 vaccine (Ad26.COV2.S) and have not been associated with mRNA COVID-19 vaccines to date.

Commented [RH(11]: include total number of vaers reports yearly to highlight overall magnitude through 6 months?
Rosenblum, Hannah (CDC)
2021-09-07 14:51:00

Commented [S(12R11]: Yes, i think that's a good idea
Shay, David (CDC/DDID/NCIRD/ID)
2021-09-08 08:02:00

Commented [RH(13R11]: Tried to use CDC wonder for this and did some literature search, but unable to find a readily available resource to cite- wondering if comparing to the # of deaths is enough to make this point

Rosenblum, Hannah (CDC)
2021-09-08 10:37:00

Commented [RH(14R11]: Received this data from Paige- not sure If I can cite Wonder- asking her…
Rosenblum, Hannah (CDC)

Commented [BR(15]: Required:
Rates per million doses administered is not the same as rates per million persons vaccinated. Rates per doses can

Commented [RH(16R15]: Thanks so much – we've removed the text about rate comparisons for the reasons you outlined

Commented [RH(17]: Framing around deaths evaluated after pediatric vaccination-
Vaers hasn't been used to evaluate deaths following

Commented [BR(18]: Required:
This reporting may also reflect a true event. This hypothesis can easily be tested in VAERS. Please discuss whether a

Commented [RH(19R18]: I don't think the other VAERS published data I've seen focuses on the timing of reported deaths so I'm not sure how easily tested this is—any

Commented [S(20R18]: the temp scan results may help here. also, please note that the total number of deaths reported to VAERS following Covid vaccination is far larger

Commented [RH(21R18]: Have the # of deaths exceeded by 8 times in this period—but still not sure how to address OS here- could use help here

Commented [RH(22]: This paragraph is really where I could use some help – I tried to sum some of what we've discussed, and Tom pointed out in his last email- but I feel

Commented [RH(23]: I think this paragraph should come before the deaths paragraph now.
Rosenblum, Hannah (CDC)

Commented [BR(24]: Editorial:
Redundant to say "observation of a post-observation". Please revise.

Commented [RH(25R24]: Thank you - edited
Rosenblum, Hannah (CDC)
2021-08-11 17:07:00

This study has several strengths and several limitations. Strengths include a large sample size and comprehensive capture of national data from two complementary surveillance systems. Data on doses administered are available for estimating reporting rates for VAERS, as the U.S. government provides all COVID-19 and collects administration data from jurisdictions. Therefore, the reporting rates calculated here use the number of mRNA vaccine doses administered as a denominator,[50] while for other vaccines the only denominators available are doses distributed, which is variably larger than dose administered. V-safe data illustrating the effects of mRNA vaccination on daily activities and work during the week following vaccination provide new information has not previously available. Limitations include that VAERS data are based on passive surveillance, and may therefore be subject to underreporting, and variable or incomplete reporting.[8] For this analysis, reports of death in VAERS were individually reviewed by physicians and follow-up is ongoing to obtain additional records for reports of death missing death certificates, autopsy reports, or other medical records; however, not all other serious AE reports were individually reviewed. VAERS reports require interpretation to determine if AE reports meet clinical case definitions.[51] ~~Though~~ EB data mining has multiple limitations[22] including that ~~is used to screen for safety alerts,~~ an absence of a~~n~~ 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. Routine screening of VAERS reports may also not be sensitive enough to pick up true associations, particularly if they occur in specific age groups. V-safe is voluntary and requires smartphone access. Participants are asked about pre-specified reactions; this report focused on the first 7 days post-vaccination. Because a subset of all vaccine recipients chose to participate in v-safe, the results likely are not generalizable to the entire vaccinated population in the United States. Participants in v-safe may also be lost to follow up as there is not a requirement for continuous enrollment.

Commented [BR(26]: Please explain if the v-safe system allows for free text reporting in addition to a list of predefined adverse events.
Office of Science
2021-08-11 10:53:00

Commented [RH(27R26]: Thanks added
Rosenblum, Hannah (CDC)
2021-08-11 17:07:00

Commented [BR(28]: Consider including another limitation of this report: it focused only on AE reported in v-safe for the first 7 days post-vaccination.
Office of Science
2021-08-11 10:55:00

Commented [RH(29R28]: Added to limitations
Rosenblum, Hannah (CDC)
2021-08-11 17:07:00

During the first six months of the U.S. COVID-19 vaccination program, more than 50% of the eligible population received at least one dose of COVID-19 vaccine.[20] VAERS and v-safe data from this period demonstrate a post-authorization safety profile for mRNA COVID-19 vaccines that is consistent with pre-authorization trials[3,4] and early post-authorization surveillance reports.[10,11] Serious AEs have been identified following mRNA vaccinations; however, based on the most current information, these events are rare. Vaccines are the most effective tool to preventing serious COVID-19 disease outcomes and the benefits of immunization in preventing serious morbidity and mortality clearly favor vaccination.[26-28] VAERS and v-safe, two complementary surveillance systems, will continue to provide data needed to inform immunization policy makers, medical and immunization providers, and the public about the safety of COVID-19 vaccination.

Research in context

Evidence before this study

We searched PubMed for articles published through July 12, 2021, using the terms ("BNT162b2" or "mRNA-1273" or "mRNA COVID-19 vaccine") AND ("reactogenicity" or "side-effects" or "adverse effects" or "health impact") not restricted by language or type of publication. Among 100 results, publications describing the health impacts following vaccination by BNT162b2 or mRNA-1273 are limited. Available literature from the United States included reports of manufacturer-sponsored phase 1–3 clinical trials. Additionally, we found seven published articles from the United States, one published article from United Kingdom and two preprints from the United States investigating reactogenicity and adverse events in mRNA vaccination. These articles discussed reactogenicity and adverse events following mRNA vaccination. No study included the period through June 2021.

Added value of this study

In this large, observational study, we assessed reactogenicity, health impact, and adverse events reported following mRNA COVID-19 vaccination during the first six months of the U.S. vaccination program. We found that reported reactions to mRNA vaccination were mostly mild in severity and transient in duration, and the great majority of reports were non-serious. Reactions and health impact were reported more frequently in females compared to males, and in individuals aged <65 years compared to older individuals. Health impact information for adults from v-safe is presented here for the first time. Deaths, overall and for specific causes by age, were reported.

Implications of all the available evidence

The findings from complementary surveillance systems from the first six months of mRNA vaccination in the United States are consistent with pre-authorization clinical trials and early post-authorization reports. Mild-to-moderate transient reactogenicity should be anticipated, particularly among younger recipients and female recipients. . As these data inform immunization policy recommendations and clinical considerations, the federal monitoring system continues to update the benefit-risk balance of vaccine

Commented [BR(30]: Suggest being more specific, e.g., younger recipients, females.
Office of Science
2021-08-11 11:01:00

Commented [RH(31R30]: edited
Rosenblum, Hannah (CDC)
2021-08-17 14:37:00

recommendations, particularly in the setting of the association of specific serious adverse events and COVID-19 vaccination.

References

1. Oliver SE, Gargano JW, Marin M, et al. The Advisory Committee on Immunization Practices' Interim Recommendation for Use of Pfizer-BioNTech COVID-19 Vaccine - United States, December 2020. *MMWR Morbidity and mortality weekly report 2020*; **69**(50): 1922-4.

2. Oliver SE, Gargano JW, Marin M, et al. The Advisory Committee on Immunization Practices' Interim Recommendation for Use of Moderna COVID-19 Vaccine - United States, December 2020. *MMWR Morbidity and mortality weekly report* 2021; **69**(5152): 1653-6.

3. Polack FP, Thomas SJ, Kitchin N, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *The New England Journal of Medicine* 2020; **383**(27): 2603-15.

4. Baden LR, El Sahly HM, Essink B, et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *The New England Journal of Medicine* 2021; **384**(5): 403-16.

5. Anderson EJ, Rouphael NG, Widge AT, et al. Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. *New England Journal of Medicine* 2020; **383**(25): 2427-38.

6. Chen RT, Davis RL, Rhodes PH. Special methodological issues in pharmacoepidemiology studies of vaccine safety. In: Strom BL, editor. Pharmacoepidemiology. 4th John Wiley & Sons; Sussex: 2005

7. Dooling K, McClung N, Chamberland M, et al. The Advisory Committee on Immunization Practices' Interim Recommendation for Allocating Initial Supplies of COVID-19 Vaccine - United States, 2020. *MMWR Morbidity and mortality weekly report* 2020; **69**(49): 1857-9.

8. Shimabukuro TT, Nguyen M, Martin D, DeStefano F. Safety monitoring in the Vaccine Adverse Event Reporting System (VAERS). *Vaccine* 2015; **33**(36): 4398-405.

9. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/vsafe.html

10. Gee J, Marquez P, Su J, et al. First Month of COVID-19 Vaccine Safety Monitoring - United States, December 14, 2020-January 13, 2021. *MMWR Morbidity and mortality weekly report* 2021; **70**(8): 283-8.

11. Chapin-Bardales J, Gee J, Myers T. Reactogenicity Following Receipt of mRNA-Based COVID-19 Vaccines. *JAMA* 2021; **325**(21): 2201-2.

12. Shimabukuro TT, Cole M, Su JR. Reports of Anaphylaxis After Receipt of mRNA COVID-19 Vaccines in the US—December 14, 2020-January 18, 2021. *JAMA* 2021; **325**(11): 1101-2.

13. Shay DK, Gee J, Su JR, et al. Safety Monitoring of the Janssen (Johnson & Johnson) COVID-19 Vaccine - United States, March-April 2021. *MMWR Morbidity and mortality weekly report* 2021; **70**(18): 680-4.

14. Allergic Reactions Including Anaphylaxis After Receipt of the First Dose of Pfizer-BioNTech COVID-19 Vaccine - United States, December 14-23, 2020. *MMWR Morbidity and mortality weekly report* 2021; **70**(2): 46-51.

15. https://www.cdc.gov/vaccinesafety/ensuringsafety/monitoring/vaers/index.html

16. Code of Federal Regulations Title 21
https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr

17. https://www.cdc.gov/vaccinesafety/pdf/VAERS-v2-SOP.pdf

18. https://www.cdc.gov/nchs/nvss/leading-causes-of-death.htm

19. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/vsafe.html

20. https://covid.cdc.gov/covid-data-tracker/#datatracker-home 21. 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.

22. 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; Jul:**36**(7):547-56.

23. https://www.cdc.gov/vaccinesafety/ensuringsafety/monitoring/vsd/index.html

24. https://www.cdc.gov/vaccines/acip/work-groups-vast/index.html

25. https://www.cdc.gov/vaccines/acip/meetings/index.html

26. MacNeil JR, Su JR, Broder KR, et al. Updated Recommendations from the Advisory Committee on Immunization Practices for Use of the Janssen (Johnson & Johnson) COVID-19 Vaccine After Reports of Thrombosis with Thrombocytopenia Syndrome Among Vaccine Recipients - United States, April 2021. *MMWR Morbidity and mortality weekly report* 2021; **70**(17): 651-6.

27. Gargano JW, Wallace M, Hadler SC, et al. Use of mRNA COVID-19 Vaccine After Reports of Myocarditis Among Vaccine Recipients: Update from the Advisory Committee on Immunization Practices - United States, June 2021. *MMWR Morbidity and mortality weekly report* 2021; **70**(27): 977-82.

28. Rosenblum HG, Hadler SC, Moulia D et al. Use of COVID-19 Vaccines After Reports of Adverse Events Among Adult Recipients of Janssen (Johnson & Johnson) and mRNA COVID-19 Vaccines (Pfizer-BioNTech and Moderna): Update from the Advisory Committee on Immunization Practices – United States, July 2021. *MMWR Morbidity and mortality weekly report* 2021: **70**(32); 1094-1099.

29. Menni C, Klaser K, May A, et al. Vaccine side-effects and SARS-CoV-2 infection after vaccination in users of the COVID Symptom Study app in the UK: a prospective observational study. *Lancet Infect Dis.* 2021; **21**(7):939-949

30. Hervé C, Laupèze B, Del Giudice G, Didierlaurent AM, Tavares Da Silva F. The how's and what's of vaccine reactogenicity. *NPJ Vaccines* 2019;**4**: 39.

31. Zimmermann P, Curtis N. Factors That Influence the Immune Response to Vaccination. *Clin Microbiol Rev* 2019; **32**(2).

32. Klein SL, Jedlicka A, Pekosz A. The Xs and Y of immune responses to viral vaccines. *Lancet Infect Dis* 2010; **10**(5): 338-49.

33. Pathiravasan CH, Zhang Y, Trinquart L, et al. Adherence of Mobile App-Based Surveys and Comparison With Traditional Surveys: eCohort Study. *J Med Internet Res* 2021; **23**(1): e24773.

34. Guo X, Vittinghoff E, Olgin JE, Marcus GM, Pletcher MJ. Volunteer Participation in the Health eHeart Study: A Comparison with the US Population. *Sci Rep* 2017; **7**(1): 1956.

35. Millar MM, Elena JW, Gallicchio L, et al. The feasibility of web surveys for obtaining patient-reported outcomes from cancer survivors: a randomized experiment comparing survey modes and brochure enclosures. *BMC Med Res Methodol* 2019; **19**(1): 208.

36. Messer BL, Dillman DA. Surveying the General Public over the Internet Using Address-Based Sampling and Mail Contact Procedures. *Public Opinion Quarterly* 2011; **75**(3): 429-57.

37. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-april-2021/

38. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-june-2021/

39. https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/21/fact-sheet-president-biden-to-call-on-all-employers-to-provide-paid-time-off-for-employees-to-get-vaccinated-after-meeting-goal-of-200-million-shots-in-the-first-100-days/

40. Moro PL, Arana J, Cano M, Lewis P, Shimabukuro TT. Deaths Reported to the Vaccine Adverse Event Reporting System, United States, 1997-2013. Clin Infect Dis 2015; 61(6): 980-7.

41. Murphy SL, Xu J, Kochanek KD, Arias E, Tejada-Vera B. Deaths: Final Data for 2018. Natl Vital Stat Rep 2021; **69**(13): 1-83.

42. Miller ER, Lewis P, Shimabukuro TT, et al. Post-licensure safety surveillance of zoster vaccine live (Zostavax®) in the United States, Vaccine Adverse Event Reporting System (VAERS), 2006-2015. *Human vaccines & immunotherapeutics* 2018; **14**(8): 1963-9

43. Moro PL, Haber P, McNeil MM. Challenges in evaluating post-licensure vaccine safety: observations from the Centers for Disease Control and Prevention. *Expert Review of Vaccines* 2019; **18**(10): 1091-101

44. Haber P, Moro PL, Ng C, et al. Safety review of tetanus toxoid, reduced diphtheria toxoid, acellular pertussis vaccines (Tdap) in adults aged ≥65 years, Vaccine Adverse Event reporting System (VAERS), United States, September 2010-December 2018. *Vaccine* 2020; **38**(6): 1476-80

45. Woo EJ, Moro PL. Postmarketing safety surveillance of quadrivalent recombinant influenza vaccine: Reports to the vaccine adverse event reporting system. *Vaccine* 2021; **39**(13): 1812-7.

46. https://www.cdc.gov/vaccines/covid-19/clinical-considerations/managing-anaphylaxis.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fvaccines%2Fcovid-19%2Finfo-by-product%2Fpfizer%2Fanaphylaxis-management.html

47. Abu Mouch S, Roguin A, Hellou E, et al. Myocarditis following COVID-19 mRNA vaccination. Vaccine 2021.

48. Marshall M, Ferguson ID, Lewis P, et al. Symptomatic Acute Myocarditis in Seven Adolescents Following Pfizer-BioNTech COVID-19 Vaccination. *Pediatrics* 2021 Jun 4;e2021052478. *Online ahead of print*

49. https://www.cdc.gov/vaccines/covid-19/clinical-considerations/myocarditis.html

50. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/distributing/first-doses.html

51. https://vaers.hhs.gov/data/dataguide.html

52. https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/fully-vaccinated-people.html

***Table* 1: Characteristics of reports received and processed by Vaccine Adverse Events Reporting System (VAERS) for mRNA COVID-19 vaccines—December 14, 2020–June 14, 2021**

	Both mRNA vaccines† (n=340,522)	BNT162b2 vaccine (n=164,669)	mRNA-1273 vaccine (n=175,816)
Reports			
Non-serious adverse event reports	313,499 (92·1)	150,486 (91·4)	162,977 (92·7)
Serious reports, including death	27,023 (7·9)	14,183 (8·6)	12,839 (7·3)
Serious, non-death adverse event reports	22,527 (6·6)	12,078 (7·3)	10,448 (5·9)
Death	4,496 (1·3)	2,105 (1·3)	2,391 (1·4)
Sex			
Female	246,085 (72·3)	116,587 (70·8)	129,475 (73·6)
Male	88,311 (25·9)	45,157 (27·4)	43,140 (24·5)
Unknown	6,126 (1·8)	2,925 (1·8)	3,201 (1·8)
Age (years)			
16–17	6,874 (2·0)	3,283 (2·0)	3,591 (2·0)
18–49	154,171 (45·3)	76,385 (46·4)	77,773 (44·2)
50–64	84,949 (24·9)	40,367 (24·5)	44,572 (25·4)
65–74	49,755 (14·6)	20,048 (12·2)	29,702 (16·9)
75–84	21,418 (6·3)	9,021 (5·5)	12,392 (7·1)
≥85	7,595 (2·2)	3,564 (2·2)	4,027 (2·3)
Unknown	15,760 (4·6)	12,001 (7·3)	3,759 (2·1)
Race/Ethnicity			
Hispanic/Latino	23,480 (6·9)	11,217 (6·8)	12,260 (7·0)
Non-Hispanic			
White	169,877 (49·9)	73,398 (44·6)	96,469 (54·9)
Black	10,446 (3·1)	5,104 (3·1)	5,342 (3·0)
Asian	10,172 (3·0)	5,038 (3·1)	5,131 (2·9)
American Indian or Alaska Native	1,414 (0·4)	615 (0·4)	799 (0·5)
Native Hawaiian or Other Pacific Islander	441 (0·1)	209 (0·1)	232 (0·1)
Multiple races	3,542 (1·0)	1,578 (1·0)	1,964 (1·1)
Other races	1,684 (0·5)	808 (0·5)	876 (0·5)
Unknown race	2,593 (0·8)	1,422 (0·9)	1,171 (0·7)
Unknown ethnicity			
White	28,787 (8·5)	15,497 (9·4)	13,289 (7·6)
Black	4,189 (1·2)	2,524 (1·5)	1,662 (1·0)
Asian	2,435 (0·7)	1,396 (0·9)	1,039 (0·6)
American Indian or Alaska Native	724 (0·2)	348 (0·2)	375 (0·2)
Native Hawaiian or Other Pacific Islander	105 (0·03)	56 (0·03)	49 (0·03)
Multiple races	590 (0·2)	301 (0·2)	289 (0·2)
Other races	4,709 (1·4)	2,838 (1·7)	1,870 (1·1)
Unknown race and ethnicity	75,334 (22·1)	42,320 (25·7)	32,999 (18·8)
Signs or symptoms most frequently reported, non-serious*			
Headache	64,064 (20·4)	30,907 (20·5)	33,154 (20·3)
Fatigue	52,048 (16·6)	24,805 (16·5)	27,241 (16·7)
Pyrexia	51,023 (16·3)	22,185 (14·7)	28,837 (17·7)
Chills	49,234 (15·7)	21,638 (14·4)	27,595 (16·9)
Pain	47,745 (15·2)	21,506 (14·3)	26,238 (16·1)
Nausea	37,333 (11·9)	18,066 (12·0)	19,267 (11·8)
Dizziness	37,257 (11·9)	20,307 (13·5)	16,950 (10·4)
Pain in extremity	31,753 (10·1)	14,098 (9·4)	17,653 (10·8)
Injection site pain	28,949 (9·2)	10,462 (7·0)	18,487 (11·3)
Injection site erythema	22,351 (7·1)	2,991 (2·0)	19,360 (11·9)
Signs or symptoms most frequently reported, serious*			
Dyspnea	4,175 (15·4)	2,210 (15·6)	1,965 (15·3)
Death‡	3,802 (14·1)	1,753 (12·4)	2,039 (15·9)
Pyrexia	2,986 (11·0)	1,469 (10·4)	1,517 (11·8)
Fatigue	2,608 (9·7)	1,395 (9·8)	1,213 (9·4)
Headache	2,567 (9·5)	1,360 (9·6)	1,207 (9·4)
Chest pain	2,300 (8·5)	1,310 (9·2)	990 (7·7)
Nausea	2,228 (8·2)	1,160 (8·2)	1,068 (8·3)
Pain	2,222 (8·2)	1,195 (8·4)	1,027 (8·0)
Asthenia	2,194 (8·1)	1,084 (7·6)	1,110 (8·6)
Dizziness	2,069 (7·7)	1,111 (7·8)	958 (7·5)

Data are n (%).
*Symptoms refers to MedDRA preferred terms (PTs) and are ordered by most frequently reported for both vaccines. MedDRA PTs are not mutually exclusive.
†Total includes reports without a vaccine manufacturer listed.
‡Not all reports of death were coded with the MedDRA PT of 'death'

***Table* 2: Frequency and reporting rates of adverse events of special interest reported to Vaccine Adverse Event Reporting System (VAERS) by recipients of mRNA COVID-19 vaccine—December 14, 2020–June 14, 2021**

	Both mRNA vaccines		BNT162b2 vaccine		mRNA-1273 vaccine	
	n	Reports per million doses*	n	Reports per million doses†	n	Reports per million doses‡
Non-serious adverse event reports	313,499	1,049·2	150,486	900·2	162,977	1,238·1
Serious reports, including death	27,023	90·4	14,183	84 8	12,839	97·5
Serious, non-death adverse event reports	22,527	75·4	12,078	72 2	10,448	79·4
Reports§ of adverse events of special interest**						
COVID-19	9,344	31·3	7,184	43 0	2,160	16·4
Coagulopathy††	4,320	14·5	2,343	14 0	1,977	15·0
Seizure	2,733	9·1	1,478	8·8	1,255	9·5
Stroke‡‡	1,937	6·5	981	5·9	955	7·3
Bells' Palsy	1,918	6·4	1,057	6·3	861	6·5
Anaphylaxis	1,639	5·5	972	5·8	667	5·1
Myopericarditis	1,307	4·4	813	4·9	494	3·8
Acute Myocardial Infarction	1,118	3·7	610	3·6	508	3·9
Appendicitis	383	1·3	258	1·5	125	1·0
Guillain-Barré Syndrome	293	1·0	154	0·9	139	1·1
Multisystem Inflammatory Syndrome in Adults	119	0·4	60	0·4	59	0·4
Transverse Myelitis	98	0·3	55	0·3	43	0·3
Narcolepsy	21	0·1	12	0·1	9	0·1

*298,792,852 doses of mRNA vaccine were administered in the study period.
†167,177,332 doses of BNT162b2 vaccine were administered in the study period.
‡131,639,515 doses of mRNA-1273 vaccine were administered in the study period.
§These represent reports, not confirmed by case definition.
**Reported death is an adverse event of special interest but counts appear in following tables. Events are not mutually exclusive.
††Coagulopathy is an aggregate term capturing three specific adverse events: 1) thrombocytopenia, 2) deep venous thrombosis/pulmonary embolism, and 3) disseminated intravascular coagulopathy.
‡‡No vaccine manufacturer was provided for one report of stroke.

***Table* 3: Characteristics of deaths reported to Vaccine Adverse Event Reporting System (VAERS) by recipients of mRNA COVID-19 vaccine—December 14, 2020–June 14, 2021**

	Both mRNA vaccines (n=4,472*)		**BNT162b2 vaccine** (n=2,087)		**mRNA-1273 vaccine** (n=2,385)	
	n (%)	Reports per million doses†	n (%)	Reports per million doses‡	n (%)	Reports per million doses §
Sex						
Female	1,906 (42·6)	6·4	918 (44·0)	5·5	988 (41·4)	7·5
Male	2,486 (55·6)	8·3	1,117 (53·5)	6·7	1,369 (57·4)	10·4
Unknown	80 (1·8)	0·3	52 (2·5)	0·3	28 (1·2)	0·2
Age (years)						
16–17	6 (0·1)	0·02	6 (0·3)	0·04	..	..
18–29	51 (1·1)	0·2	27 (1·3)	0·2	24 (1·0)	0·2
30–39	94 (2·1)	0·3	50 (2·4)	0·3	44 (1·8)	0·3
40–49	151 (3·4)	0·5	74 (3·5)	0·4	77 (3·2)	0·6
50–59	328 (7·3)	1·1	132 (6·3)	0·8	196 (8·2)	1·5
60–69	765 (17·1)	2·6	354 (17·0)	2·1	411 (17·2)	3·1
70–79	1,118 (25·0)	3·7	497 (23·8)	3·0	621 (26·0)	4·7
80–89	1,128 (25·2)	3·8	529 (25·3)	3·2	599 (25·1)	4·6
≥90	637 (14·2)	2·1	302 (14·5)	1·8	335 (14·0)	2·5
Unknown	194 (4·3)	0·6	116 (5·6)	0·7	78 (3·3)	0·6

*Of 4,496 deaths, 24 were excluded as they could not be confirmed or were duplicate reports upon review.
†298,792,852 doses of mRNA vaccine were administered in the study period.
‡167,177,332 doses of BNT162b2 vaccine were administered in the study period.
§131,639,515 doses of mRNA-1273 vaccine were administered in the study period.

***Table* 4: Most common causes of death among reports received and processed by Vaccine Adverse Event Reporting System (VAERS) following mRNA COVID-19 vaccination (n=4,472)—December 14, 2020–June 14, 2021**

	Death or autopsy certificate available			No death certificate or autopsy available		
ICD-10 Major Group	**Both mRNA vaccines** (n=808)	**BNT162b2 vaccine** (n=401)	**mRNA-1273 vaccine** (n=407)	**Both mRNA vaccines** (n=3,664)	**BNT162b2 vaccine** (n=1,686)	**mRNA-1273 vaccine** (n=1,978)
All reported deaths						
Diseases of the heart	376 (46·5)	161 (40·1)	215 (52·8)	622 (17·0)	296 (17·6)	326 (16·5)
COVID-19 disease	102 (12·6)	62 (15·5)	40 (9·8)	317 (8·7)	178 (10·6)	139 (7·0)
Other	68 (8·4)	38 (9·5)	30 (7·4)	141 (3·8)	68 (4·0)	73 (3·7)
Cerebrovascular diseases	53 (6·6)	28 (7·0)	25 (6·1)	207 (5·6)	101 (6·0)	106 (5·4)
Dementia	41 (5·1)	20 (5·0)	21 (5·2)	9 (0·2)	3 (0·2)	6 (0·3)
Chronic lower respiratory diseases	28 (3·5)	17 (4·2)	11 (2·7)	29 (0·8)	10 (0·6)	19 (1·0)
Malignant neoplasms	27 (3·3)	15 (3·7)	12 (2·9)	68 (1·9)	42 (2·5)	26 (1·3)
Unknown/unclear	27 (3·3)	9 (2·2)	18 (4·4)	1,984 (54·1)	844 (50·1)	1,140 (57·6)
Septicemia	23 (2·8)	12 (3·0)	11 (2·7)	72 (2·0)	47 (2·8)	25 (1·3)
Influenza and pneumonia	22 (2·7)	18 (4·5)	4 (1·0)	113 (3·1)	52 (3·1)	61 (3·1)
Accidents/unintentional injuries	11 (1·4)	3 (0·7)	8 (2·0)	22 (0·6)	8 (0·5)	14 (0·7)
Renal disease	8 (1·0)	5 (1·2)	3 (0·7)	25 (0·7)	7 (0·4)	18 (0·9)
Hematologic disease, other than malignancy	7 (0·9)	5 (1·2)	2 (0·5)	19 (0·5)	9 (0·5)	10 (0·5)
Pneumonitis due to solids and liquids	6 (0·7)	3 (0·7)	3 (0·7)	8 (0·2)	5 (0·3)	3 (0·2)
Diabetes mellitus	4 (0·5)	1 (0·2)	3 (0·7)	6 (0·2)	4 (0·2)	2 (0·1)
Chronic liver disease and cirrhosis	4 (0·5)	3 (0·7)	1 (0·2)	7 (0·2)	4 (0·2)	3 (0·2)
Intentional self-harm	1 (0·1)	1 (0·2)	0 (0·0)	15 (0·4)	8 (0·5)	7 (0·4)

Data are n (%).

(new) Table 5: **Observed deaths vs. expected deaths in a 7-day risk period (need footnotes to Abara paper)**

Commented [RH(32]: Note this is the 5th VAERS table (compared to 2 v-safe and 1 v-safe figure) and the 3rd table regarding death

Rosenblum, Hannah (CDC)
2021-09-09 10:39:00

	Expected		Observed (reported in VAERS)					
	All-cause death rates per million vaccinated persons		Both mRNA vaccines n (rate per vaccinated persons)		BNT162b2 vaccine n (rate per million doses administered)		mRNA-1273 vaccine n (rate per million doses administered)	
	Within 7 days of vaccination	Within 42 days of vaccination	Within 7 days of vaccination	Within 42 days of vaccination	Within 7 days of vaccination	Within 42 days of vaccination	Within 7 days of vaccination	Within 42 days of vaccination
Age (years)								
16–24**	14.2	85.1	14	26	9	17	5	9
25–34	25.5	152.7	32	56	18	31	14	25
35–44	37.4	224.5	54	97	28	47	26	50
45–54	77.0	461.7	118	200	55	91	63	109
55–64	169.8	1,018.6	247	453	99	201	148	252
65–74	343.2	2,059.4	457	878	207	385	250	493
75–84	857.2	5,143.0	412	908	180	396	232	512
≥85	2,601.4	15,608.3	437	980	202	448	235	532
Unknown	--	--	28	40	13	16	15	24
Total	165.6	993.3	1,799	3,638	811	1,632	988	2,006

*calculated from Abara paper per 10,000,000

**Abara paper is 15-24. Age 15 was not included in this paper

Need 42 day columns?

***Table* 5: Demographic characteristics of v-safe participants reporting receipt of mRNA COVID-19 vaccine and completing at least one health survey 0-7 days after vaccination—December 14, 2020–June 14, 2021**

Characteristics	Both mRNA vaccines (n = 7,914,583)	BNT162b2 vaccine		mRNA-1273 vaccine	
		Dose 1 n=3,455,778	Dose 2 n=2,920,526	Dose 1 n= 3,319,737	Dose 2 n=2,753,894
Sex					
Female	4,975,209 (62·9)	2,150,068 (62·2)	1,861,599 (63·7)	2,073,542 (62·5)	1,779,200 (64·6)
Male	2,860,738 (36·1)	1,272,011 (36·8)	1,032,941 (35·4)	1,210,622 (36·5)	947,612 (34·4)
Other	8,872 (0·1)	4,027 (0·1)	3,464 (0·1)	3,443 (0·1)	2,947 (0·1)
Unknown	69,764 (0·9)	29,672 (0·9)	22,522 (0·8)	32,130 (1·0)	24,135 (0·9)
Age (years)					
16–17	73,347 (0·9)	63,865 (1·8)	38,530 (1·3)	946 (0·03)	473 (0·02)
18–49	3,791,839 (47·9)	1,726,465 (50·0)	1,431,627 (49·0)	1,505,760 (45·4)	1,219,210 (44·3)
50–59	1,500,981 (19·0)	653,799 (18·9)	574,422 (19·7)	627,214 (18·9)	531,200 (19·3)
60–64	739,381 (9·3)	315,404 (9·1)	279,350 (9·6)	316,768 (9·5)	270,831 (9·8)
65–74	1,344,721 (17·0)	516,227 (14·9)	452,928 (15·5)	643,663 (19·4)	557,279 (20·2)
≥75	464,314 (5·9)	180,018 (5·2)	143,669 (4·9)	225,386 (6·8)	174,901 (6·4)
Race/Ethnicity					
Hispanic	782,301 (9·9)	346,197 (10·0)	288,263 (9·9)	316,460 (9·5)	256,185 (9·3)
Non-Hispanic					
White	4,701,715 (59·4)	2,059,560 (59·6)	1,896,823 (64·9)	1,979,056 (59·6)	1,830,413 (66·5)
Black	443,938 (5·6)	202,598 (5·9)	176,164 (6·0)	178,981 (5·4)	153,667 (5·6)
Asian	467,932 (5·9)	215,713 (6·2)	196,173 (6·7)	154,498 (4·7)	138,793 (5·0)
American Indian or Alaska Native	27,899 (0·4)	11,161 (0·3)	9,194 (0·3)	13,486 (0·4)	11,410 (0·4)
Native Hawaiian or Other Pacific Islander	19,393 (0·2)	8,500 (0·2)	7,373 (0·3)	7,689 (0·2)	6,664 (0·2)
Multiple races	110,326 (1·4)	50,954 (1·5)	46,129 (1·6)	41,977 (1·3)	38,772 (1·4)
Other races	42,230 (0·5)	19,252 (0·6)	16,757 (0·6)	15,885 (0·5)	13,880 (0·5)
Unknown race	23,420 (0·3)	10,249 (0·3)	9,090 (0·3)	9,502 (0·3)	8,270 (0·3)
Unknown ethnicity*					
White	115,766 (1·5)	48,084 (1·4)	38,674 (1·3)	52,143 (1·6)	42,070 (1·5)
Black	26,865 (0·3)	11,602 (0·3)	8,570 (0·3)	11,993 (0·4)	8,406 (0·3)
Asian	33,146 (0·4)	14,134 (0·4)	11,844 (0·4)	11,356 (0·3)	9,153 (0·3)
American Indian or Alaska Native	3,142 (0·04)	1,206 (0·03)	848 (0·03)	1,582 (0·05)	1,151 (0·04)
Native Hawaiian or Other Pacific Islander	1,945 (0·02)	815 (0·02)	659 (0·02)	800 (0·02)	613 (0·02)
Multiple races	6,370 (0·1)	2,902 (0·1)	2,408 (0·1)	2,478 (0·1)	2,041 (0·1)
Other races	13,148 (0·2)	5,681 (0·2)	4,528 (0·2)	5,414 (0·2)	4,263 (0·2)
Unknown race and ethnicity*	129,647 (1·6)	56,481 (1·6)	45,410 (1·6)	54,969 (1·7)	44,340 (1·6)
Unavailable†	965,400 (12·2)	390,689 (11·3)	161,619 (5·5)	461,468 (13·9)	183,803 (6·7)
Pregnant at time of vaccination	86,801 (1·1)	39,884 (1·2)	39,163 (1·3)	25,255 (0·8)	25,428 (0·9)
Pregnancy test positive after vaccination	27,370 (0·3)	1,548 (0·04)	11,677 (0·4)	4,009 (0·1)	10,199 (0·4)

Data are n (%).

*Unknown indicates that v-safe participants selected unknown or preferred not to say.
†Unavailable refers to information that was not collected or missing in v-safe.

***Table 6*: Reported local and systemic reactions*, and reported health impact following mRNA COVID-19 vaccines reported days 0–7 after vaccination to v-safe, by manufacturer and dose—December 14, 2020 – June 14, 2021**

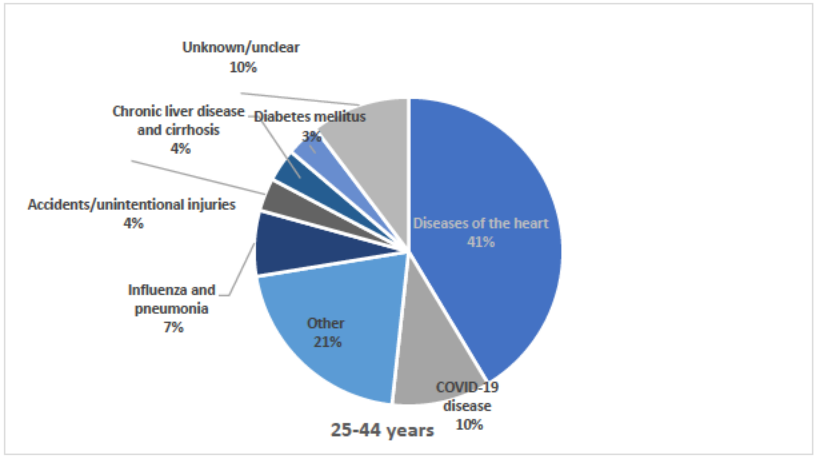
	Both mRNA vaccines		BNT162b2 vaccine		mRNA-1273 vaccine	
	Dose 1 (n=6,775,515)	Dose 2 (n=5,674,420)	Dose 1 (n=3,455,778)	Dose 2 (n=2,920,526)	Dose 1 (n=3,319,737)	Dose 2 (n=2,753,894)
Any injection site reaction	4,644,989 (68·6)	4,068,447 (71·7)	2,212,051 (64·0)	1,908,124 (65·3)	2,432,938 (73·3)	2,160,323 (78·4)
Injection site pain	4,488,402 (66·2)	3,890,848 (68·6)	2,140,843 (61·9)	1,835,398 (62·8)	2,347,559 (70·7)	2,055,450 (74·6)
Swelling	703,790 (10·4)	976,946 (17·2)	246,230 (7·1)	309,718 (10·6)	457,560 (13·8)	667,228 (24·2)
Redness	353,788 (5·2)	640,739 (11·3)	116,108 (3·4)	167,127 (5·7)	237,680 (7·2)	473,612 (17·2)
Itching	376,076 (5·6)	605,633 (10·7)	145,596 (4·2)	191,132 (6·5)	230,480 (6·9)	414,501 (15·1)
Any systemic reaction	3,573,429 (52·7)	4,018,920 (70·8)	1,771,509 (51·3)	1,931,643 (66·1)	1,801,920 (54·3)	2,087,277 (75·8)
Fatigue	2,295,205 (33·9)	3,158,299 (55·7)	1,127,904 (32·6)	1,475,646 (50·5)	1,167,301 (35·2)	1,682,653 (61·1)

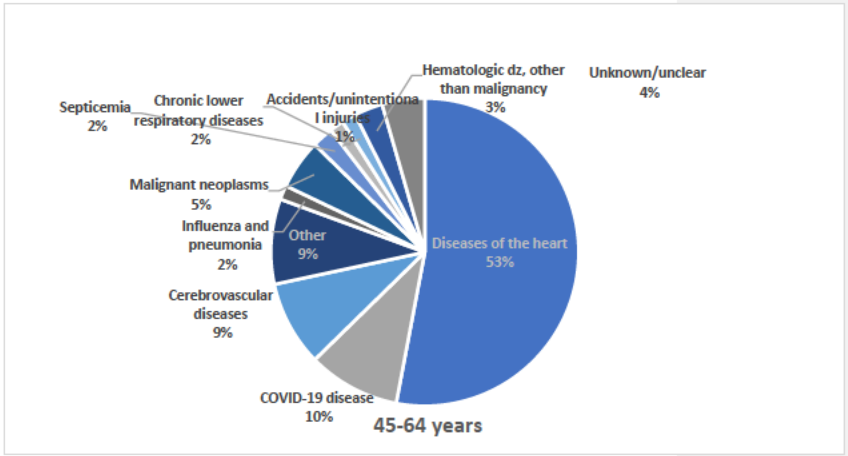
Headache	1,831,471 (27·0)	2,623,721 (46·2)	893,992 (25·9)	1,189,444 (40·7)	937,479 (28·2)	1,434,277 (52·1)
Myalgia	1,423,336 (21·0)	2,478,170 (43·7)	653,821 (18·9)	1,085,365 (37·2)	769,515 (23·2)	1,392,805 (50·6)
Chills	631,546 (9·3)	1,680,185 (29·6)	263,617 (7·6)	642,856 (22·0)	367,929 (11·1)	1,037,329 (37·7)
Fever	642,092 (9·5)	1,679,577 (29·6)	274,650 (7·9)	656,454 (22·5)	367,442 (11·1)	1,023,123 (37·2)
Joint pain	642,006 (9·5)	1,440,927 (25·4)	285,812 (8·3)	591,877 (20·3)	356,194 (10·7)	849,050 (30·8)
Nausea	562,273 (8·3)	901,103 (15·9)	267,160 (7·7)	384,525 (13·2)	295,113 (8·9)	516,578 (18·8)
Diarrhea	383,576 (5·7)	419,044 (7·4)	190,542 (5·5)	198,618 (6·8)	193,034 (5·8)	220,426 (8·0)
Abdominal pain	233,511 (3·4)	359,107 (6·3)	113,872 (3·3)	158,251 (5·4)	119,639 (3·6)	200,856 (7·3)
Rash	85,766 (1·3)	99,878 (1·8)	41,565 (1·2)	42,662 (1·5)	44,201 (1·3)	57,216 (2·1)
Vomiting	55,710 (0·8)	91,727 (1·6)	25,336 (0·7)	36,761 (1·3)	30,374 (0·9)	54,966 (2·0)
With reported health impact*	808,963 (11·9)	1,821,421 (32·1)	361,834 (10·5)	740,529 (25·4)	447,129 (13·5)	1,080,892 (39·2)
Unable to do normal activity	658,330 (9·7)	1,501,679 (26·5)	290,207 (8·4)	598,584 (20·5)	368,123 (11·1)	903,095 (32·8)
Unable to work	305,709 (4·5)	911,366 (16·1)	135,063 (3·9)	360,411 (12·3)	170,646 (5·1)	550,955 (20·0)
Reported medical care	56,647 (0·8)	53,077 (0·9)	27,358 (0·8)	25,568 (0·9)	29,289 (0·9)	27,509 (1·0)
Telehealth	19,562 (0·3)	19,770 (0·3)	9,318 (0·3)	9,238 (0·3)	10,244 (0·3)	10,532 (0·4)
Clinic	18,671 (0·3)	16,793 (0·3)	9,109 (0·3)	8,487 (0·3)	9,562 (0·3)	8,306 (0·3)
Emergency visit	9,907 (0·1)	8,907 (0·2)	5,087 (0·1)	4,494 (0·2)	4,820 (0·1)	4,413 (0·2)
Hospitalization	1,896 (0·03)	2,053 (0·04)	915 (0·03)	1,001 (0·03)	981 (0·03)	1,052 (0·04)

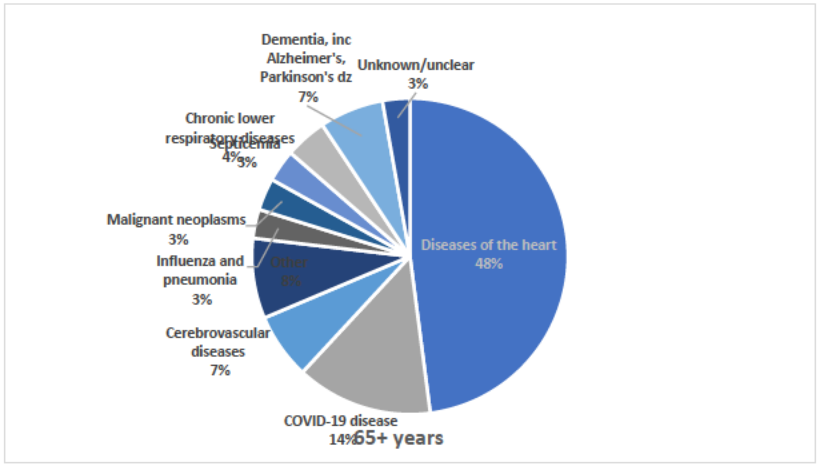
Data are n (%).
*Reports of local and systemic reactions, and reports of health impact are not mutually exclusive.

***Figure 1*: Percent distribution of the 10 leading causes of death, by age, among reported deaths with death certificate or autopsy to Vaccine Adverse Event Reporting System (VAERS) December 14, 2020–June 14, 2021 following mRNA vaccination**

Commented [RH(33]: make supplemental or delete?
Data is captured in supplemental table 2
Rosenblum, Hannah (CDC)
2021-09-08 16:03:00







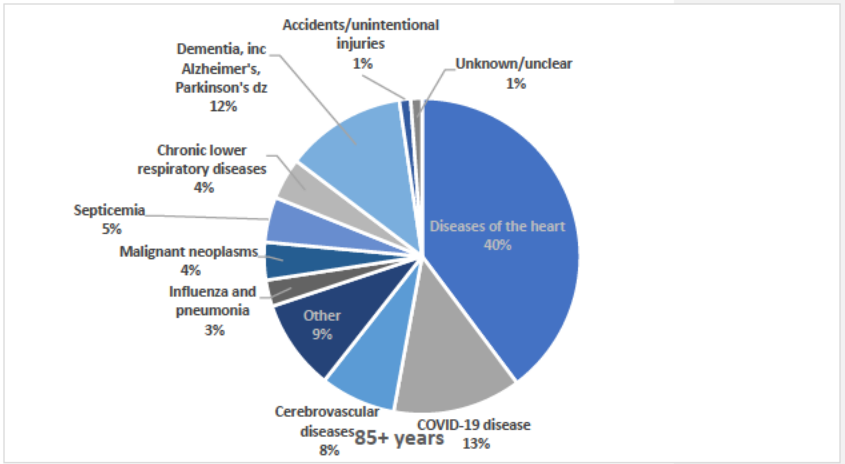
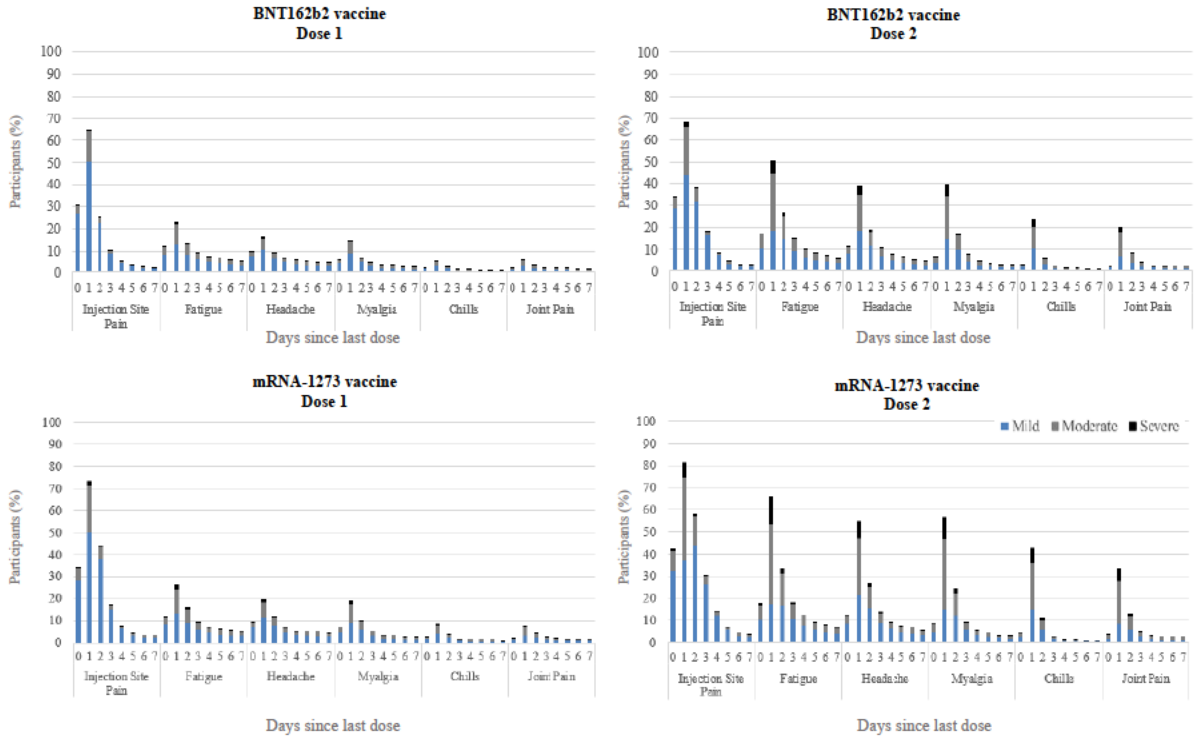


***Figure* 2: Local and systemic reactions* to mRNA COVID-19 vaccine reported in v-safe, by manufacturer, dose, days after vaccination, and severity†**



*Top five reactions determined by reported frequency after second dose of both mRNA COVID-19 vaccines in v-safe, excluding fever because it was not rated mild, moderate, or severe.
†Mild was defined as "noticeable symptoms but they aren't a problem", moderate was defined as "symptoms that limit normal activities, and severe symptoms "make normal daily activities difficult or impossible"

***Supplemental Table* 1: mRNA COVID-19 vaccine doses administered in the United States—December 14, 2020–June 14, 2021**

Characteristics	Both mRNA vaccines† (n=298,792,852)	BNT162b2 vaccine† (n=167,177,332)	mRNA-1273 vaccine† (n=131,639,515)
Sex			
Female	155,969,573 (53·2)	86,507,992 (53·5)	69,461,582 (52·8)
Male	134,373,958 (45·8)	73,768,602 (45·6)	60,605,356 (46·1)
Unknown	2,868,979 (1·0)	1,452,344 (0·9)	1,416,634 (1·1)
Age (years)			
16–17*	5,506,763 (1·8)	5,365,855 (3·2)	140,908 (0·1)
18–49	126,288,626 (42·3)	74,999,327 (44·9)	51,289,299 (39·0)
50–64	79,207,752 (26·5)	43,595,972 (26·1)	35,611,780 (27·1)
65–74	51,699,307 (17·3)	25,402,217 (15·2)	26,297,090 (20·0)
75–84	27,731,181 (9·3)	13,555,128 (8·1)	14,176,053 (10·8)
≥85	8,359,223 (2·8)	4,248,648 (2·5)	4,110,575 (3·1)
Unknown	23,995 (0·01)	10,185 (0·01)	13,810 (0·01)
Race/Ethnicity			
Hispanic	31,599,632 (10·8)	17,964,345 (11·1)	13,635,287 (10·4)
Non-Hispanic			
White	112,698,875 (38·4)	61,996,607 (38·3)	50,702,268 (38·6)
Asian	11,789,429 (4·0)	7,258,033 (4·5)	4,531,396 (3·4)
Black	16,848,436 (5·7)	9,665,586 (6·0)	7,182,849 (5·5)
American Indian or Alaska Native	1,738,938 (0·6)	842,263 (0·5)	896,674 (0·7)
Native Hawaiian or Other Pacific Islander	508,285 (0·2)	295,634 (0·2)	212,651 (0·2)
Multiple races	8,856,800 (3·0)	5,037,828 (3·1)	3,818,972 (2·9)
Other races	6,949,404 (2·4)	4,161,353 (2·6)	2,788,051 (2·1)
Unknown race and ethnicity	102,227,532 (34·9)	54,511,493 (33·7)	47,716,039 (36·3)

Data are n (%).
*mRNA-1273 vaccine was not authorized for individuals <18 years during this period, reported mRNA-1273 doses are either from clinical trials or were administered or reported in error
†Totals reflect the number of doses in age categories. Missing doses for sex and race/ethnicity are due to certain jurisdictions that report data in aggregate.

***Supplemental Table* 2: Causes of death among reported deaths to Vaccine Adverse Event Reporting System (VAERS) December 14, 2020–June 14, 2021 following mRNA vaccination, by age**

Commented [RH(34]: Combine 16-24 and 25-34? n=9
Rosenblum, Hannah (CDC)
2021-08-27 21:56:00

Reports of death with death cert or autopsy ICD-10 Major Group All	All 808	16-24 n=4		25-34 n=5		35-44 n=24		45-54 n=37		55-64 n=101		65-74 n=171		75-84 n=202		85+ n=264	
		n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Diseases of the heart	376	1	25.0	2	40.0	10	41.7	21	56.8	50	49.5	98	57.3	91	45.0	103	39.0
COVID-19 disease	102	1	25.0	1	20.0	2	8.3	2	5.4	11	10.9	19	11.1	32	15.8	34	12.9
Cerebrovascular diseases	53	0	0	0	0	0	0	4	10.8	8	7.9	8	4.7	13	6.4	20	7.6
Other	68	0	0	1	20.0	5	20.8	2	5.4	10	9.9	11	6.4	15	7.4	24	9.1
Influenza and pneumonia	22	0	0	0	0	2	8.3	0	0	2	2.0	4	2.3	7	3.5	7	2.7
Malignant neoplasms	27	0	0	0	0	0	0	2	5.4	5	5.0	2	1.2	8	4.0	10	3.8
Septicemia	23	0	0	0	0	0	0	1	2.7	2	2.0	2	1.2	6	3.0	12	4.5
Chronic lower respiratory diseases	28	0	0	0	0	0	0	0	0	2	2.0	11	6.4	4	2.0	11	4.2
Dementia, inc Alzheimer's, Parkinson's dz	41	0	0	0	0	0	0	0	0	1	1.0	1	0.6	7	3.5	32	12.1
Accidents/unintentional injuries	11	0	0	0	0	1	4.2	0	0	2	2.0	2	1.2	3	1.5	3	1.1
Renal dz, incl nephritis and chronic dz	8	0	0	0	0	0	0	1	2.7	0	0	3	1.8	3	1.5	1	0.4
Hematologic dz, other than malignancy	7	0	0	0	0	0	0	1	2.7	3	3.0	0	0	1	0.5	2	0.8
Intentional self-harm	1	1	25.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pneumonitis due to solids and liquids	6	0	0	0	0	0	0	0	0	1	1.0	0	0	3	1.5	2	0.8
Chronic liver disease and cirrhosis	4	0	0	1	20.0	0	0	0	0	1	1.0	1	0.6	1	0.5	0	0
Diabetes mellitus	4	0	0	0	0	1	4.2	0	0	0	0	2	1.2	1	0.5	0	0
Unknown/unclear	27	1	25.0	0	0	3	12.5	3	8.1	3	3.0	7	4.1	7	3.5	3	1.1

***Supplemental Table 3*: Causes and impressions of death among reported deaths to Vaccine Adverse Event Reporting System (VAERS) December 14, 2020–June 14, 2021 following mRNA vaccination**

ICD-10 Major Group and impression	All reports of death	Reported deaths with death certificate or autopsy
All reported deaths, n	4,472	808
Diseases of the heart	998 (22.3)	376 (46.5)
Aortic dissection, aneurysm or aortitis	13 (0.3)	7 (0.9)
Arrhythmia	42 (0.9)	18 (2.2)
Atherosclerotic cardiovascular or hypertensive cardiovascular disease	129 (2.9)	98 (12.1)
Cardiac arrest	321 (7.2)	79 (9.8)
Cardiomyopathy or hypertrophy	17 (0.4)	12 (1.5)
Heart failure	104 (2.3)	47 (5.8)
Myocardial infarction	247 (5.5)	87 (10.8)
Myocarditis	4 (0.1)	0 (0.0)
Pulmonary embolism	84 (1.9)	17 (2.1)
Other cardiac cause	37 (0.8)	11 (1.4)
COVID-19 disease	419 (9.4)	102 (12.6)
Cerebrovascular diseases	260 (5.8)	53 (6.6)
Other	209 (4.7)	68 (8.4)
Disseminated herpes zoster	2 (0.04)	1 (0.1)
Drug overdose/intoxication	7 (0.2)	5 (0.6)
Failure to thrive	9 (0.2)	9 (1.1)
Gastrointestinal[*]	36 (0.8)	8 (1.0)
Hemorrhage/Hemorrhagic shock	4 (0.1)	2 (0.2)
Metabolic derangement	4 (0.1)	1 (0.1)
Multiorgan failure	28 (0.6)	7 (0.9)
Natural	2 (0.04)	2 (0.2)
Neurologic[†]	28 (0.6)	6 (0.7)
Obesity	2 (0.04)	2 (0.2)
Respiratory failure	63 (1.4)	22 (2.7)
Vaccine related[‡]	4 (0.1)	3 (0.4)
Influenza and pneumonia	135 (3.0)	22 (2.7)
Malignant neoplasms	95 (2.1)	27 (3.3)
Septicemia	95 (2.1)	23 (2.8)
Chronic lower respiratory diseases	57 (1.3)	28 (3.5)
Dementia	50 (1.1)	41 (5.1)
Accidents/unintentional injuries	33 (0.7)	11 (1.4)
Renal disease	33 (0.7)	8 (1.0)
Hematologic disease, other than malignancy	26 (0.6)	7 (0.9)
Intentional self-harm	16 (0.4)	1 (0.1)
Pneumonitis due to solids and liquids	14 (0.3)	6 (0.7)
Chronic liver disease and cirrhosis	11 (0.2)	4 (0.5)
Diabetes mellitus	10 (0.2)	4 (0.5)
Unknown/unclear	2,011 (45.0)	27 (3.3)

[†]Data are n (%) unless otherwise stated.
[*]Gastrointestinal includes gastrointestinal bleeding, bowel obstruction/perforation, mesenteric ischemia, pancreatitis.
[†]Neurologic includes amyotrophic lateral sclerosis, encephalopathy, hydrocephalus, Guillain-Barré syndrome, seizure.
[‡]Vaccine related includes systemic inflammatory response syndrome from vaccine reaction, anaphylaxis post-COVID-19 vaccination

***Supplemental Table* 4: Local and systemic reactions* 0–7 days after vaccination by sex, age, and dose number, reported in v-safe—December 14, 2020–June 14, 2021**

	Female		Male		<65 years		≥65 years	
	Dose 1	**Dose 2**	**Dose 1**	**Dose 2**	**Dose 1**	**Dose 2**	**Dose 1**	**Dose 2**
	(n=4,223,610)	(n=3,640,799)	(n=2,482,633)	(n=1,980,553)	(n=5,210,221)	(n=4,345,643)	(n=1,565,294)	(n=1,328,777)
Any injection site reaction	3,095,194 (73·3)	2,792,488 (76·7)	1,498,108 (60·3)	1,235,278 (62·4)	3,835,618 (73·6)	3,290,206 (75·7)	809,371 (51·7)	778,241 (58·6)
Injection site pain	2,989,733 (70·8)	2,666,734 (73·3)	1,448,440 (58·3)	1,184,914 (59·8)	3,728,795 (71·6)	3,179,024 (73·2)	759,607 (48·5)	711,824 (53·6)
Swelling	539,793 (12·8)	771,962 (21·2)	154,980 (6·2)	194,033 (9·8)	604,868 (11·6)	812,126 (18·7)	98,922 (6·3)	164,820 (12·4)
Redness	283,345 (6·7)	529,175 (14·5)	66,134 (2·7)	104,933 (5·3)	295,413 (5·7)	512,516 (11·8)	58,375 (3·7)	128,223 (9·6)
Itching	299,407 (7·1)	504,016 (13·8)	72,095 (2·9)	95,356 (4·8)	309,607 (5·9)	466,319 (10·7)	66,469 (4·2)	139,314 (10·5)
Any systemic reaction	2,444,362 (57·9)	2,752,592 (75·6)	1,088,296 (43·8)	1,226,561 (61·9)	2,972,931 (57·1)	3,237,621 (74·5)	600,498 (38·4)	781,299 (58·8)
Fatigue	1,624,531 (38·5)	2,221,361 (61·0)	643,206 (25·9)	904,556 (45·7)	1,941,979 (37·3)	2,588,541 (59·6)	353,226 (22·6)	569,758 (42·9)
Headache	1,349,155 (31·9)	1,906,337 (52·4)	460,786 (18·6)	690,138 (34·8)	1,595,091 (30·6)	2,226,046 (51·2)	236,380 (15·1)	397,675 (29·9)
Myalgia	954,469 (22·6)	1,724,474 (47·4)	450,562 (18·1)	726,994 (36·7)	1,219,190 (23·4)	2,085,722 (48·0)	204,146 (13·0)	392,448 (29·5)
Chills	451,583 (10·7)	1,202,364 (33·0)	172,283 (6·9)	459,577 (23·2)	542,285 (10·4)	1,426,710 (32·8)	89,261 (5·7)	253,475 (19·1)
Fever	446,178 (10·6)	1,182,201 (32·5)	187,713 (7·6)	478,912 (24·2)	565,804 (10·9)	1,449,504 (33·4)	76,288 (4·9)	230,073 (17·3)
Joint pain	444,630 (10·5)	1,023,525 (28·1)	188,846 (7·6)	400,963 (20·2)	539,196 (10·3)	1,214,624 (28·0)	102,810 (6·6)	226,303 (17·0)
Nausea	447,766 (10·6)	728,730 (20·0)	106,872 (4·3)	161,455 (8·2)	500,782 (9·6)	794,450 (18·3)	61,491 (3·9)	106,653 (8·0)
Diarrhea	272,890 (6·5)	313,252 (8·6)	106,079 (4·3)	101,107 (5·1)	323,773 (6·2)	352,077 (8·1)	59,803 (3·8)	66,967 (5·0)
Abdominal pain	179,210 (4·2)	283,422 (7·8)	50,991 (2·1)	71,115 (3·6)	203,575 (3·9)	316,165 (7·3)	29,936 (1·9)	42,942 (3·2)
Rash	65,498 (1·6)	79,092 (2·2)	19,193 (0·8)	19,735 (1·0)	70,985 (1·4)	79,913 (1·8)	14,781 (0·9)	19,965 (1·5)
Vomiting	43,998 (1·0)	75,650 (2·1)	10,936 (0·4)	14,915 (0·8)	49,483 (0·9)	81,733 (1·9)	6,227 (0·4)	9,994 (0·8)

Data are n (%).
*Reports of local and systemic reactions are not mutually exclusive.

***Supplemental Table* 5: Most common local and systemic reactions* to mRNA COVID-19 vaccine reported in v-safe, by dose and severity,† 0-7 days after vaccination with BNT162b2 vaccine**

		Dose 1				Dose 2			
	Day	All, n	Severe	Moderate	Mild	All, n	Severe	Moderate	Mild
Injection site pain	0	2,272,335	2,533 (0·1)	80,358 (3·5)	599,511 (26·4)	1,766,510	4,359 (0·2)	94,156 (5·3)	503,779 (28·5)
	1	2,545,271	18,827 (0·7)	334,755 (13·2)	1,289,293 (50·7)	2,027,330	48,810 (2·4)	453,726 (22·4)	885,434 (43·7)
	2	2,545,434	4,356 (0·2)	62,838 (2·5)	565,455 (22·2)	2,116,614	10,391 (0·5)	126,741 (6·0)	680,408 (32·1)
	3	2,507,344	2,119 (0·1)	26,602 (1·1)	216,785 (8·6)	2,067,908	3,332 (0·2)	43,037 (2·1)	336,222 (16·3)
	4	2,436,977	1,420 (0·1)	16,710 (0·7)	102,077 (4·2)	2,028,926	1,820 (0·1)	20,548 (1·0)	149,408 (7·4)
	5	2,332,032	1,138 (0·05)	11,965 (0·5)	60,401 (2·6)	2,000,426	1,272 (0·1)	12,719 (0·6)	73,486 (3·7)
	6	2,249,409	909 (0·04)	8,901 (0·4)	40,597 (1·8)	2,000,472	1,106 (0·1)	11,008 (0·6)	46,707 (2·3)
	7	2,198,611	768 (0·03)	7,783 (0·4)	32,967 (1·5)	2,067,201	1,557 (0·1)	14,159 (0·7)	44,322 (2·1)
Fatigue	0	2,272,335	7,280 (0·3)	79,232 (3·5)	193,192 (8·5)	1,766,510	11,293 (0·6)	98,802 (5·6)	185,051 (10·5)
	1	2,545,271	35,734 (1·4)	229,606 (9·0)	326,015 (12·8)	2,027,330	135,581 (6·7)	523,998 (25·8)	373,601 (18·4)
	2	2,545,434	16,936 (0·7)	114,562 (4·5)	211,046 (8·3)	2,116,614	39,668 (1·9)	217,269 (10·3)	308,126 (14·6)
	3	2,507,344	10,636 (0·4)	74,341 (3·0)	145,114 (5·8)	2,067,908	15,361 (0·7)	104,673 (5·1)	193,174 (9·3)
	4	2,436,977	8,275 (0·3)	58,170 (2·4)	109,266 (4·5)	2,028,926	10,280 (0·5)	69,886 (3·4)	133,603 (6·6)
	5	2,332,032	7,062 (0·3)	49,739 (2·1)	88,721 (3·8)	2,000,426	8,089 (0·4)	55,840 (2·8)	103,919 (5·2)
	6	2,249,409	6,428 (0·3)	44,044 (2·0)	76,633 (3·4)	2,000,472	7,200 (0·4)	48,388 (2·4)	86,907 (4·3)
	7	2,198,611	6,027 (0·3)	40,428 (1·8)	67,168 (3·1)	2,067,201	7,528 (0·4)	46,669 (2·3)	78,361 (3·8)
Headache	0	2,272,335	3,394 (0·1)	42,501 (1·9)	167,985 (7·4)	1,766,510	5,217 (0·3)	52,759 (3·0)	144,892 (8·2)
	1	2,545,271	20,011 (0·8)	129,629 (5·1)	265,970 (10·4)	2,027,330	82,393 (4·1)	333,605 (16·5)	381,368 (18·8)
	2	2,545,434	10,458 (0·4)	69,347 (2·7)	162,658 (6·4)	2,116,614	24,063 (1·1)	134,054 (6·3)	249,895 (11·8)
	3	2,507,344	6,670 (0·3)	46,850 (1·9)	110,115 (4·4)	2,067,908	10,356 (0·5)	68,461 (3·3)	148,990 (7·2)
	4	2,436,977	5,552 (0·2)	38,319 (1·6)	85,635 (3·5)	2,028,926	7,238 (0·4)	47,550 (2·3)	103,204 (5·1)
	5	2,332,032	4,911 (0·2)	34,379 (1·5)	72,831 (3·1)	2,000,426	6,154 (0·3)	40,322 (2·0)	82,191 (4·1)
	6	2,249,409	4,733 (0·2)	31,540 (1·4)	64,890 (2·9)	2,000,472	5,467 (0·3)	35,177 (1·8)	69,168 (3·5)
	7	2,198,611	4,381 (0·2)	29,475 (1·3)	58,752 (2·7)	2,067,201	5,372 (0·3)	34,057 (1·6)	63,628 (3·1)
Myalgia	0	2,272,335	1,999 (0·1)	29,601 (1·3)	96,095 (4·2)	1,766,510	4,001 (0·2)	38,960 (2·2)	75,790 (4·3)
	1	2,545,271	18,440 (0·7)	136,939 (5·4)	219,125 (8·6)	2,027,330	101,801 (5·0)	408,637 (20·2)	293,241 (14·5)
	2	2,545,434	7,441 (0·3)	56,954 (2·2)	112,788 (4·4)	2,116,614	23,521 (1·1)	140,700 (6·6)	209,074 (9·9)
	3	2,507,344	4,200 (0·2)	33,605 (1·3)	65,696 (2·6)	2,067,908	6,925 (0·3)	54,206 (2·6)	100,982 (4·9)
	4	2,436,977	3,255 (0·1)	25,814 (1·1)	46,369 (1·9)	2,028,926	4,146 (0·2)	32,786 (1·6)	60,489 (3·0)
	5	2,332,032	2,831 (0·1)	22,598 (1·0)	37,598 (1·6)	2,000,426	3,239 (0·2)	25,326 (1·3)	44,242 (2·2)
	6	2,249,409	2,543 (0·1)	20,904 (0·9)	33,016 (1·5)	2,000,472	2,973 (0·1)	22,422 (1·1)	36,522 (1·8)
	7	2,198,611	2,504 (0·1)	19,474 (0·9)	30,222 (1·4)	2,067,201	3,379 (0·2)	23,046 (1·1)	33,563 (1·6)
Chills	0	2,272,335	879 (0·04)	8,246 (0·4)	34,000 (1·5)	1,766,510	2,091 (0·1)	14,428 (0·8)	38,195 (2·2)
	1	2,545,271	8,558 (0·3)	45,518 (1·8)	78,033 (3·1)	2,027,330	62,884 (3·1)	210,579 (10·4)	207,218 (10·2)
	2	2,545,434	3,371 (0·1)	18,659 (0·7)	36,412 (1·4)	2,116,614	11,744 (0·6)	51,490 (2·4)	76,276 (3·6)
	3	2,507,344	1,462 (0·1)	9,241 (0·4)	19,569 (0·8)	2,067,908	2,582 (0·1)	13,423 (0·6)	25,421 (1·2)
	4	2,436,977	1,051 (0·04)	6,915 (0·3)	13,967 (0·6)	2,028,926	1,336 (0·1)	7,424 (0·4)	14,223 (0·7)
	5	2,332,032	863 (0·04)	5,531 (0·2)	11,284 (0·5)	2,000,426	955 (0·05)	5,423 (0·3)	10,583 (0·5)
	6	2,249,409	779 (0·03)	5,048 (0·2)	9,932 (0·4)	2,000,472	851 (0·04)	4,763 (0·2)	9,029 (0·5)
	7	2,198,611	752 (0·03)	4,645 (0·2)	8,889 (0·4)	2,067,201	1,222 (0·1)	5,515 (0·3)	9,039 (0·4)
Joint pain	0	2,272,335	1,069 (0·05)	11,375 (0·5)	24,689 (1·1)	1,766,510	2,396 (0·1)	18,677 (1·1)	25,699 (1·5)
	1	2,545,271	9,676 (0·4)	61,691 (2·4)	69,532 (2·7)	2,027,330	55,446 (2·7)	225,949 (11·2)	137,601 (6·8)
	2	2,545,434	4,608 (0·2)	31,238 (1·2)	44,072 (1·7)	2,116,614	14,386 (0·7)	80,490 (3·8)	90,461 (4·3)
	3	2,507,344	2,675 (0·1)	19,912 (0·8)	29,313 (1·2)	2,067,908	4,624 (0·2)	32,971 (1·6)	46,467 (2·2)
	4	2,436,977	2,165 (0·1)	15,923 (0·7)	22,386 (0·9)	2,028,926	2,882 (0·1)	20,861 (1·0)	29,916 (1·5)
	5	2,332,032	1,999 (0·1)	13,922 (0·6)	18,869 (0·8)	2,000,426	2,341 (0·1)	16,528 (0·8)	23,366 (1·2)
	6	2,249,409	1,773 (0·1)	13,018 (0·6)	16,874 (0·8)	2,000,472	2,138 (0·1)	15,046 (0·8)	19,649 (1·0)
	7	2,198,611	1,686 (0·1)	12,245 (0·6)	15,605 (0·7)	2,067,201	2,462 (0·1)	15,782 (0·8)	18,678 (0·9)

Data are n (%) unless otherwise stated.

*Top five reactions determined by reported frequency after second dose of both mRNA COVID-19 vaccines in v-safe, excluding fever because it was not rated mild/moderate/severe. Symptoms are not mutually exclusive. †Mild was defined as "noticeable symptoms but they aren't a problem", moderate was defined as "symptoms that limit normal activities, and severe symptoms", and severe symptoms "make normal daily activities difficult or impossible".

***Supplemental Table* 6: Most common local and systemic reactions* to mRNA COVID-19 vaccine reported in v-safe, by dose and severity,† 0-7 days after vaccination with mRNA-1273 vaccine**

		Dose 1				Dose 2			
	Day	All, n	Severe	Moderate	Mild	All, n	Severe	Moderate	Mild
Injection site pain	0	2,112,380	4,971 (0·2)	113,992 (5·4)	595,108 (28·2)	1,656,723	11,174 (0·7)	155,045 (9·4)	535,260 (32·3)
	1	2,424,231	49,225 (2·0)	512,076 (21·1)	1,214,808 (50·1)	1,937,029	131,379 (6·8)	735,284 (38·0)	713,164 (36·8)
	2	2,474,399	13,723 (0·6)	152,289 (6·2)	932,770 (37·7)	2,035,773	26,351 (1·3)	268,371 (13·2)	890,618 (43·7)
	3	2,459,431	4,778 (0·2)	47,625 (1·9)	370,863 (15·1)	1,993,354	6,469 (0·3)	78,053 (3·9)	525,158 (26·3)
	4	2,390,709	2,881 (0·1)	25,370 (1·1)	148,288 (6·2)	1,960,829	3,393 (0·2)	33,315 (1·7)	239,358 (12·2)
	5	2,285,185	2,087 (0·1)	17,079 (0·7)	77,452 (3·4)	1,939,300	2,493 (0·1)	19,783 (1·0)	109,652 (5·7)
	6	2,196,757	1,512 (0·1)	12,520 (0·6)	50,245 (2·3)	1,949,754	2,496 (0·1)	18,002 (0·9)	61,677 (3·2)
	7	2,157,101	1,595 (0·1)	13,301 (0·6)	45,669 (2·1)	2,019,370	3,443 (0·2)	22,300 (1·1)	49,711 (2·5)
Fatigue	0	2,112,380	7,221 (0·3)	74,690 (3·5)	170,757 (8·1)	1,656,723	13,938 (0·8)	104,699 (6·3)	173,749 (10·5)
	1	2,424,231	54,659 (2·3)	275,510 (11·4)	315,454 (13·0)	1,937,029	240,342 (12·4)	706,424 (36·5)	330,245 (17·0)
	2	2,474,399	23,894 (1·0)	143,988 (5·8)	228,253 (9·2)	2,035,773	60,995 (3·0)	292,539 (14·4)	337,135 (16·6)
	3	2,459,431	11,711 (0·5)	78,023 (3·2)	147,930 (6·0)	1,993,354	19,031 (1·0)	126,765 (6·4)	213,889 (10·7)
	4	2,390,709	8,430 (0·4)	57,900 (2·4)	107,495 (4·5)	1,960,829	11,806 (0·6)	80,578 (4·1)	146,796 (7·5)
	5	2,285,185	7,252 (0·3)	48,602 (2·1)	86,144 (3·8)	1,939,300	9,238 (0·5)	61,449 (3·2)	110,778 (5·7)
	6	2,196,757	6,486 (0·3)	43,439 (2·0)	73,380 (3·3)	1,949,754	8,051 (0·4)	52,161 (2·7)	91,466 (4·7)
	7	2,157,101	6,309 (0·3)	41,022 (1·9)	65,817 (3·1)	2,019,370	8,702 (0·4)	50,008 (2·5)	80,798 (4·0)
Headache	0	2,112,380	3,475 (0·2)	40,930 (1·9)	153,086 (7·2)	1,656,723	6,812 (0·4)	58,638 (3·5)	143,161 (8·6)
	1	2,424,231	33,272 (1·4)	164,590 (6·8)	273,687 (11·3)	1,937,029	154,933 (8·0)	497,171 (25·7)	411,266 (21·2)
	2	2,474,399	15,614 (0·6)	91,066 (3·7)	186,713 (7·5)	2,035,773	39,797 (2·0)	195,074 (9·6)	311,816 (15·3)
	3	2,459,431	7,782 (0·3)	50,386 (2·0)	114,639 (4·7)	1,993,354	13,752 (0·7)	85,153 (4·3)	177,740 (8·9)
	4	2,390,709	5,818 (0·2)	38,567 (1·6)	86,140 (3·6)	1,960,829	8,903 (0·5)	56,974 (2·9)	120,773 (6·2)
	5	2,285,185	5,316 (0·2)	34,399 (1·5)	72,082 (3·2)	1,939,300	7,429 (0·4)	46,575 (2·4)	93,323 (4·8)
	6	2,196,757	4,732 (0·2)	31,891 (1·5)	63,922 (2·9)	1,949,754	6,576 (0·3)	40,092 (2·1)	77,243 (4·0)
	7	2,157,101	4,585 (0·2)	30,705 (1·4)	59,396 (2·8)	2,019,370	6,705 (0·3)	37,634 (1·9)	68,333 (3·4)
Myalgia	0	2,112,380	3,011 (0·1)	35,724 (1·7)	95,048 (4·5)	1,656,723	7,757 (0·5)	54,566 (3·3)	82,160 (5·0)
	1	2,424,231	38,950 (1·6)	197,871 (8·2)	226,700 (9·4)	1,937,029	198,988 (10·3)	610,812 (31·5)	292,422 (15·1)
	2	2,474,399	13,531 (0·5)	86,102 (3·5)	150,896 (6·1)	2,035,773	39,990 (2·0)	203,682 (10·0)	251,594 (12·4)
	3	2,459,431	5,264 (0·2)	38,086 (1·6)	74,647 (3·0)	1,993,354	9,143 (0·5)	65,042 (3·3)	114,111 (5·7)
	4	2,390,709	3,627 (0·2)	26,656 (1·1)	47,845 (2·0)	1,960,829	5,041 (0·3)	36,805 (1·9)	65,072 (3·3)
	5	2,285,185	2,989 (0·1)	22,955 (1·0)	37,471 (1·6)	1,939,300	3,796 (0·2)	27,445 (1·4)	45,771 (2·4)
	6	2,196,757	2,667 (0·1)	21,040 (1·0)	33,060 (1·5)	1,949,754	3,592 (0·2)	24,073 (1·2)	37,073 (1·9)
	7	2,157,101	2,731 (0·1)	20,799 (1·0)	31,659 (1·5)	2,019,370	4,415 (0·2)	25,026 (1·2)	33,652 (1·7)
Chills	0	2,112,380	1,395 (0·1)	10,125 (0·5)	35,178 (1·7)	1,656,723	4,685 (0·3)	23,091 (1·4)	45,194 (2·7)
	1	2,424,231	23,553 (1·0)	84,997 (3·5)	101,682 (4·2)	1,937,029	137,685 (7·1)	402,336 (20·8)	291,400 (15·0)
	2	2,474,399	7,601 (0·3)	33,238 (1·3)	52,246 (2·1)	2,035,773	23,939 (1·2)	90,569 (4·5)	113,067 (5·6)
	3	2,459,431	2,358 (0·1)	11,569 (0·5)	21,723 (0·9)	1,993,354	4,164 (0·2)	18,349 (0·9)	31,919 (1·6)
	4	2,390,709	1,374 (0·1)	7,398 (0·3)	14,470 (0·6)	1,960,829	2,100 (0·1)	9,271 (0·5)	16,372 (0·8)
	5	2,285,185	1,022 (0·04)	5,986 (0·3)	11,645 (0·5)	1,939,300	1,547 (0·1)	6,581 (0·3)	11,521 (0·6)
	6	2,196,757	838 (0·04)	5,231 (0·2)	10,167 (0·5)	1,949,754	1,500 (0·1)	6,158 (0·3)	9,622 (0·5)
	7	2,157,101	891 (0·04)	5,094 (0·2)	9,317 (0·4)	2,019,370	2,376 (0·1)	7,405 (0·4)	9,915 (0·5)
Joint pain	0	2,112,380	1,448 (0·1)	13,228 (0·6)	24,666 (1·2)	1,656,723	4,387 (0·3)	26,725 (1·6)	29,864 (1·8)
	1	2,424,231	20,384 (0·8)	92,808 (3·8)	79,034 (3·3)	1,937,029	115,152 (5·9)	375,004 (19·4)	163,832 (8·5)
	2	2,474,399	8,102 (0·3)	46,926 (1·9)	57,163 (2·3)	2,035,773	25,592 (1·3)	123,386 (6·1)	116,891 (5·7)
	3	2,459,431	3,509 (0·1)	23,226 (0·9)	33,006 (1·3)	1,993,354	6,277 (0·3)	41,550 (2·1)	55,855 (2·8)
	4	2,390,709	2,400 (0·1)	16,852 (0·7)	23,553 (1·0)	1,960,829	3,601 (0·2)	24,570 (1·3)	34,118 (1·7)
	5	2,285,185	2,010 (0·1)	14,447 (0·6)	19,143 (0·8)	1,939,300	2,831 (0·1)	18,829 (1·0)	25,132 (1·3)
	6	2,196,757	1,821 (0·1)	13,250 (0·6)	16,821 (0·8)	1,949,754	2,582 (0·1)	16,587 (0·9)	21,152 (1·1)
	7	2,157,101	1,874 (0·1)	13,157 (0·6)	16,230 (0·8)	2,019,370	3,240 (0·2)	17,390 (0·9)	19,800 (1·0)

Data are n (%) unless otherwise stated.
*Top five reactions determined by reported frequency after second dose of both mRNA COVID-19 vaccines in v-safe, excluding fever because it was not rated mild/moderate/severe. Symptoms are not mutually exclusive. †Mild was defined as "noticeable symptoms but they aren't a problem", moderate was defined as "symptoms that limit normal activities, and severe symptoms", and severe symptoms "make normal daily activities difficult or impossible".

***Supplemental Table* 7: Reported health impact* 0-7 days after vaccination by mRNA COVID-19 vaccine manufacturer, dose, and sex reported in v-safe—December 14, 2020–June 14, 2021**

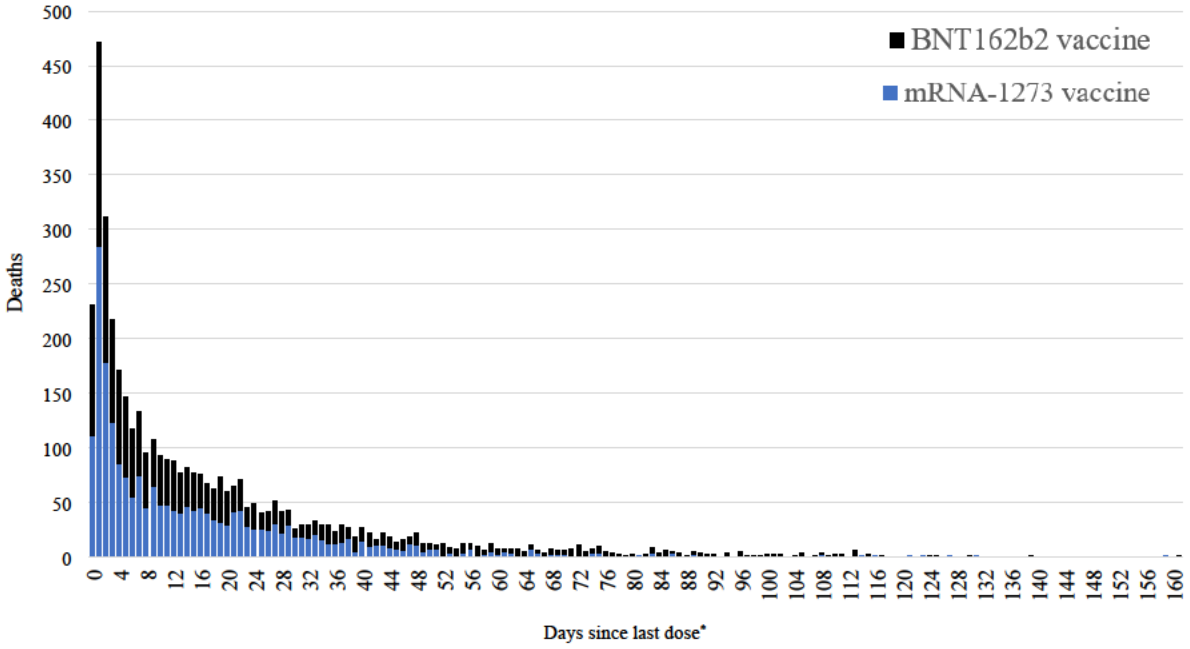
	BNT162b2 vaccine						mRNA-1273 vaccine					
	Dose 1 (n=3,455,778)			Dose 2 (n=2,920,526)			Dose 1 (n=3,319,737)			Dose 2 (n=2,753,894)		
Sex/day	Unable to do normal activity	Unable to work	Reported medical care	Unable to do normal activity	Unable to work	Reported medical care	Unable to do normal activity	Unable to work	Reported medical care	Unable to do normal activity	Unable to work	Reported medical care
Female												
Day 0	29,039 (2·1)	11,480 (0·8)	2,486 (0·2)	38,593 (3·5)	18,411 (1·7)	1,567 (0·1)	30,334 (2·4)	10,888 (0·8)	2,099 (0·2)	48,485 (4·6)	22,274 (2·1)	1,471 (0·1)
Day 1	105,123 (6·6)	41,398 (2·6)	2,810 (0·2)	324,559 (24·9)	178,887 (13·7)	3,276 (0·3)	151,710 (10·0)	60,161 (4·0)	3,370 (0·2)	522,192 (41·4)	296,178 (23·5)	4,470 (0·4)
Day 2	53,456 (3·4)	22,016 (1·4)	2,912 (0·2)	121,302 (8·9)	67,628 (5·0)	3,335 (0·2)	74,193 (4·8)	31,936 (2·1)	3,432 (0·2)	188,421 (14·2)	107,761 (8·1)	3,767 (0·3)
Day 3	35,994 (2·3)	13,399 (0·9)	3,387 (0·2)	55,917 (4·2)	24,334 (1·8)	3,695 (0·3)	40,305 (2·6)	15,496 (1·0)	3,345 (0·2)	73,844 (5·7)	32,292 (2·5)	4,052 (0·3)
Day 4	28,877 (1·9)	10,799 (0·7)	3,572 (0·2)	36,450 (2·8)	14,154 (1·1)	3,377 (0·3)	29,880 (2·0)	10,773 (0·7)	3,351 (0·2)	43,833 (3·4)	16,702 (1·3)	3,541 (0·3)
Day 5	24,765 (1·7)	9,468 (0·6)	3,548 (0·2)	29,069 (2·3)	10,747 (0·8)	3,125 (0·2)	25,056 (1·7)	9,512 (0·7)	3,467 (0·2)	32,958 (2·6)	12,195 (1·0)	3,065 (0·2)
Day 6	22,401 (1·6)	9,187 (0·7)	3,621 (0·3)	25,167 (2·0)	9,669 (0·8)	3,157 (0·2)	22,502 (1·6)	9,271 (0·7)	3,733 (0·3)	28,146 (2·2)	10,840 (0·9)	3,133 (0·2)
Day 7	20,820 (1·5)	8,801 (0·6)	3,811 (0·3)	24,955 (1·9)	10,060 (0·8)	3,419 (0·3)	21,804 (1·6)	9,242 (0·7)	4,483 (0·3)	28,538 (2·2)	12,066 (0·9)	3,272 (0·3)
Male												
Day 0	8,905 (1·0)	5,711 (0·7)	569 (0·1)	11,137 (1·7)	8,208 (1·3)	380 (0·1)	8,954 (1·1)	5,339 (0·7)	479 (0·1)	13,450 (2·3)	8,955 (1·5)	337 (0·1)
Day 1	30,240 (3·2)	16,781 (1·8)	656 (0·1)	93,820 (13·3)	66,375 (9·4)	820 (0·1)	46,535 (5·3)	24,313 (2·8)	955 (0·1)	167,957 (25·6)	110,868 (16·9)	1,104 (0·2)
Day 2	13,698 (1·5)	7,846 (0·8)	767 (0·1)	29,528 (4·0)	21,766 (3·0)	768 (0·1)	21,696 (2·4)	12,307 (1·4)	836 (0·1)	47,601 (6·9)	33,333 (4·9)	785 (0·1)
Day 3	8,925 (1·0)	4,669 (0·5)	827 (0·1)	12,163 (1·7)	7,101 (1·0)	788 (0·1)	10,625 (1·2)	5,218 (0·6)	865 (0·1)	15,542 (2·3)	8,550 (1·3)	784 (0·1)
Day 4	7,267 (0·8)	3,667 (0·4)	967 (0·1)	7,978 (1·1)	4,250 (0·6)	843 (0·1)	7,670 (0·9)	3,801 (0·4)	867 (0·1)	9,428 (1·4)	4,613 (0·7)	754 (0·1)
Day 5	6,180 (0·7)	3,207 (0·4)	981 (0·1)	6,319 (0·9)	3,224 (0·5)	901 (0·1)	6,516 (0·8)	3,376 (0·4)	932 (0·1)	7,156 (1·1)	3,425 (0·5)	785 (0·1)
Day 6	5,696 (0·7)	3,019 (0·4)	1,022 (0·1)	5,790 (0·8)	2,902 (0·4)	868 (0·1)	5,829 (0·7)	3,107 (0·4)	1,035 (0·1)	6,433 (1·0)	3,146 (0·5)	793 (0·1)
Day 7	5,324 (0·7)	2,937 (0·4)	1,050 (0·1)	5,873 (0·8)	3,147 (0·4)	975 (0·1)	5,443 (0·7)	3,047 (0·4)	1,094 (0·1)	6,651 (0·9)	3,547 (0·5)	886 (0·1)

Data are n (%)†.

*Reports of health impacts are not mutually exclusive.

†Percent corresponds to number of respondents by sex and day.

***Supplemental Figure* 1: Number of reports of death per day following vaccination, by manufacturer, to Vaccine Adverse Event Reporting System (VAERS)—December 14, 2020–June 14, 2021**

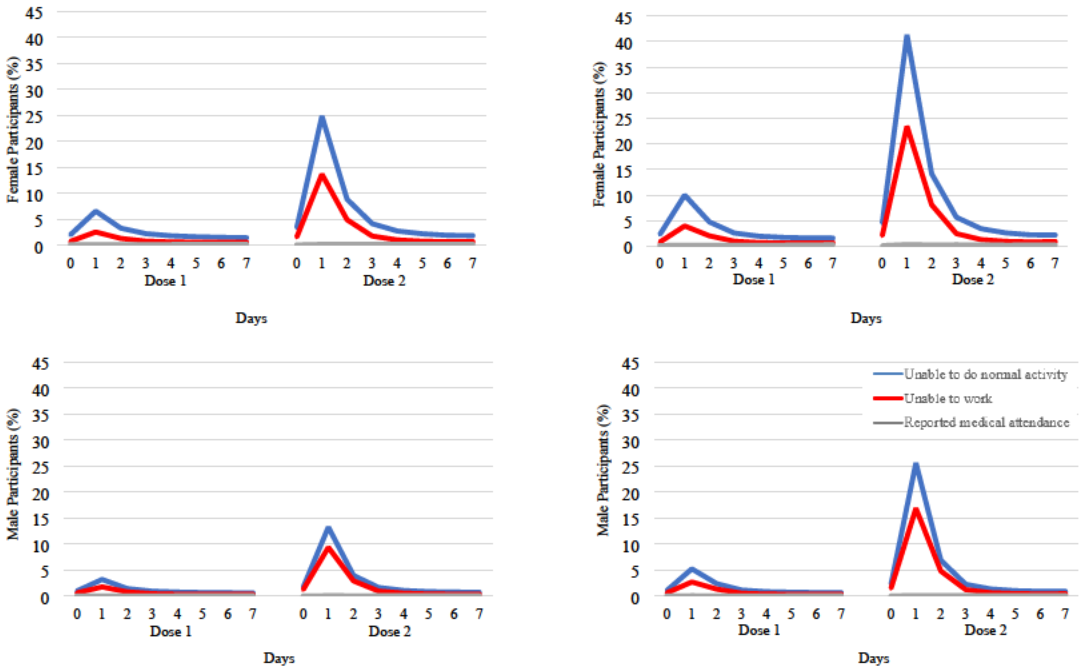


*x-axis reports through 161 days since last dose.

Commented [BR(35]: Editorial suggestion:
This wording sounds awkward and confusing. Please revise for clarity. Consider something like "Number of reports per day of onset" .....
Office of Science
2021-08-11 11:45:00

Commented [RH(36R35]: Thanks. done
Rosenblum, Hannah (CDC)
2021-08-31 12:09:00

***Supplemental Figure* 2: Reported health impact 0-7 days after mRNA COVID-19 vaccination by manufacturer, type of impact, and sex reported in v-safe—December 14, 2020–June 14, 2021**



Top left: Female participants reporting health impact after receiving BNT162b2 vaccine. *Top right:* Female participants reporting health impact after receiving mRNA-1273 vaccine. *Bottom left:* Male participants reporting health impact after receiving BNT162b2 vaccine. *Bottom right*: Male participants reporting health impact after receiving mRNA-1273 vaccine.

From: "Miller, Elaine R. (CDC/DDID/NCEZID/DHQP)" <[redacted]>
To: "Su, John (CDC/DDID/NCEZID/DHQP)" <[redacted]>
Subject: RE: Distributing weekly data mining raw output
Date: Wed, 11 Aug 2021 20:46:08 +0000
Importance: Normal

Thanks-Can I share with Jonathan and Pedro since they are responding to inquiries?

From: Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Wednesday, August 11, 2021 4:42 PM
To: Miller, Elaine R. (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: RE: Distributing weekly data mining raw output

Hi Elaine,

Please see enclosed, and (per FDA's request to keep these data closer hold) don't share. Thanks!

- John

From: Miller, Elaine R. (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Wednesday, August 11, 2021 4:33 PM
To: Menschik, David (FDA/CBER) <[redacted]>; Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: RE: Distributing weekly data mining raw output

Thanks David.

John-please share this data with me.

From: Menschik, David <[redacted]>
Sent: Wednesday, August 11, 2021 4:26 PM
To: Miller, Elaine R. (CDC/DDID/NCEZID/DHQP) <[redacted]>
Cc: Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: Distributing weekly data mining raw output

Hi Elaine,

I saw your email about expanding the data mining raw output distribution list. Our plan is actually to limit its distribution, largely for data security reasons.
Moving forward I'll be forwarding this output to John Su as your group's POC and he has discretion to share with other team members (e.g., upon request by you/others in the group) as needed. Sorry for any inconvenience.

Best,
David

From: "Su, John (CDC/DDID/NCEZID/DHQP)" <[redacted]>
To: "Miller, Elaine R. (CDC/DDID/NCEZID/DHQP)" <[redacted]>
Subject: RE: Distributing weekly data mining raw output
Date: Wed, 11 Aug 2021 20:41:44 +0000
Importance: Normal
Attachments: USST_20210806.xls

Hi Elaine,

Please see enclosed, and (per FDA's request to keep these data closer hold) don't share. Thanks!

- John

From: Miller, Elaine R. (CDC/DDID/NCEZID/DHQP) <[redacted]>
Sent: Wednesday, August 11, 2021 4:33 PM
To: Menschik, David (FDA/CBER) <[redacted]>; Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: RE: Distributing weekly data mining raw output

Thanks David.

John-please share this data with me.

From: Menschik, David <[redacted]>
Sent: Wednesday, August 11, 2021 4:26 PM
To: Miller, Elaine R. (CDC/DDID/NCEZID/DHQP) <[redacted]>
Cc: Su, John (CDC/DDID/NCEZID/DHQP) <[redacted]>
Subject: Distributing weekly data mining raw output

Hi Elaine,

I saw your email about expanding the data mining raw output distribution list. Our plan is actually to limit its distribution, largely for data security reasons.
Moving forward I'll be forwarding this output to John Su as your group's POC and he has discretion to share with other team members (e.g., upon request by you/others in the group) as needed. Sorry for any inconvenience.

Best,
David

From: "Menschik, David" <[redacted]>
To: "Zinderman, Craig E" <[redacted]>
Subject: suggested edits as discussed...
Date: Fri, 07 May 2021 16:32:56 -0000
Importance: Normal
Attachments: Good_Morning_Ana.docx

...attached...

Good AfternoonMorning 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 that you to please hold on stop 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 have not contributed to our alreadywe already have a robust process for continuous monitoring ofreviewing incoming and aggregated VAERS (and other vaccine safety) data. I will describe a little bit below: including:

-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 sSerious reports coming into VAERS daily until meeting specified milestone and then ; for the first couple months with the mRNA vaccines, and still currently for JnJ, FDA MOs review each serious report as it is processed into VAERS. For the older mRNA vaccines, MOs now conduct weekly review of aggregated PT counts 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.. Some alerts that you have sent, like for AMI, have already temporarily signaled in population-based surveillance, which has already triggered a more thorough review of VAERS data.

Also as you knowFurther, 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 and varying the stratifications or other parameters 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, and predictable, and road-tested. Results from Aadjusting parameters thato raise or lower sensitivity of the alerts as the vaccination campaign is underway could lead to some artificial creation of alerts and an apparent, but not real, sudden change in safety results. 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, ~~for all of the above reasons,~~ in the above context, we have found ~~that examining the analyses you have been sending has largely reflected events otherwise under evaluation~~no indication for action based on your findings, which have been consuming a lot of resources at a time when resources are stretched across a preexisting robust pharmacovigilance infrastructure. ~~We haven't seen a proportionate increase in efficiency or yield, given the robust screening and scrutiny of VAERS data already in place as described above. We'd hate for you to be wasting your time and efforts, so we thought we should suggest that it might be a better use of resources for you to refocus your data mining efforts on other product types. Perhaps there are CDER products and drug related data in your Center that could benefit more from your continued data mining explorations and analyses.~~ Thanks much for your time and ~~attention~~understanding, and~~;~~ sorry for the long email.

Kind regards,~~Thanks~~,

Craig Zinderman, MD, MPH
Associate Director for Medical Policy
Office of Biostatistics and Epidemiology
FDA/Center for Biologics Evaluation and Research
[redacted]

From: "Menschik, David" <[redacted]>
To: "Baer, Bethany" <[redacted]>, "Zinderman, Craig E" <[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 15:06:11 -0000
Importance: Normal
Inline-Images: image001.png

Sorry – didn't see that – will reschedule

From: Baer, Bethany <[redacted]>
Sent: Thursday, April 15, 2021 11:05 AM
To: Menschik, David <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

I am on leave tomorrow so can't attend. I am free most of Monday and Tuesday except for a few meetings.
Bethany

From: Menschik, David <[redacted]>
Sent: Thursday, April 15, 2021 10:41 AM
To: Baer, Bethany <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

Sending invite for tomorrow at noon

From: Baer, Bethany <[redacted]>
Sent: Thursday, April 15, 2021 10:07 AM
To: Menschik, David <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

Sounds good. Happy to meet and discuss anytime open on my calendar.
Bethany

From: Menschik, David <[redacted]>
Sent: Thursday, April 15, 2021 9:31 AM
To: Baer, Bethany <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Important analysis by DuMouchel -> Peculiarities of disproportionality statistics when the product of interest is in almost all of the reports

Before we potentially reach out to Ana, we should meet internally – many considerations not suited to email…

From: Niu, Manette <[redacted]>
Sent: Thursday, April 15, 2021 9:27 AM
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

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]gov>
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]

[redacted]

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: "Menschik, David" <[redacted]>
To: "Zinderman, Craig E" <[redacted]>
Bcc: "Menschik, David" <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts
Date: Wed, 08 Sep 2021 10:49:16 -0000
Importance: Normal
Attachments: Draft_proposed_respponse_on_age_stratification.docx
Inline-Images: image001.png; image002.png

Hi Craig, I made some edits to draft proposed response (per attached) in advance of discussing…

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]

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 <[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
[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
[redacted]

U.S. FOOD & DRUG ADMINISTRATION

Draft proposed response:

Hi Ana,

I certainly understand where you are coming from.

While we must limit access to VAERS Empirica data mining data to those within the CBER/OBE/DE chain of command for reasons including CBER contractual/budgetary policies and data security, I can in general speak to the understandable concern that you raise regarding stratifying by year.

We have previously thought about and discussed this issue, recognizing that the vast-majority of reports received in VAERS this year have involved COVID-19 vaccines which could drive PT-vaccine disproportionately scores towards the null by contributing substantially to the comparator group, particularly if there is a class-effect (e.g., if all COVID vaccines are associated with the same adverse event). While you have laid out reasonable arguments (which had generally occurred to us) why stratifying VAERS data by year raises new limitations in interpreting data output using existing methods, there are sound reasons for retaining adjustment by year.

From an epidemiology standpoint, exposures and health outcomes (not to mention public perceptions, behaviors, practices, etc. whether stimulated or not) can vary dramatically from one year to the next (independent of vaccinations) and such disparities by year can increase with increasing number of years. For example, during the 'COVID era,' circulating SARS-CoV-2 disease can drive a substantial increase in reported specific AEs, independently of AEs that may be associated with vaccinations; these AEs would most likely be over-represented in individual COVID vaccine-AE disproportionality scores if the comparison group were expanded to include reports from increasing time periods prior to the 'COVID-era.'

On a related note, MedDRA terms are continuously being updated and can regularly have substantial updates introducing new AE terms not available when reports were coded during prior time periods. For instance, this past week (under version 24.1 release), a new preferred term (PT), "multisystem inflammatory syndrome" was added to MedDRA. Without controlling for time (e.g., year) of vaccination, there would likely be inflated disproportionality for newer MedDRA AE terms in association with COVID vaccines since an expanded comparison group would include substantially more VAERS reports that have no chance of having such newer MedDRA terms due to being coded prior to the availability of such a term.

We will plan to discuss internally within CBER/OBE/DE and with Commonwealth options and associated feasibilities, impacts, etc. for potential approaches to addressing the age-stratification issue. Any further discussion on VAERS data mining methods/findings outside my chain of command (for reasons including data security) will have to be offline and in general terms, as well as without reference to any specific VAERS vaccine-PT pair outputs.

Thank you for your understanding,

David

From: "Menschik, David" <[redacted]>
To: "Zinderman, Craig E" <[redacted]>
Bcc: "Menschik, David" <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts
Date: Mon, 06 Sep 2021 13:03:27 -0000
Importance: Normal
Attachments: Draft_proposed_respponse_on_age_stratification.docx
Inline-Images: image001.png; image002.png

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]

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 <[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
[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
[redacted]

U.S. FOOD & DRUG ADMINISTRATION

Draft proposed response:

Hi Ana,

I certainly understand where you are coming from.

While we must limit access to VAERS Empirica data mining data to those within the CBER/OBE/DE chain of command for reasons including contractual financial as well as data security reasons, I can in general speak to the concern that you raise regarding stratifying by year.

We have previously thought about and discussed this issue, recognizing that the vast-majority of reports in VAERS this year have involved COVID-19 vaccines which could drive PT-vaccine disproportionately scores towards the null by contributing substantially to the comparator group, particularly if there is a class-effect (e.g., if all COVID vaccines are associated with the same adverse event). While you have laid out reasonable arguments (which had generally occurred to us) why stratifying VAERS data by year raises new limitations in interpreting data output using existing methods, there are sound reasons for retaining adjustment by year including:

From an epidemiology standpoint, exposures and health outcomes (not to mention public perceptions, behaviors, practices, etc. whether stimulated or not) can vary dramatically from one year to the next (independent of vaccinations) and such disparities by year can increase with increasing number of years. For example, during the 'COVID era,' circulating SARS-CoV-2 disease can drive a substantial increase in reported specific AEs, independently of AEs that may be associated with vaccinations; these AEs would most likely be over-represented in individual COVID vaccine-AE disproportionality scores if the comparison group were expanded to include reports from increasing time periods prior to the 'COVID-era.'

On a related note, MedDRA terms are continuously being updated and can regularly have substantial updates introducing new AE terms not available when reports were coded during prior time periods. For instance, this past week (under version 24.1 release), a new preferred term (PT), "multisystem inflammatory syndrome" was added to MedDRA. Without controlling for time (e.g., year) of vaccination, there would likely be inflated disproportionality for newer MedDRA AE terms in association with COVID vaccines since an expanded comparison group would include substantially more VAERS reports that have no chance of having such newer MedDRA terms due to being coded prior to the availability of a given term.

We will plan to discuss internally within DE and with Commonwealth options and associated feasibilities, impacts, etc. for potential approaches to addressing the age-stratification issue (e.g., exploring adjustment stratifications of more than one year). Any further discussion on VAERS data mining methods outside my chain of command (for reasons including data security) will have to be offline and in general terms, without reference to any specific VAERS vaccine-PT pairs.

Thank you for your understanding,

David

From: "Menschik, David" <[redacted]>
To: "Baer, Bethany" <[redacted]>
Subject: RE: [External] : CORRECTION: Following up on today's discussion
Date: Thu, 26 Sep 2024 17:55:58 -0000
Importance: Normal
Inline-Images: image001.jpg

Thanks for sharing – I also explored a bit and found that in general the ER05 appears way more sensitive in that its scores are generally higher than corresponding EB05 scores when sampling different PTs – this resulted in a lot more 'statistical signals' using their SDR threshold in the sandbox. We can discuss more offline though agree for now our focus should be on preserving current functions, features, outputs, etc.
Wishing you fantastic travels!
David

From: Baer, Bethany <[redacted]>
Sent: Thursday, September 26, 2024 1:46 PM
To: Menschik, David <[redacted]>
Subject: RE: [External] : CORRECTION: Following up on today's discussion

Hi David,
I spent some time yesterday and today exploring the ER05. I have previously read the Harpaz/Szarfman publication and have now read the DuMouchel white paper. I feel I understand the concept and big picture, but then I was completely out of my league on p. 6 of the white paper when the detailed methodology part started. I added the ER05 column to my signals view and saw that for Gardasil the ER05 is significantly higher (numbers in teens-twenties) than the EB05 for the same PTs and Ns (EB05s were 3-5). Then for 1 PT (psychogenic pseudosyncope), the ER05 and EB05 were the exact same. The Pfizer covid bivalent had numbers that were overall closer to each other for the ER05 and EB05 of many PTs, with the ER05 frequently being around 0.5-1 higher than the EB05.

So, I understand the theory behind masking and trying to adjust for it, but I feel that comprehending the details of the approach and, importantly, which approach is "better," is beyond my training and experience. I think someone with more data mining expertise would have to be involved in that decision. I don't think I have more to add regarding the ER05. I also don't have any specific questions for the Oracle team for Friday. From my question last week, it didn't sound like that is the group to have an in-depth discussion regarding ER05. I really appreciate the documents they forwarded to us.

Let me know if you have any questions regarding this. I wanted to get you my thoughts before I head out on leave next week.
Thanks,
Bethany

From: Menschik, David <[redacted]>
Sent: Tuesday, September 24, 2024 2:32 PM
To: Baer, Bethany <[redacted]>; Panchanathan, Sarada <[redacted]>; Thompson, Deborah <[redacted]>
Cc: Zinderman, Craig <[redacted]>
Subject: RE: [External] : CORRECTION: Following up on today's discussion

Thanks all for the prompt feedback. I checked WONDER and found comparable counts (e.g., n=49 for 'drug ineffective and Pfizer bivalent) to what we've observed in new Empirica so New Empirica's not picking up the extra cases is not its fault and it does indicate to me that the difference is related to differences between our internal data set and the public data set. Based on this, I'm ok to move on beyond this discrepancy issue. Agree with not meeting and also please advise if you

have any items to discuss on Friday. I don't have any new items to discuss with Oracle and if none of you do by Thursday afternoon, I can propose to cancel the meeting.
Thanks,
David

From: Baer, Bethany <[redacted]>
Sent: Tuesday, September 24, 2024 1:07 PM
To: Panchanathan, Sarada <[redacted]>; Thompson, Deborah <[redacted]>; Menschik, David <[redacted]>
Subject: RE: [External] : CORRECTION: Following up on today's discussion

I don't need to meet either. My thought about what they said about the brain fog cases was just the question of would those different procedures be aligned if we went with New Empirica and they had our in-house data? I don't know the procedure for when a new PT is added to Meddra and how old reports are handled – but I understand if they say that is the difference between the processes. The issue would then only show up for relatively new PTs with a lot of reports but also some older reports in the system not previously coded– like brain fog.
-Bethany

From: Panchanathan, Sarada <[redacted]>
Sent: Tuesday, September 24, 2024 12:12 PM
To: Thompson, Deborah <[redacted]>; Menschik, David <[redacted]>; Baer, Bethany <[redacted]>
Subject: RE: [External] : CORRECTION: Following up on today's discussion

I don't need to meet, but also happy to meet if needed.

Warm regards,
Soumya

From: Thompson, Deborah <[redacted]>
Sent: Tuesday, September 24, 2024 11:47 AM
To: Menschik, David <[redacted]>; Baer, Bethany <[redacted]>; Panchanathan, Sarada <[redacted]>
Subject: RE: [External] : CORRECTION: Following up on today's discussion

Thanks, David. Agree. I don't think we need a huddle before meeting with Oracle on Friday, but am happy to meet if preferred by others.

Thanks,

Deb

From: Menschik, David <[redacted]>
Sent: Tuesday, September 24, 2024 11:19 AM
To: Thompson, Deborah <[redacted]>; Baer, Bethany <[redacted]>; Panchanathan, Sarada <[redacted]>
Subject: FW: [External] : CORRECTION: Following up on today's discussion

FYI - don't think it makes sense to share thoughts with Oracle over email though please advise if you have any thoughts and would like to huddle before our meeting with Oracle on Friday…

From: Menschik, David
Sent: Tuesday, September 24, 2024 11:17 AM

To: Robert Weber [redacted]>; Philip Sheridan <[redacted]>
Cc: Alimchandani, Meghna <[redacted]>
Subject: RE: [External] : CORRECTION: Following up on today's discussion

Thank you Robert for sharing these helpful observations and thoughts. We'll also plan to take a closer look and looking forward to regrouping on Friday.

Best regards,
David

From: Robert Weber <[redacted]>
Sent: Monday, September 23, 2024 8:57 AM
To: Menschik, David <[redacted]>; Philip Sheridan <[redacted]>
Cc: Alimchandani, Meghna <[redacted]>
Subject: Re: [External] : CORRECTION: Following up on today's discussion

Thank you, David.

After a first pass, I noticed the following points – sharing observations early in case you want to add further details at this stage. We will keep investigating this more thoroughly.

#1: The list of cases shared shows up as "COVID19 (PFIZER-BIONTECH)", i.e. the monovalent product, in our database. When looking at a few examples, I noticed that the narrative was indicating uncertainty about the exact product/batch no – have these cases potentially ben re-coded in your inhouse database due some other information that might not be available in the public datasets? I also found cases clearly coded as bivalent Pfizer missing in the list you provided.
With the numbers currently coded as bivalent in our database, the disproportionality scores for bivalent Pfizer are low due to the large number of drug ineffective cases attributed to other COVID vaccines, in particular monovalent Pfizer.

#2: Here, the number of Brain fog cases (25) appears to be limited to those reported after the PT being added to MedDRA, while your list also has older cases mentioning Brain fog in the narrative. My suspicion is that the discrepancy is due to data preparation differences between your inhouse database and the public datasets we are using for the VAERS data loaded into the sandbox. We are looking into this further.

Best regards,
Robert

From: Menschik, David <[redacted]>
Date: Friday, 20. September 2024 at 19:39
To: Philip Sheridan <[redacted]>, Robert Weber <[redacted]>
Cc: Alimchandani, Meghna <[redacted]>
Subject: [External] : CORRECTION: Following up on today's discussion

Apologies: I had an error in my earlier email, the sandbox sample size in example two was 25.

Sorry for my mistake,
David

From: Menschik, David
Sent: Friday, September 20, 2024 1:21 PM
To: Philip Sheridan <[redacted]>; Robert Weber <[redacted]>

Cc: Alimchandani, Meghna <[redacted]>
Subject: Following up on today's discussion

Hi Robert and Phil,

Thanks for a helpful workshop session earlier today. As discussed, attached please find VAERS IDs with received date through 1/31/2024 for the two examples we discussed with corresponding sandbox output as follows:

1. Pfizer Bivalent and PT 'Drug ineffective' using "VAERS 202401: US Only" run (N=49, EB05=0.060)
   Note: expected much higher EB05 (>2)
2. Gardasil and PT 'Brain fog" using "VAERS 202401: All" run (N=~~211~~ 25, EB05=7.78)
   Note: expected much lower EB05 (close to 2)

As you can observe, our counts are substantially higher for both and wondering why so many VAERS IDs are missing in the sandbox.

Thanks and have a great weekend,
David

David Menschik, MD, MPH
Associate Director for Surveillance Informatics
Division of Pharmacovigilance/Office of Biostatistics and Pharmacovigilance
Center for Biologics Evaluation and Research/FDA
[redacted]

From: "Menschik, David" <[redacted]>
To: "Zinderman, Craig E" <[redacted]>
Subject: RE: suggested edits as discussed...
Date: Fri, 07 May 2021 18:58:47 -0000
Importance: Normal

---

thanks

From: Zinderman, Craig E <[redacted]>
Sent: Friday, May 07, 2021 2:58 PM
To: Menschik, David <[redacted]>; Nair, Narayan <[redacted]>
Subject: RE: suggested edits as discussed...

Yup, agree. I will send a new version to NN.

Thanks,
Craig

From: Menschik, David <[redacted]>
Sent: Friday, May 07, 2021 2:56 PM
To: Zinderman, Craig E <[redacted]>
Subject: RE: suggested edits as discussed...

Yes I like that, followed by “Thanks much for your understanding…”

From: Zinderman, Craig E <[redacted]>
Sent: Friday, May 07, 2021 2:54 PM
To: Menschik, David <[redacted]>
Subject: RE: suggested edits as discussed...

Got it; understood. Maybe we should just restate the ask at the end of the last paragraph:
“So, we are asking that you please hold on creating and sending data mining results for COVID-19 vaccine AE data.“

Thanks,
Craig

From: Menschik, David <[redacted]>
Sent: Friday, May 07, 2021 2:51 PM
To: Zinderman, Craig E <[redacted]>
Subject: RE: suggested edits as discussed...

Thanks – I removed the language in the last paragraph since I didn’t want it to be misconstrued as patronizing (obviously far from your intention…)

From: Zinderman, Craig E <[redacted]>
Sent: Friday, May 07, 2021 2:49 PM

To: Menschik, David <[redacted]>; Nair, Narayan <[redacted]>
Subject: RE: suggested edits as discussed...

Narayan:
I drafted, and David edited, the attached message to Ana as discussed. Please feel free to edits as you see fit. It's pretty long so feel to shorten if you can see any opportunity for that.

I put it in Word for ease of tracked changes.

Thanks,
Craig

From: Menschik, David <[redacted]>
Sent: Friday, May 07, 2021 12:33 PM
To: Zinderman, Craig E <[redacted]>
Subject: suggested edits as discussed...

…attached…

From: "Menschik, David" <[redacted]>
To: "Baer, Bethany" [redacted]
Subject: RE: Signal Management
Date: Fri, 08 Jan 2021 16:56:52 -0000
Importance: Normal
Attachments: VAERS_data_mining_20210106.pptx
Inline-Images: image001.png

Thanks! I revised the second slide accordingly – could you please look at this slide and let me know if it looks ok? Also wanted to confirm if I have the correct reference in the footnote.
Thanks,
David

From: Baer, Bethany <[redacted]>
Sent: Friday, January 08, 2021 11:25 AM
To: Menschik, David <[redacted]>
Subject: RE: Signal Management

I think the slides look good. I looked back at a few classic data mining references (some Szarfman and DuMouchel papers) and saw that the most common EBGM definition seems to be "Empiric Bayes Geometric Mean." Sometimes it isn't all capitalized and sometimes "empirical" or "Bayesian" is used, but the "Empiric" as the first word form seems most common. I don't think any of them are wrong, and I admit that I have been inconsistent about their use myself.

My only other comment is that if the audience is more sophisticated or wants more statistical info, the presenter should know the term Multi-item Gamma Poisson Shrinker (MGPS) – the algorithm that our data mining uses. As Szarfman's 2004 Pharmacotherapy paper explains: "The MGPS systematically identifies and 'shrinks' the very common and volatile observed:expected ratios with the smaller number of events and expectations. This process guards against generating multiple false-positive signals due to multiple independent comparisons." Referring to the MGPS provides that next level of statistical details that folks like Paige at the CDC have asked about in the past.

Thanks,
Bethany

From: Menschik, David <[redacted]>
Sent: Friday, January 8, 2021 10:43 AM
To: Baer, Bethany <[redacted]>
Subject: RE: Signal Management

Ahhh…thanks!

On different note, any feedback or edits for the 3 slides I drafted based on your slides appreciated… (will likely be sharing with CDC next week…)

Thanks,
David

From: Baer, Bethany <[redacted]>
Sent: Friday, January 08, 2021 10:41 AM
To: Menschik, David <[redacted]>
Subject: RE: Signal Management

It's a foreign brand name not licensed in the US. We see different ones like that every months or two.

From: Menschik, David <[redacted]>
Sent: Friday, January 8, 2021 10:34 AM
To: Baer, Bethany <[redacted]>
Subject: RE: Signal Management

I'm not familiar with Fluenz tetra – is that a new product and should it be added too? (indicates year 2020-2021)

From: Baer, Bethany <[redacted]>
Sent: Friday, January 08, 2021 10:08 AM
To: Hendrix, Brian * <[redacted]>
Cc: Menschik, David <[redacted]>
Subject: RE: Signal Management

Hi Brian,
Thanks for adding back on the list. I had been looking for it. Yes, we'd like the Covid19 (Moderna) and the Covid19 (Pfizer-BioNTech) added as highlighted below. Nothing else on the list needs to be added for this round.
Thanks!
Bethany

From: Hendrix, Brian * <[redacted]>
Sent: Friday, January 8, 2021 10:00 AM
To: Baer, Bethany <[redacted]>
Cc: Menschik, David <[redacted]>
Subject: RE: Signal Management

Hi Bethany,

Not sure how it disappeared from the email, but here's the full list. So I'll add Pfizer and Moderna. Let me know if you want any of the flu vaccines.

-Brian

VAX_NAME	VAX_NAME_FDA	FIRST_APPEARED
COVID19 (COVID19 (UNKNOWN))	COVID19 (COVID19 (UNKNOWN))	12/30/2020
COVID19 (COVID19 (MODERNA))	COVID19 (COVID19 (MODERNA))	12/23/2020
INFLUENZA (SEASONAL) (AFLURIA 03-04)	INFLUENZA (SEASONAL) (AFLURIA)	12/23/2020
COVID19 (COVID19 (PFIZER-BIONTECH))	COVID19 (COVID19 (PFIZER-BIONTECH))	12/17/2020
INFLUENZA (SEASONAL) (FLUZONE HIGH-DOSE QUADRIVALENT 14-15)	INFLUENZA (SEASONAL) (FLUZONE HIGH-DOSE QUADRIVALENT)	12/17/2020
INFLUENZA (SEASONAL) (FLUZONE HIGH-DOSE QUADRIVALENT 17-18)	INFLUENZA (SEASONAL) (FLUZONE HIGH-DOSE QUADRIVALENT)	12/9/2020
INFLUENZA (SEASONAL) (AFLURIA QUADRIVALENT 09-10)	INFLUENZA (SEASONAL) (AFLURIA QUADRIVALENT)	12/3/2020
INFLUENZA (SEASONAL) (FLUENZ TETRA 20-21)	INFLUENZA (SEASONAL) (FLUENZ TETRA)	11/24/2020

From: Baer, Bethany <[redacted]>
Sent: Friday, January 8, 2021 9:58 AM
To: Hendrix, Brian * <[redacted]>
Cc: Menschik, David <[redacted]>
Subject: RE: Signal Management

Hi Brian,
We had decided to include the two brand names (Pfizer-BioNTech and Moderna) but not the COVID (no brand name category) for the Signals table. So I think there should be two rather than three.
Is there some other category I am missing?
Thanks,
Bethany

From: Hendrix, Brian * <[redacted]>
Sent: Friday, January 8, 2021 9:43 AM
To: Baer, Bethany <[redacted]>
Subject: Signal Management

Hi Bethany,

I'll plan to add the 3 Covid entries. Do we need any of the others?

-Brian

Brian Hendrix (contractor)
Commonwealth Informatics, Inc.

Empirica Signal Support Team

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

U.S. FOOD & DRUG ADMINISTRATION

Data Mining Introduction*

- Statistical method for identifying disproportionality (excess of reported adverse event [AE] for a product relative to other products)
- Hypothesis generating
  - Statistical signal of disproportional reporting (SSDR) ≠ safety signal
- Absence of disproportionality does not confirm absence of safety signal nor negate a signal otherwise detected

*Best Practices in Drug and Biological Product Postmarket Safety Surveillance for FDA Staff (November 2019; Draft). Available at https://www.fda.gov/media/130216/download

DE Data Mining Methods

- Empirica™ Signal software (Oracle)
- Calculates Empiric Bayes Geometric Mean (EBGM) using observed to expected (O/E) vaccine-AE pair ratios
  - EBGM derived from a statistical model (Multi-item Gamma Poisson Shrinker; MGPS) that accounts for instability from small numbers by "shrinking" O/E ratios*
- Results adjusted by gender, year received and age
- Vaccine-AE pairs ranked by lower 5% bound of CI of EBGM (EB05)
- Standard threshold for SSDR: EB05 ≥2

*Szarfman A, Tonning JM, Doraiswamy PM. Pharmacovigilance in the 21st century: new systematic tools for an old problem. Pharmacotherapy. 2004 Sep;24(9):1099-104. doi: 10.1592/phco.24.13.1099.38090. PMID: 15460169.

Limitations of Hypotheses Generated by Data Mining Include:

- Impacted by stimulated reporting (e.g., V-safe program)
- Potential statistical interaction (e.g., If vaccines X and Y often given concomitantly, statistical signal for vaccine X and AE Z may be driven by vaccine Y)
- VAERS limitations (e.g., passive reporting, variable reporting by report source, duplicate reports, missing data etc.)

From: "Menschik, David" <[redacted]>
To: "Zinderman, Craig E" <[redacted]>
Subject: FW: CBER VAERS Signal Management Liaisons/Contacts
Date: Thu, 09 Sep 2021 14:02:21 -0000
Importance: Normal
Inline-Images: image004.png; image003.png

FYI

From: Menschik, David
Sent: Wednesday, September 08, 2021 3:43 PM
To: Szarfman, Ana <[redacted]>
Cc: Hendrix, Brian * <[redacted]>; Sydnor, James * <[redacted]>; Lebow, William * <[redacted]>; Baer, Bethany <[redacted]>; Siegel, Jeffrey <J[redacted]>; Stockbridge, Norman L <[redacted]>; Narayan Nair ([redacted]) <[redacted]>
Subject: RE: CBER VAERS Signal Management Liaisons/Contacts

Hi Ana,

I certainly understand where you are coming from. I can in general speak to the understandable concern that you raise regarding stratifying by year. While we must limit conduct, analyses, and discussion of VAERS data mining to CBER/OBE/DE, I can discuss this general issue a bit more.
We have previously thought about and discussed this issue, recognizing that the vast-majority of reports received in VAERS this year have involved COVID-19 vaccines which could drive PT-vaccine disproportionately scores towards the null by contributing substantially to the comparator group, particularly if there is a class-effect (e.g., if all COVID vaccines are associated with the same adverse event). While you have laid out reasonable arguments (which had generally occurred to us) why stratifying VAERS data by year raises new limitations in interpreting data output using existing methods, there are sound reasons for retaining adjustment by year.
From an epidemiology standpoint, exposures and health outcomes (not to mention public perceptions, behaviors, practices, etc. whether stimulated or not) can vary dramatically from one year to the next (independent of vaccinations) and such disparities by year can increase with increasing number of years. For example, during the 'COVID era,' circulating SARS-CoV-2 disease can drive a substantial increase in reported specific AEs, independently of AEs that may be associated with vaccinations; these AEs would most likely be over-represented in individual COVID vaccine-AE disproportionality scores if the comparison group were expanded to include reports from increasing time periods prior to the 'COVID-era.'
On a related note, MedDRA terms are continuously being updated and can regularly have substantial updates introducing new AE terms not available when reports were coded during prior time periods. For instance, this past week (under version 24.1 release), a new preferred term (PT), "multisystem inflammatory syndrome" was added to MedDRA. Without controlling for time (e.g., year) of vaccination, there would likely be inflated disproportionality for newer MedDRA AE terms in association with COVID vaccines since an expanded comparison group would include substantially more VAERS reports that have no chance of having such newer MedDRA terms due to being coded prior to the availability of such a term.
We will plan to discuss internally within CBER/OBE/DE and with Commonwealth options and associated feasibilities, impacts, etc. for potential approaches to addressing the year-stratification issue. Any further discussion on VAERS data mining methods/findings outside my chain of command will have to be offline and in general terms, as well as without reference to any specific VAERS vaccine-PT pair outputs.

Thank you for your understanding,
David

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]

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 <David.Menschik@fda.hhs.gov>; Lebow, William * <[redacted]>; Baer, Bethany <[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
[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
[redacted]

U.S. FOOD & DRUG ADMINISTRATION

From: "Menschik, David" <[redacted]>
To: "Alimchandani, Meghna" <[redacted]>, "Zinderman, Craig E" <[redacted]>
Bcc: "Menschik, David" <[redacted]>
Subject: Coverage through 8/18
Date: Fri, 05 Aug 2022 12:22:50 -0000
Importance: Normal
Embedded: unnamed

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: "Menschik, David" <[redacted]>
To: "Shimabukuro, Tom (CDC)" <[redacted]>, "Su, John (CDC)" <[redacted]>
Cc: "Zinderman, Craig E" <[redacted]>, "Nair, Narayan" <[redacted]>, "Alimchandani, Meghna" <[redacted]>
Bcc: "Menschik, David" <[redacted]>
Subject: Jynneos DM
Date: Tue, 02 Aug 2022 11:35:17 -0000
Importance: Normal
Attachments: USST_JYNNEOS_20220729.xls

Good morning Tom and John,

As per recent leadership meeting request, attached please find a list of all (i.e., unvetted, regardless of notability, etc.) PTs with data mining alerts (i.e., EB05 ≥2) for Jynneos VAERS reports from our weekly 'US Signals Summary Table' ('as of date' 7/29/22). 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: "Baer, Bethany" <[redacted]>
To: "Niu, Manette" <[redacted]>
Cc: "Menschik, David" <[redacted]>, "Zinderman, Craig E" <[redacted]>
Subject: RE: Data mining
Date: Tue, 16 Mar 2021 23:00:24 +0000
Importance: Normal

Sounds good. Thanks, Manette.
Bethany

From: Niu, Manette <[redacted]>
Sent: Tuesday, March 16, 2021 6:38 AM
To: Baer, Bethany <[redacted]>
Cc: Menschik, David <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Data mining

Bethany,
Craig and I discussed this issue yesterday. I'll try and set up a meeting for us to meet with Ana to discuss her VAERS objectives/rationale.
Thank you!
Manette

From: Baer, Bethany <[redacted]>
Sent: Sunday, March 14, 2021 5:11 PM
To: Zinderman, Craig E <[redacted]>; Niu, Manette <[redacted]>
Cc: Menschik, David <[redacted]>
Subject: RE: Data mining

Thanks, Craig. I understand that this is a complex issue and I just wanted to make sure you all were involved and aware of the situation. Her comments were only made on the CDER/CBER call with the contractor, Commonwealth, and were brief mentions. Potentially, she was just brainstorming and theorizing, but it struck me as unusual compared to the typical topics covered on the call. She has been at FDA a long time and involved in many projects, so I would certainly hope that she would reach out to someone in CBER to collaborate if she did want to move forward with something. A joint project could be great. I have had only limited interactions with Ana during CDER/CBER Empirica calls so I do not know her well. I know that she knows a lot more about data mining than I do!
Thanks,
Bethany

From: Zinderman, Craig E <[redacted]>
Sent: Friday, March 12, 2021 3:08 PM
To: Baer, Bethany <[redacted]>; Niu, Manette <[redacted]>
Cc: Menschik, David <[redacted]>
Subject: RE: Data mining

Bethany:
I can understand your concerns, but I'm not sure that there is an obvious solution. Refusing her access just for vaccines seems a little disingenuous since she has full access to FAERS data for all of CBER's >200 non-vaccine products. Seems reasonable to try to understand why she wants to use VAERS data instead of her own Center's data, and to caution her that while its fine for her to do methodological work, we aren't interested in additional data mining studies of COVID data

outside of CBER's usual processes. Did she make these statements publicly? Is there someone who has some sort of relationship/collaboration with her that could approach her on these issues? I don't know her and don't have any interactions with her. Anyone know who she works for?

Thanks,
Craig

From: Baer, Bethany <[redacted]>
Sent: Friday, March 12, 2021 12:42 PM
To: Zinderman, Craig E <[redacted]>; Niu, Manette <[redacted]>
Subject: RE: Data mining

Hi Craig and Manette,
The funding/payment issue is still being looked into by Judy. After a little more thought, I wanted to be sure I clearly raised my other concern. If Ana is not working on a specific project with someone in our office, is there justification for her to have access to the VAERS data in Empirica? As I mentioned in the earlier email, on the Empirica CDER/CBER calls she has twice now expressed interest in the COVID vaccine data mining and made some broad statements that I don't think DE would agree with (e.g., indicating that the vaccines are causing ITP). If she is not working with someone in OBE, who is coordinating the direction of the inquiries and working to interpret the results? Who would provide clearance for sharing the information? Factors such as GDIT VAERS report processing backlogs, and other outside issues, would be important to consider when looking at this data and may not be widely understood outside of DE. Since we typically only give Empirica CBER access to DE medical officers or other CBER members working on specific projects, I defer to you both regarding this. I know Ana worked to develop the data mining system and this might be a special circumstance due to her knowledge and experience. In the last year, she was working on a duplicate detection algorithm with FAERS reports in Empirica that involved some CBER examples, so there might be administrative reasons why she needs the extra access. I wanted to make sure I had expressed my concerns based on what I had heard on some recent calls.
Thanks,
Bethany

From: Baer, Bethany
Sent: Thursday, March 11, 2021 5:51 PM
To: Zinderman, Craig E <[redacted]>; Niu, Manette <[redacted]>
Subject: RE: Data mining

Yes, that is helpful. Thanks, Craig. There were some other emails between Judy and CBER/CDER contracting/funding folks that involved trying to coordinate payments for several different projects. In the end, it sounded like, due to other forces, CDER had to already make the payment and things couldn't be balanced out this year for this between the two centers. I will check in with Judy about the bottom line regarding that. Right now, we will leave Ana's account as is with a CBER login.
Thanks,
Bethany

From: Zinderman, Craig E <[redacted]>
Sent: Thursday, March 11, 2021 3:24 PM
To: Baer, Bethany <[redacted]>; Niu, Manette <[redacted]>
Subject: RE: Data mining

Hi Bethany:
I spoke to Manette about this yesterday. While Ana has spoken to Manette about the work that Manette describes below, sounds like Manette isn't actively engaged in working on a collaboration or project. So, I would say that we don't have a business need to pay for an account for Ana.

However, I would think that we have no objections to her having access/ a CBER account. So, if the payment issue is no longer a problem, then its fine for her to have this access. If CBER is expected to pay, then I think we would not be able to justify the cost of her account (unless the payment is minimal and non-consequential enough that CBER won't need us to justify).

Does that help?

Thanks,
Craig

From: Baer, Bethany <[redacted]>
Sent: Thursday, March 11, 2021 2:43 PM
To: Niu, Manette <[redacted]>; Zinderman, Craig E <[redacted]>
Subject: RE: Data mining

Hi Manette and Craig,
It sounds like the payment issue is less of an immediate concern due to some arrangements between CDER and CBER, but I wanted to reconfirm that you would like Ana Szarfman to have a CBER Empirica account and access to VAERS through Empirica. The way Kosal has set it up, she has a CBER login. On some larger Empirica contractor calls recently, she has expressed interest in using COVID vaccines as an example of data mining and sharing results for training purposes within the FDA. I wasn't sure the process for how that would happen and if DE was wanting to do that with such new products with potentially actively changing safety profiles.
Thanks,
Bethany

From: Niu, Manette <[redacted]>
Sent: Wednesday, March 10, 2021 1:24 PM
To: Zinderman, Craig E <[redacted]>
Cc: Baer, Bethany <[redacted]>
Subject: FW: Data mining

Craig,
The background for this: Ana approached me several months ago as she is interested in testing a new process in VAERS based on new methodology proposed by Bill DuMouchel for data mining focused on signal detection for concomitant medication use. (I did ask Steve if it was alright to grant Ana VAERS access, and he agreed). The project she proposes is very preliminary in the exploratory-hypothesis stage (no protocol).
Thank you!
Manette

From: Niu, Manette
Sent: Wednesday, March 10, 2021 1:14 PM
To: Nguon, Kosal * <[redacted]>
Subject: RE: Data mining

Will CBER or CDER pay for this account?

From: Nguon, Kosal * <[redacted]>
Sent: Wednesday, March 10, 2021 1:11 PM
To: Szarfman, Ana <[redacted]>; Niu, Manette <[redacted]>
Cc: Sydnor, James * <[redacted]>
Subject: RE: Data mining

Hello Manette and Ana,

Ana has a user license through CDER, and the CDER/CBER designation is for purposes of tracking and paperwork. Ana is listed as a "CBER" login user as this is the only feasible avenue to provide access to both FAERS and VAERS data. We may have to move Ana back to the CDER login group for the annual user review and we can move her back to CBER afterwards , but it should be of minimal impact. I believe everyone was in agreement that Ana should have access to VAERS data. Thanks to you all for your patience and understanding.

Best,

Kosal

From: Szarfman, Ana <[redacted]>
Sent: Wednesday, March 10, 2021 12:54 PM
To: Niu, Manette <[redacted]>
Cc: Nguon, Kosal * <[redacted]>
Subject: RE: Data mining

As far as I can understand, I have access as a CDER user. I added Kosal in the cc because he did the programming assessment.

From: Niu, Manette <[redacted]>
Sent: Wednesday, March 10, 2021 12:52 PM
To: Szarfman, Ana <[redacted]>
Subject: RE: Data mining

Great, this is through CDER, correct?

From: Szarfman, Ana <[redacted]>
Sent: Wednesday, March 10, 2021 12:51 PM
To: Niu, Manette <[redacted]>
Subject: RE: Data mining

Yes, I now have access to the VAERS data.

From: Niu, Manette <[redacted]>
Sent: Wednesday, March 10, 2021 12:49 PM
To: Szarfman, Ana <[redacted]>
Subject: RE: Data mining

Ana,
Do you have access to VAERS in your Empirica account?
Thank you!
Manette

From: Szarfman, Ana <[redacted]>
Sent: Wednesday, March 10, 2021 12:47 PM
To: Niu, Manette <[redacted]>
Subject: RE: Data mining

Hi Manette,

Thanks you for addressing this issue! I just got access.

Do you want to participate in the meetings with Bill DuMouchel?

From: Niu, Manette <[redacted]>
Sent: Wednesday, March 10, 2021 12:45 PM
To: Szarfman, Ana <[redacted]>
Subject: Data mining

Ana,
Were you able to get your CDER Empirica account set up?
Thank you!
Manette

From: "Szarfman, Ana" <[redacted]>
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.
Date: Fri, 26 Mar 2021 19:48:39 +0000
Importance: Normal
Attachments: CovidWeek9MaskExamples.xls; AllCovid.zip; Ana_Szarfman_-_Briefing_of_Dr_Peter_Marks_-_March_1_2021_at_100_PM..pdf
Inline-Images: image003.png

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]

FDA U.S. FOOD & DRUG ADMINISTRATION

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

I am humbled and thankful for the tremendously difficult and amazingly hard work you are all doing and for all your successes

Many thanks for your invitation to exchange thoughts

Mortality data to address the COVID-19 public health and analytical needs of the users of the information

More timely detection of adverse events and associated risk factors that we may not know how to formulate *a priori*

Ana Szarfman, MD, PhD, FAMIA, Medical Officer, Diplomate by the American Board of Pathology in both, Clinical Pathology (1984) and Clinical Informatics (2017), Safety Data Mining Developer and Medical Informatics Analyst, Division of Cardiology and Nephrology, CDER, FDA

Dr. Peter Marks Briefing, February 29, 2021

Mortality data linked to EHRs and Claims data

- **There is no universal tool to CENTRALLY capture mortality data in the U.S.**
- Multiple surveillance and analytical systems cannot easily access data on *DEATHS OCCURRING IN AN OUT-OF-HEALTHCARE SETTING* using EHRs or claims data.
- If a patient *DIES WHILE IN THE HOSPITAL, the death will be coded as such in the EHR, but not if the death occurs with patients being discharged to hospice care, to a nursing home, or to their home.*
- When such information is needed for RESEARCH or to be linked to CLAIMS data (such as BCBS), it is typically obtained from *PRIVATE SERVICES* who collect the information from various sources like *FUNERAL HOMES AND OBITUARIES IN LOCAL NEWSPAPERS.* **(a very time consuming and very inefficient process)**

- Deaths occurring at home, on the street, or when the subject is homeless, as well as autopsies are not treated as medical clinical service events.
- The death certificate information does not get back to the medical record.
- Death certificates are notoriously sparse and incomplete.
  - They are collected by a multitude of governing localities, and then gathered by the State. The types of reportable deaths are determined by federal, state or local laws.

- The States receive the certificates and submit them to the National Death Index (NDI) where they get adjudicated and added to the NDI final file annually.
- The NDI website provides death information to researchers; but the process requires funding support*:*
  https://www.cdc.gov/nchs/data/ndi/ndi_application.pdf
  https://www.cdc.gov/nchs/ndi/portal.htm
- These requests are usually applied to 500 patients in a research project or clinical trial.

→We need mortality data for over 300 million individuals

- Clinical trials contain COMPREHENSIVE AUTOPSY information but not the EHRs.
- Registries, like cancer registries or transplant registries systematically collect death information, but they represent relatively small siloes of information disjoined from EHRs.
- CMS and DoD and VA hospitals GET FEEDBACK SEEDS from the SSA of *DEATHS THAT NEED TO BE REMOVED FROM THEIR BENEFICIARY LIST.*
  - *NOT SURE HOW OPTIMIZED THESE SYSTEMS ARE FOR INCLUDING DEATHS IN THEIR ANALYSES OF CLINICAL DATA.*
- Outside these Federal systems, the SSA STOPPED MAKING THIS INFORMATION AVAILABLE 5 YEARS AGO because of a potential for fraud (people applying for loans using fake Social Security codes) and would only typically provide an answer for a specific person, and inform when and how they died.

- Death information IS NOT CONSIDERED PRIVATE HEALTH INFORMATION. IT IS PUBLIC INFORMATION.
  - Family members, life insurance companies, and voter registrations are granted such access.
- The key problem is that this information is not collected and made available in a timely way.
- If there is a time, it is now for the Federal Government to act and improve the ability to centrally collect death information or to expedite the link of NDI information to address the necessary research and public health needs of all the analysts of the information.
  - The government can define the requirements and precautions that the receiving parties will need to put in place to avoid fraud.
  - The government can also monitor and prevent fraud activity.
- Correction of this situation will require awareness of this problem, know-how, efforts, funding, and regulatory support.

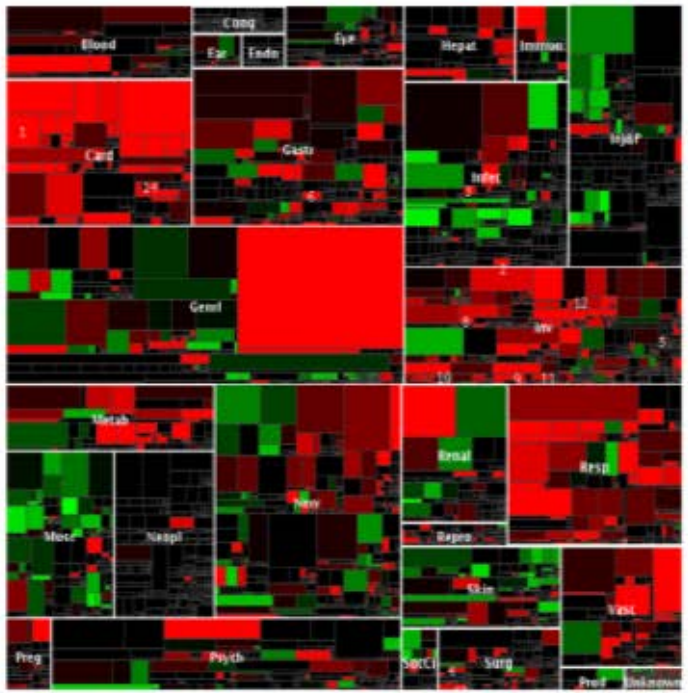
The superior performance of the RGPS algorithm for data mining spontaneous reports, currently only available outside the FDA and CDC

There is a contracting mechanism in place at the FDA to solve this problem quickly

SECTOR MAPS FOR COVID19 VACCINES IN VAERS, RESTRICTED TO AGES 18-75, FDA YEARS 2019-2021, FROM 3 DISPROPORTIONALITY ALGORITHMS

PRR

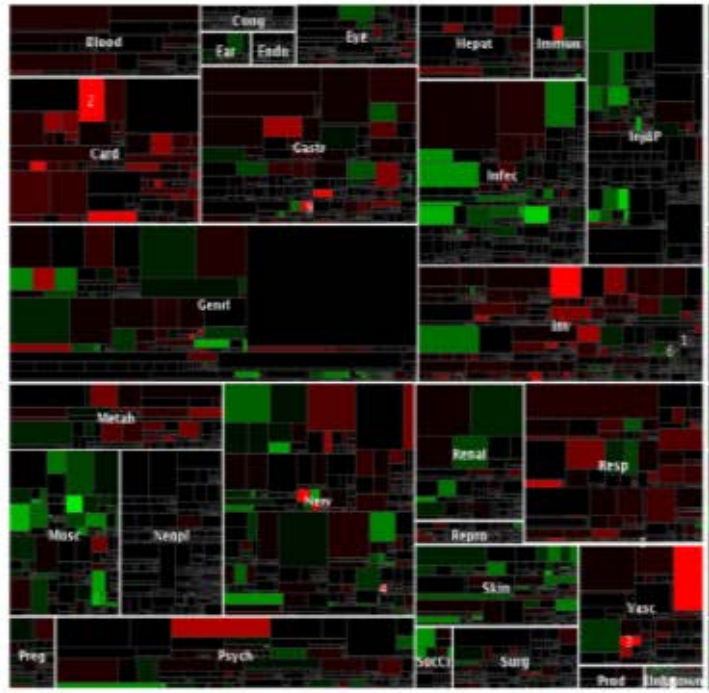
Drug=COVID19



Rank	SOC	Term (PT)	PRR
1	Card	Cardio-respiratory arrest	164.466
2	Infec	COVID-19 pneumonia	164.466
3	Infec	Appendicitis	102.248
4	Surg	Appendicectomy	93.204
5	Inv	SARS-CoV-2 test positive	84.463
6	Gastr	Small intestinal obstruction	48.895
7	Resp	Throat clearing	42.972
8	Inv	Activated partial thromboplastin time prolonged	40.005
9	Inv	Blood magnesium decreased	40.005
10	Inv	Blood pH increased	40.005

RGPS: ERAM

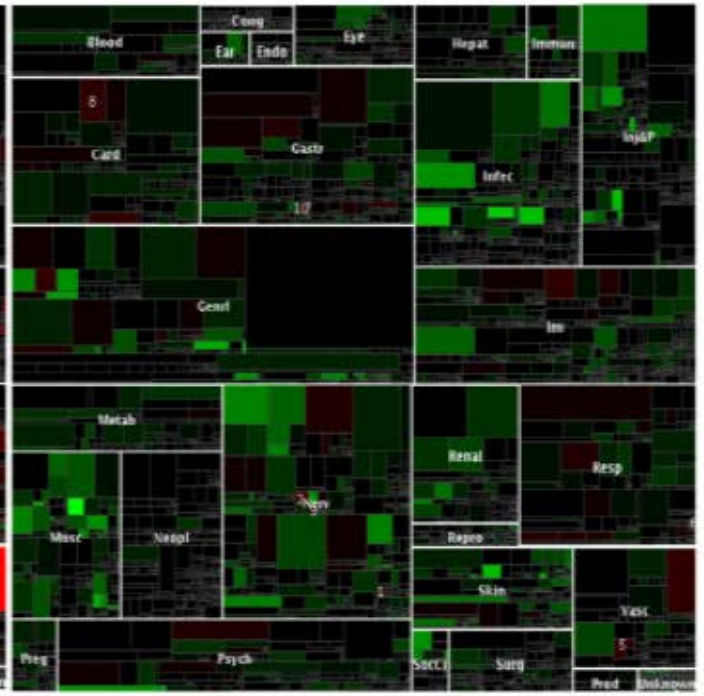
Drug=COVID19



Rank	SOC	Term (PT)	ERAM
1	Inv	SARS-CoV-2 test positive	13.728
2	Card	Tachycardia	7.770
3	Vasc	Flushing	6.754
4	Nerv	Anosmia	6.470
5	Resp	Throat clearing	6.316
6	Inv	Procalcitonin increased	5.795
7	Unknown	Hypertension [SMQ] [broad]	5.766
8	Unknown	Hypertension [SMQ] [narrow]	5.695
9	Gastr	Paraesthesia oral	5.670
10	Unknown	Gastrointestinal perforation [SMQ] [broad]	5.590

MGPS: EBGM

Drug=COVID19



Rank	SOC	Term (PT)	EBGM
1	Nerv	Anosmia	1.991
2	Nerv	Dysgeusia	1.941
3	Unknown	Taste and smell disorders [SMQ] [broad]	1.915
4	Unknown	Taste and smell disorders [SMQ] [narrow]	1.915
5	Vasc	Flushing	1.863
6	Resp	Pharyngeal paraesthesia	1.811
7	Gastr	Paraesthesia oral	1.773
8	Card	Tachycardia	1.770
9	Nerv	Agnosia	1.746
10	Gastr	Hypoaesthesia oral	1.715

The graph showing the signals of 3 Sector Maps next to each other is quite interesting

- The PRR on the left highlights almost everything
- The MGPS on the right is flat (you are not getting useful information with such low counts)
- The RGPS in the middle looks more informative for follow-up evaluation
  - This is because RGPS can better adjust for both, masking (false negatives) and confounding (false positives).

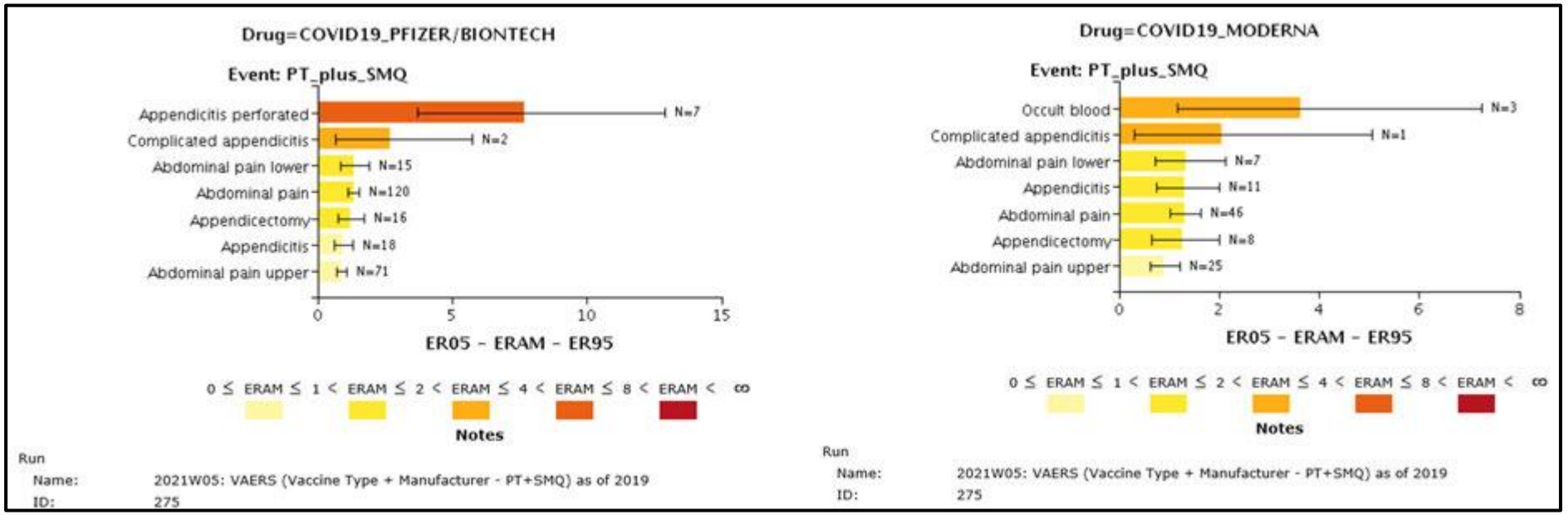
Drug=COVID19_PFIZER/BIONTECH

Rank	SOC	Term (PT)	ERAM
1	Vasc	Hypertensive emergency	31.164
2	Infec	COVID-19 pneumonia	24.861
3	Nerv	Anosmia	13.750
4	Resp	Throat clearing	11.520
5	Infec	Suspected COVID-19	10.500
6	Infec	COVID-19	9.619
7	Vasc	Pelvic venous thrombosis	9.439
8	Surg	Hospice care	8.551
9	Inv	Forced expiratory volume decreased	7.781
10	Resp	Respiratory tract irritation	7.746

Drug=COVID19_MODERNA

Rank	SOC	Term (PT)	ERAM
1	Infec	COVID-19 pneumonia	58.246
2	Resp	Lung opacity	13.824
3	Resp	Throat clearing	7.832
4	Repro	Uterine spasm	6.756
5	Inv	Electrocardiogram PR shortened	6.021
6	Inv	Computerised tomogram thorax abnormal	5.917
7	Repro	Suppressed lactation	5.719
8	Resp	Chronic obstructive pulmonary disease	5.595
9	Prod	Product container issue	5.258
10	Surg	Hospice care	5.229

Only to highlight an "Appendicitis, perforated" signal with the Pfizer vaccine that may require follow-up evaluation



For the seven Pfizer cases the onset times for appendicitis perforated are as follows:

VAERS_ID from Demo	Onset days	Onset period	Onset period
0951817	1	Days: 0-6	Days: 0-6
0952318	1	Days: 0-6	Days: 0-6
0955762	0	Days: 0-6	Days: 0-6
0961503	2	Days: 0-6	Days: 0-6
0962110	0	Days: 0-6	Days: 0-6
0974009	3	Days: 0-6	Days: 0-6
0979690	2	Days: 0-6	Days: 0-6

The MGPS data mining method currently in use at the Agency and at the CDC is not the state of the art

RGPS is the state of the art:

- Is a regression-based extension of MGPS that incorporates more information into the signal generation process. This leads to a lower rate of missed signals and less false alerts.
  - Removes the effect of products whose strong signals in the background are overwhelming the signals of the product of interest and thus uncovers signals of products being masked (the false negatives.)
  - Adjusts for the concomitant products in the same reports having strong signals to remove the confounding that generates false positives with innocent bystander products.

VAERS doesn't code concomitant medications, while FAERS does

- If we would also adjust for the concomitant meds in the narratives of VAERS reports, we could improve the estimate of confounding (false alerts or false positives) even better.
- Bill DuMouchel, who developed both, MGPS and RGPS is planning to extract and code the drugs in the narratives of VAERS reports to reduce confounding.

- Note also that REPORTS FOR THE ASTRA ZENECA VACCINE IN USE IN THE U.K. *arrive to the FAERS data instead of arriving to the VAERS data.*

Simultaneous, automated identification of adverse events and risk factors from multiple products in multiple clinical trials, that we may not know how to specify *a priori*

This is a new analytical approach by Dr. Bill DuMouchel that will require funding

Analysis of adverse event data (as opposed to efficacy) from studies of medical products involve several difficult problems

- Pre-specification of analysis end points are rarely possible for adverse events, leading to a multiple comparisons challenge whenever many different adverse events show up.
- Rare adverse event issues often show up with small counts that could be grave dangers to public health.
- Data from many clinical trials and observational studies may need to be analyzed jointly, such as for many newly developed products.

Risk prediction for diseases for which we do not understand the trajectory of individual patients is low

- There is a need for a statistical method capable of identifying rare, unbalanced risk factors that not only occur during or after treatment, **but also, more importantly, at baseline.**
- Consistent efficacy and safety signals may remain hidden in disjoined clinical trial applications or by analyses that do not properly adjust for multiplicity and small counts across all the data being generated.
- We need to implement a solution that will allow for display of the adverse event information and results of analyses in an interactive and user-friendly way that will not require a continuous and impossible-to-document customizations of the complex data and the analytical tools.

Advantages of having this automated, intuitive, and interactive analytical program in place

- Can perform of analysis of patient level data from many clinical trials and many applications at once.
- The addition or removal of some data will generate a quick analysis WITHOUT HAVING TO DO A LOT OF INTERMEDIATE ANALYSES and redo a meta-analysis based on the new summary statistics.
- Can provide TRANSPARENCY to the decision-making process and enable RAPID RE-ASSESSMENT OF THE DATA FOR INCREASED COMPREHENSION of new and evolving issues of interest

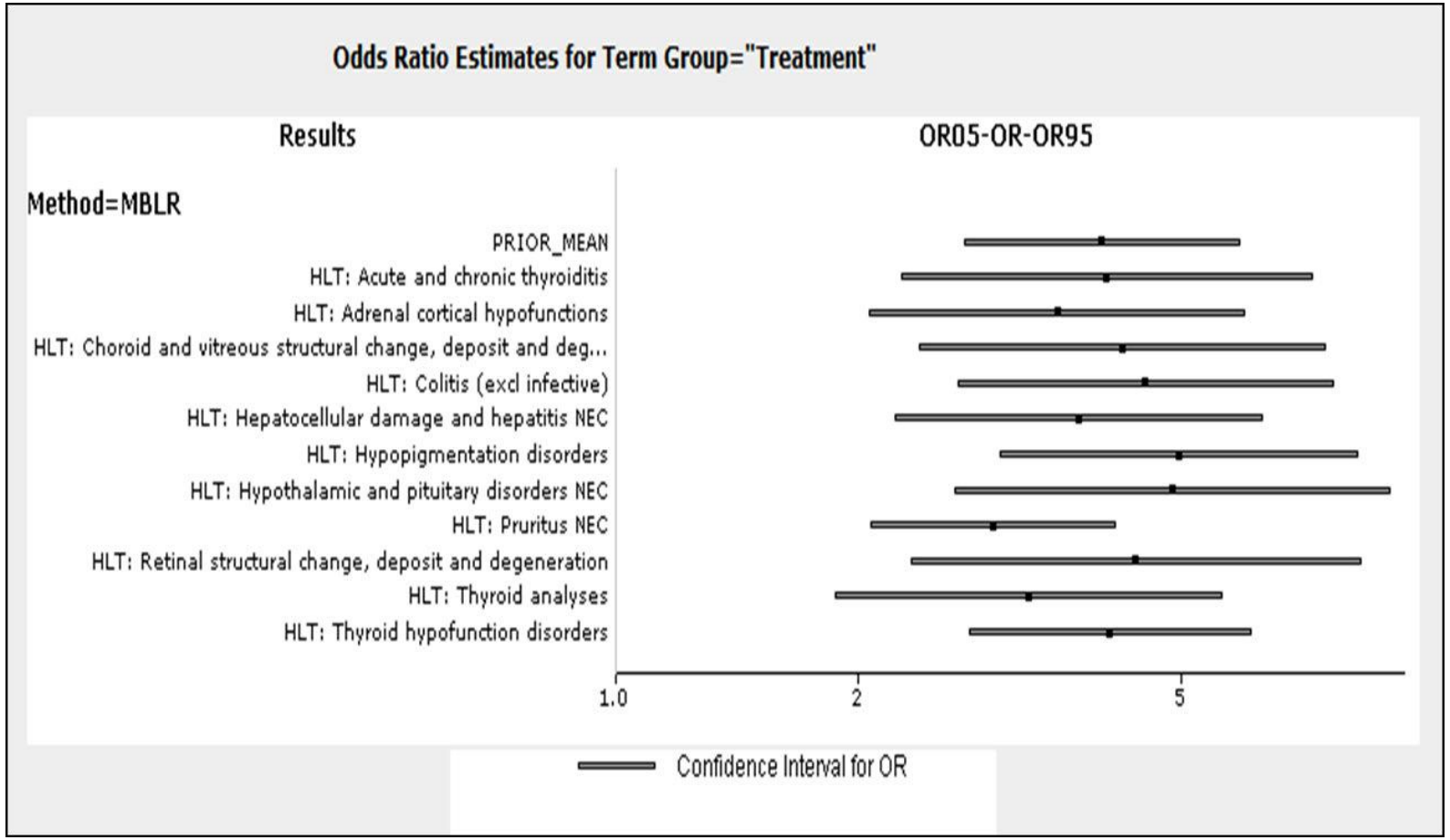
Comparing different treatments that do not all appear in any one study

- This approach will compare subgroups based on multiple treatment arms, covariates, and 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.
- **THESE ACTIVITIES NEED TO BE DONE EXPEDITIOUSLY, IN A TRANSPARENT WAY TO AVOID PUBLIC CONFUSION**
- **WE HAVE NEVER BEEN ABLE TO CONDUCT SUCH ASSESSMENTS, AND EVEN LESS TO CONDUCT SUCH ASSESSMENTS IN AN INTUITIVE AND AUTOMATED MANNER.**

Eight Studies of the ICPI Nivolumab versus other Active Comparators -- How to compare the various treatment effects?

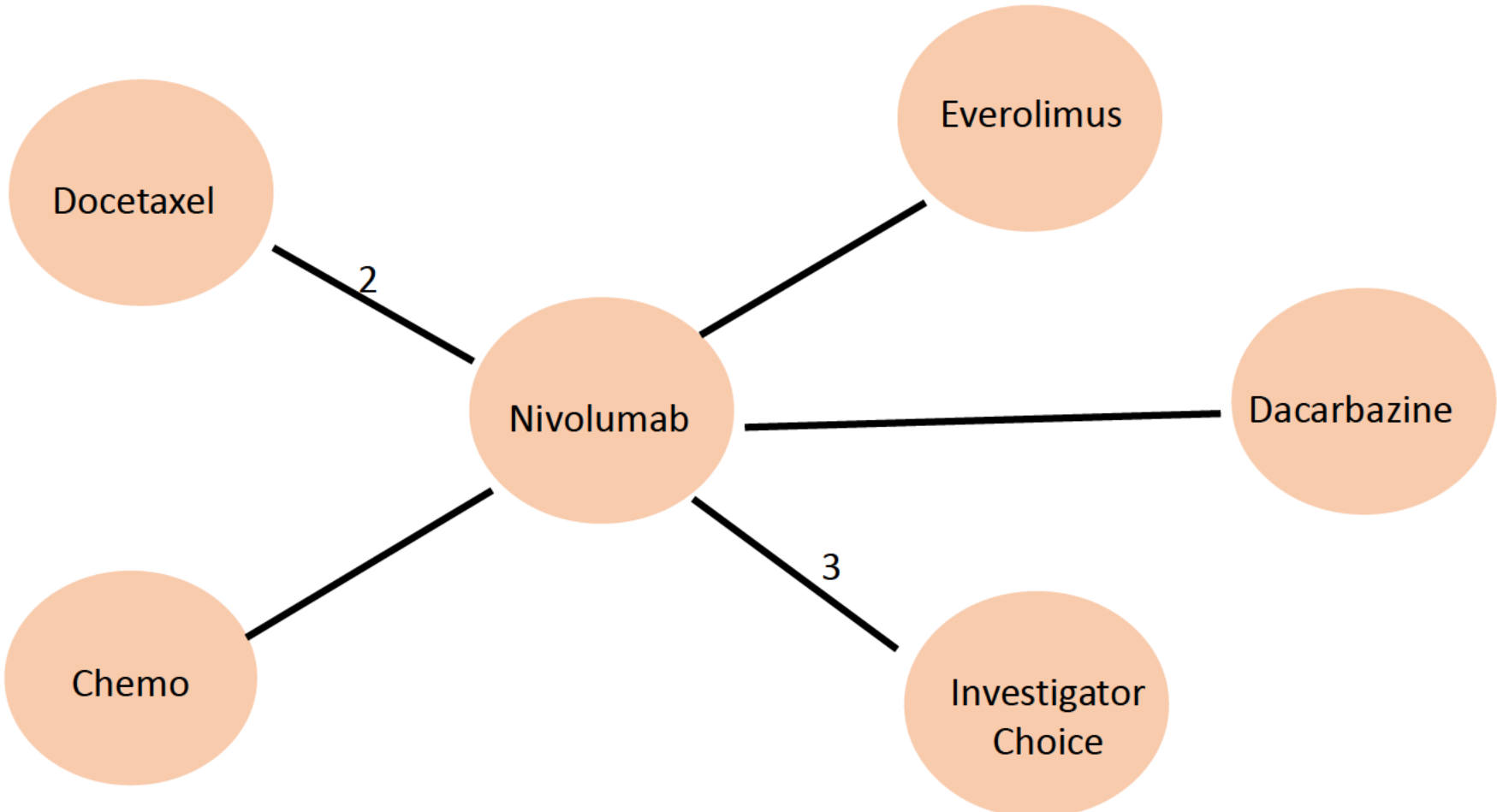
Study	ARM 1	Arm 2
37	NIVOLUMAB	INVESTIGATOR CHOICE
17		DOCETAXEL
25		EVEROLIMUS
57		DOCETAXEL
66		DACARBAZINE
26		INVESTIGATOR CHOICE
41		INVESTIGATOR CHOICE
27		CHEMO

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

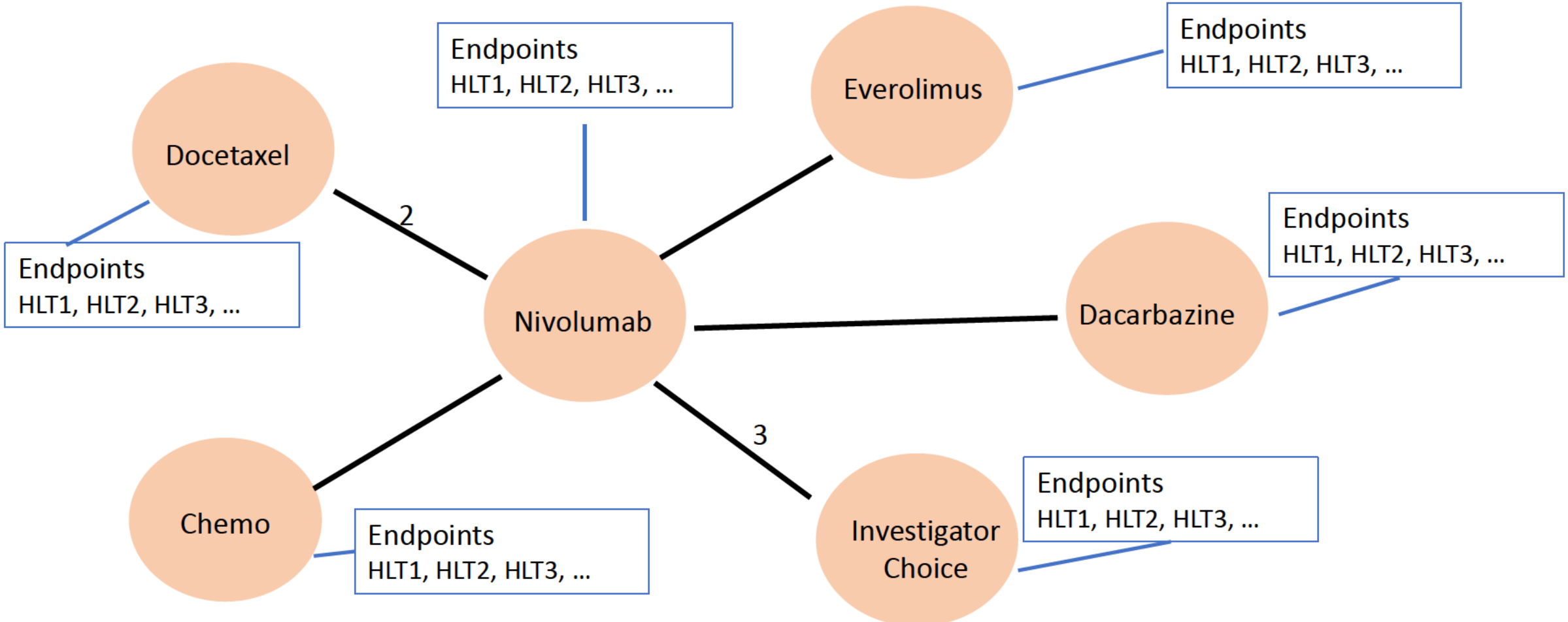


Representing a Pool of Studies as a Network

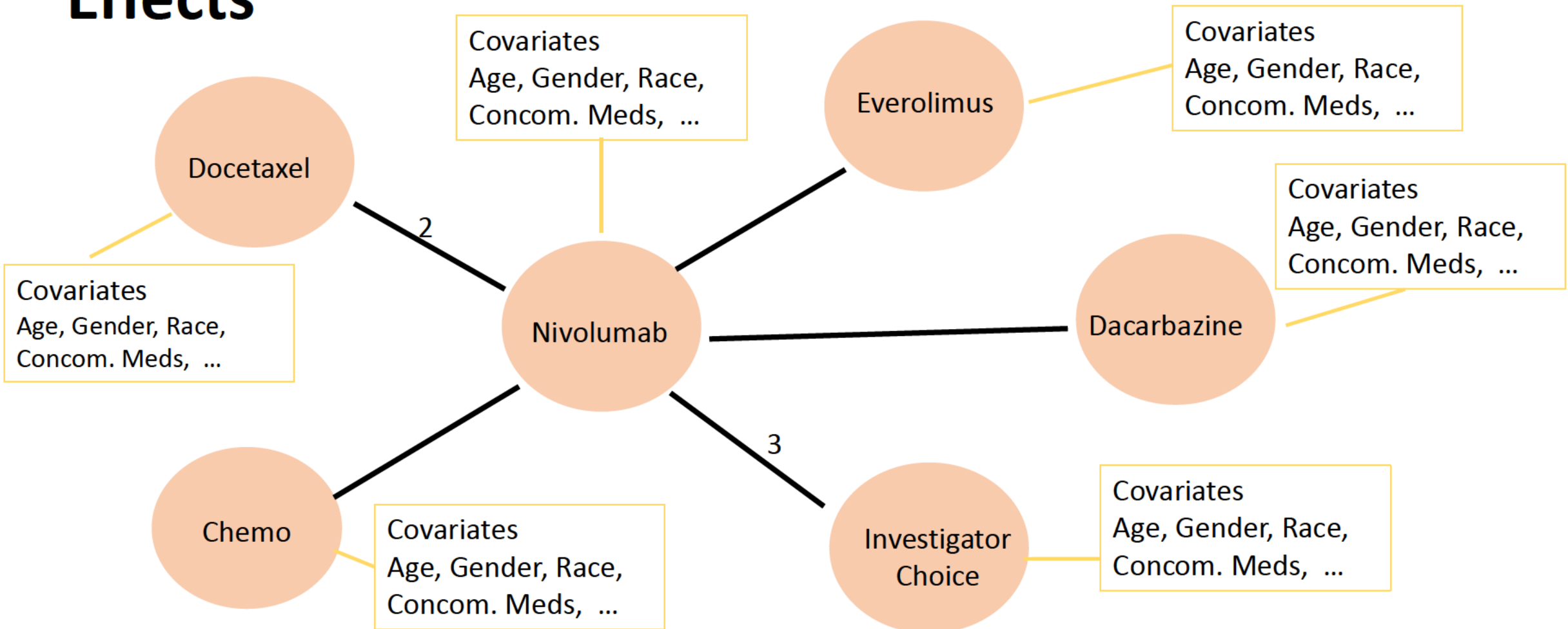
Lines Connecting Two Arms Represent within-Study Comparisons (numbers=study multiplicity)



Each Patient May Have Multiple Medically Similar Endpoint Measurements



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



Thanks to your work, there are now in place multiple approaches for passive and active surveillance for post-authorization safety signals assessments

Some critical roadblocks to consider:

- Claims not collecting vaccinations in a systematic way to understand who gets vaccinated and with which vaccine.
- No direct link/access to dead registries or EHRs to understand cause of death and underlying risk factors.
- **Viral variants need to be associated with particular clinical profiles,** a task that requires the maintenance of a high level of clinical data veracity. **However,**
- There is a lack of a **universal proactive definition and application of permissible variables and values in EHRs** and of **unique IDs for patients (linked to providers and health facilities)** that will simplify access to such data.
- There are **incomplete and not standardized data creation practices in place within and across systems** that unnecessarily delays the analytical processing.
- There are **no simple ways to follow individual patient progression in EHRs and in other sources of data**

Potential improvements:

- For passive surveillance:
  - RGPS will provide a big advantage over MGPS for signal detection
- For active surveillance:
  - The simultaneous automated assessment of safety data from multiple clinical trials and OD -- and associated risk factors that we may not be able correctly specify *a priori* -- will provide a great analytical advantage

Zip archive written by user: Bill DuMouchel on: 03/23/2021 20:20:58 UTC
Source Data: VAERS data as of March 05 of 2021 from www.vaers.hhs.gov loaded on 2021-03-07 00:00:00.0