Rolling! Ensuring High-Quality Data for Public Health Use

4:15 PM EST | December 6th, 2022 | 2022 SyS Symposium

Moderator:

- James Muncy, MPH – CSTE

Presenters:

- Kelly Carey, MPH – Division of Health Informatics and Surveillance, CDC
- Howard Burkom, PhD – Johns Hopkins Applied Physics Laboratory (JHUAPL)
- Breanna Swan Hovey, PhD – Johns Hopkins Applied Physics Laboratory (JHUAPL)

KAHUINA CONSULTING

CSTE

Council of State and Territorial Epidemiologists

Kelly Carey, MPH
Division of Health Informatics and Surveillance, CDC

CSTE

KAHUINA
CONSULTING

OPTIMIZING DATA QUALITY

Using Facility Total Volume and Availability of Discharge Diagnoses for Year-Over-Year Trend Analyses

December 6, 2022

Kelly Carey, MPH
Mathematical Statistician, Division of Health Informatics and Surveillance

National Syndromic Surveillance Program (NSSP)

NSSP data

- Are collected in near real-time
- Represent 73% of emergency departments (EDs) nationwide

From January 2018 to November 9, 2022

- More than 535 million ED visits
- 4,718 facilities reported at least 1 ED visit

ED visit information is used to conduct routine surveillance activities for a multitude of categories

$$\begin{array}{c} Surveillance \\ Trend \end{array} = \frac{\begin{array}{c} ED\ visits\ with\ a\ specific\ set\ of \\ diagnosis\ codes \end{array}}{All\ ED\ visits}$$

ED visit information is used to conduct routine surveillance activities for a multitude of categories

Discharge diagnosis (DD) field information

$$Surveillance\ Trend = \frac{ED\ visits\ with\ a\ specific\ set\ of\ diagnosis\ codes}{All\ ED\ visits}$$

ED visit information is used to conduct routine surveillance activities for a multitude of categories

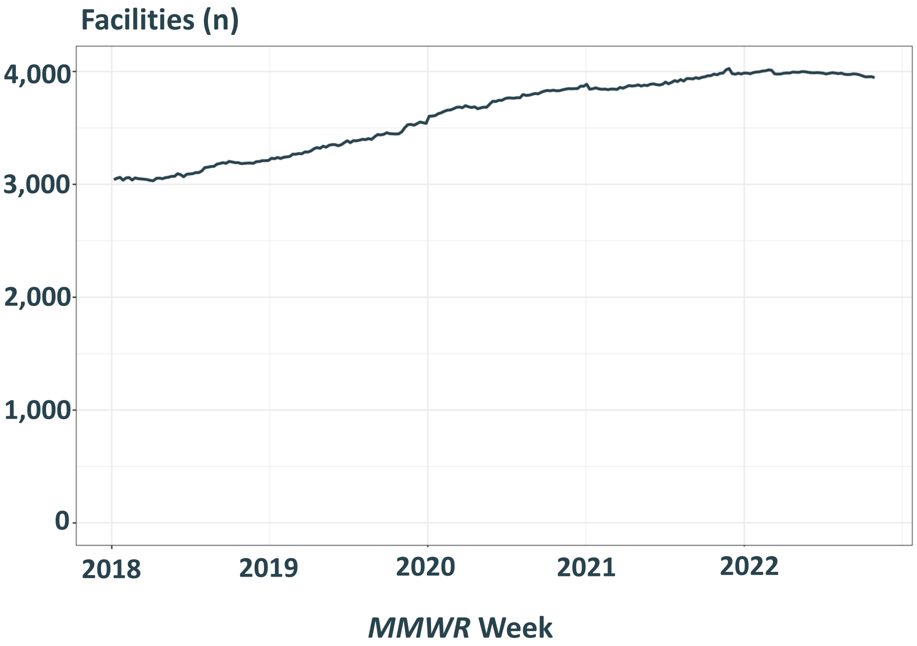
$$\text{Surveillance Trend} = \frac{\text{ED visits with a specific set of diagnosis codes}}{\text{All ED visits}}$$

Total visit volume

VARIATION IN DATA AVAILABILITY EXISTS BY FACILITY

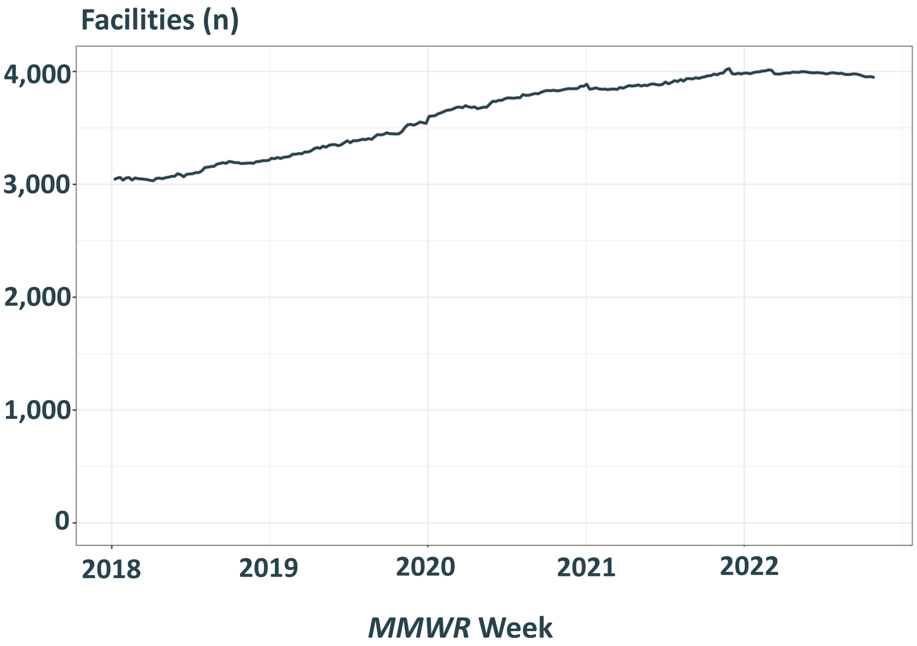
As NSSP has expanded, availability of data has increased

3,032 – 4,026 facilities reporting data per week

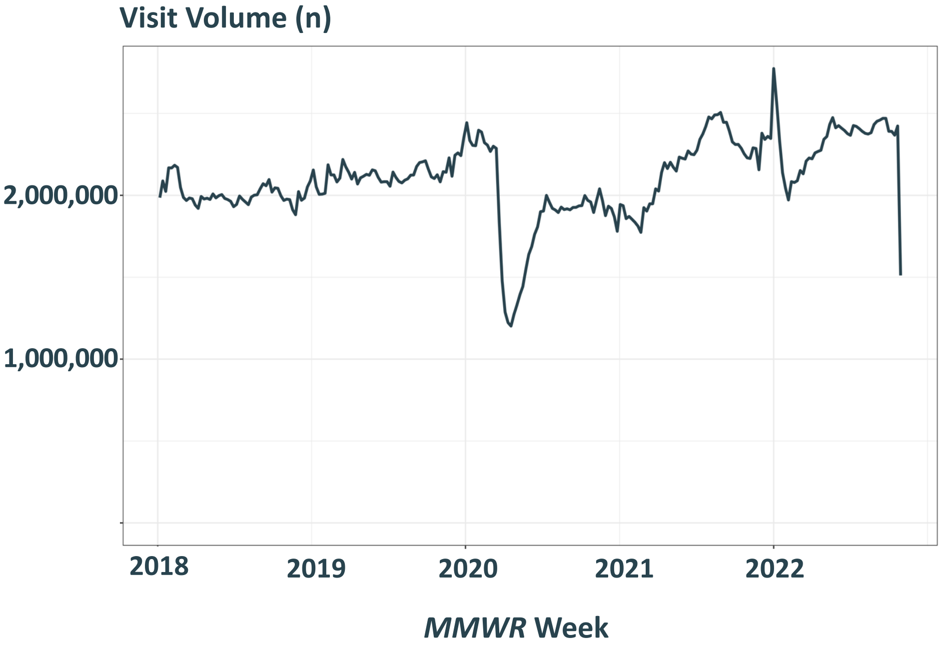


As NSSP has expanded, availability of data has increased

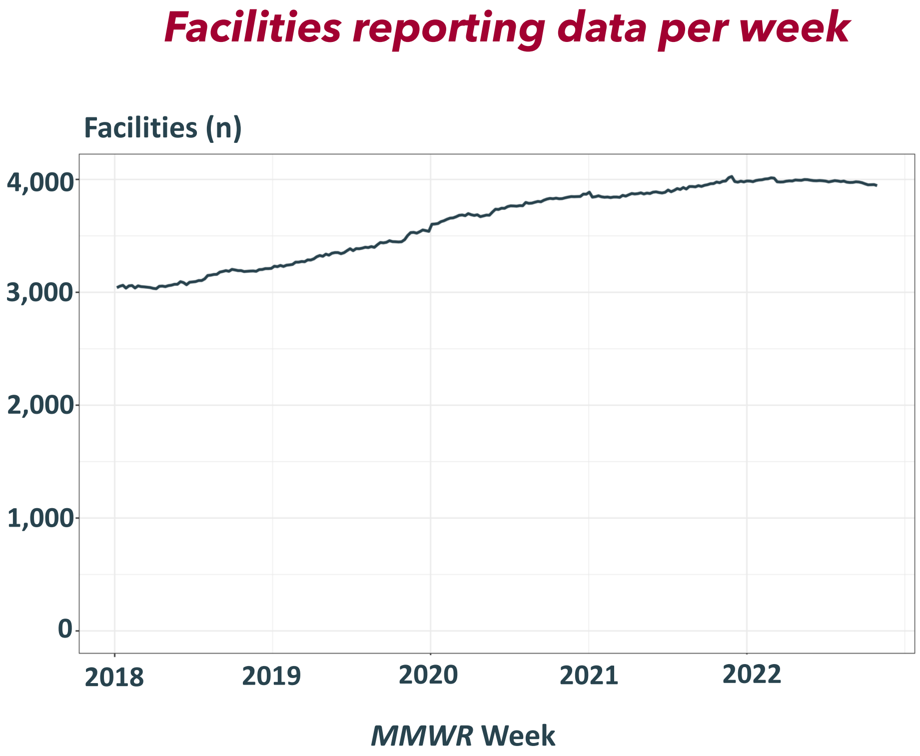
3,032 – 4,026 facilities reporting data per week



1.2 – 2.7 million visits per week



New facility onboarding increased geographic coverage



Compare 2022 data to 2018 data:

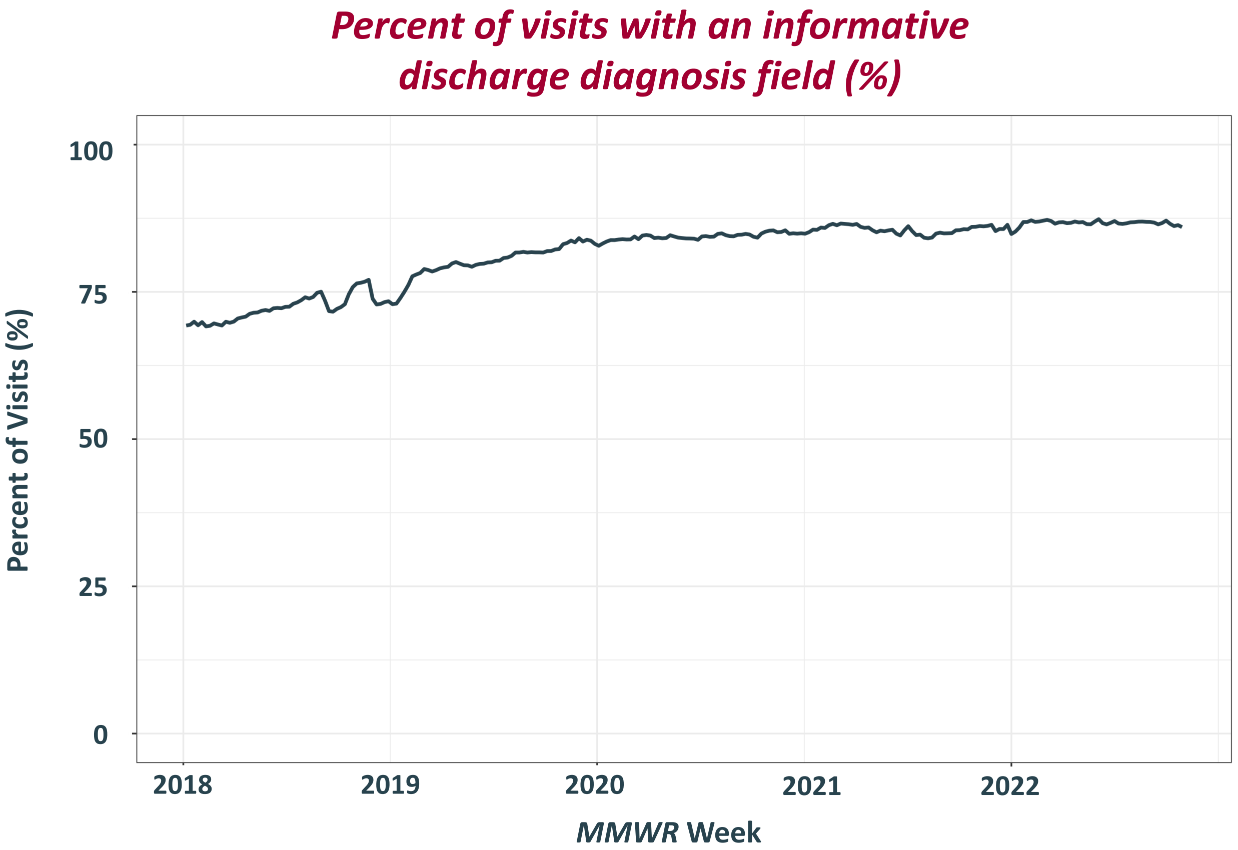
27.9% more facilities

- **15.5% more visits per week on average**

Visits from 2022 are from a more diverse set of facilities than in 2018

Increases in visits impact the denominator

Data in the discharge diagnosis field have improved



Informative

ICD-10-CM

SNOMED

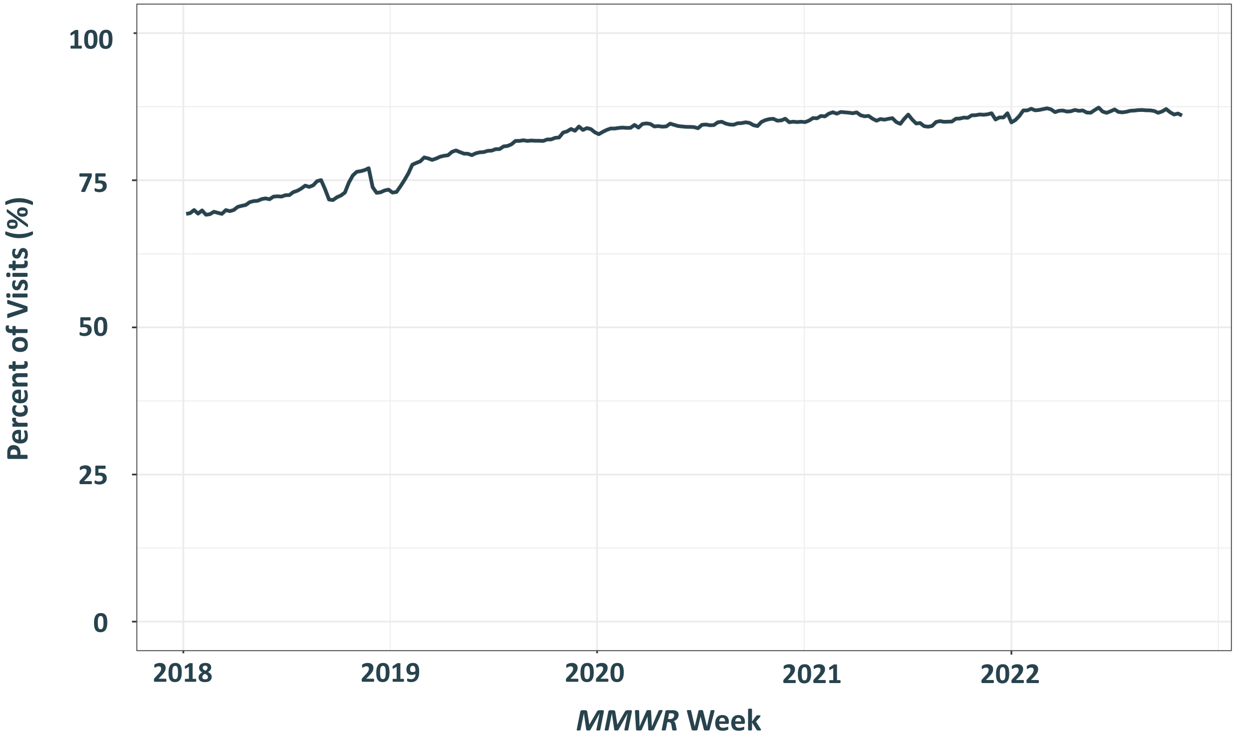
ICD-9

Uninformative

Null or blank

Data from recent years are more likely to be categorized into a surveillance category than earlier years

Percent of visits with an informative discharge diagnosis field (%)



Compare 2022 data to 2018 data:

19.9% more visits with an informative DD field

Visits from 2022 are more likely to have the DD codes needed to classify the visit into a surveillance category

When comparing surveillance trends, data available in 2022 are not the same data available in 2018

$$\text{Surveillance Trend} = \frac{\text{ED visits with a specific set of diagnosis codes}}{\text{All ED visits}}$$

When comparing surveillance trends, data available in 2022 are not the same data available in 2018

$$\text{Surveillance Trend} = \frac{\text{ED visits with a specific set of diagnosis codes}}{\text{All ED visits}}$$

2022 has 27.9% more facilities and 15.5% more visits per week than 2018
Impacts denominator

When comparing surveillance trends, data available in 2022 are not the same data available in 2018

2022 has 19.9% more visits with an informative discharge diagnosis field than 2018
Impacts numerator

$$\mathit{Surveillance\ Trend} = \frac{\mathit{ED\ visits\ with\ a\ specific\ set\ of\ diagnosis\ codes}}{\mathit{All\ ED\ visits}}$$

2022 has 27.9% more facilities and 15.5% more visits per week than 2018
Impacts denominator

VARIATION CAN BE CONTROLLED THROUGH TWO DATA QUALITY FILTERS

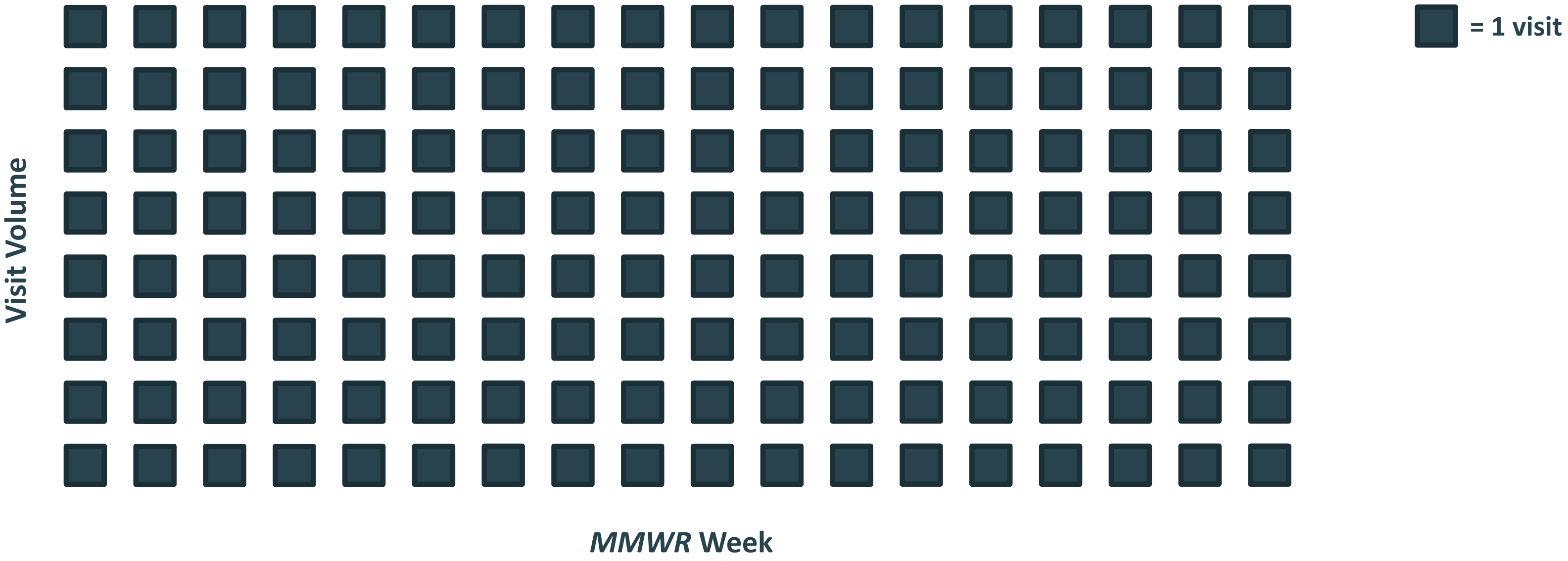
Coefficient of variation (CoV) controls for gaps in facility reporting

$$CoV = \frac{standard\ deviation\ (\sigma)}{mean\ (\mu)} * 100$$

- Measures total volume volatility over time
- Controls for facility onboarding and periods when a facility does not send data
- Recommended values ≤ 45

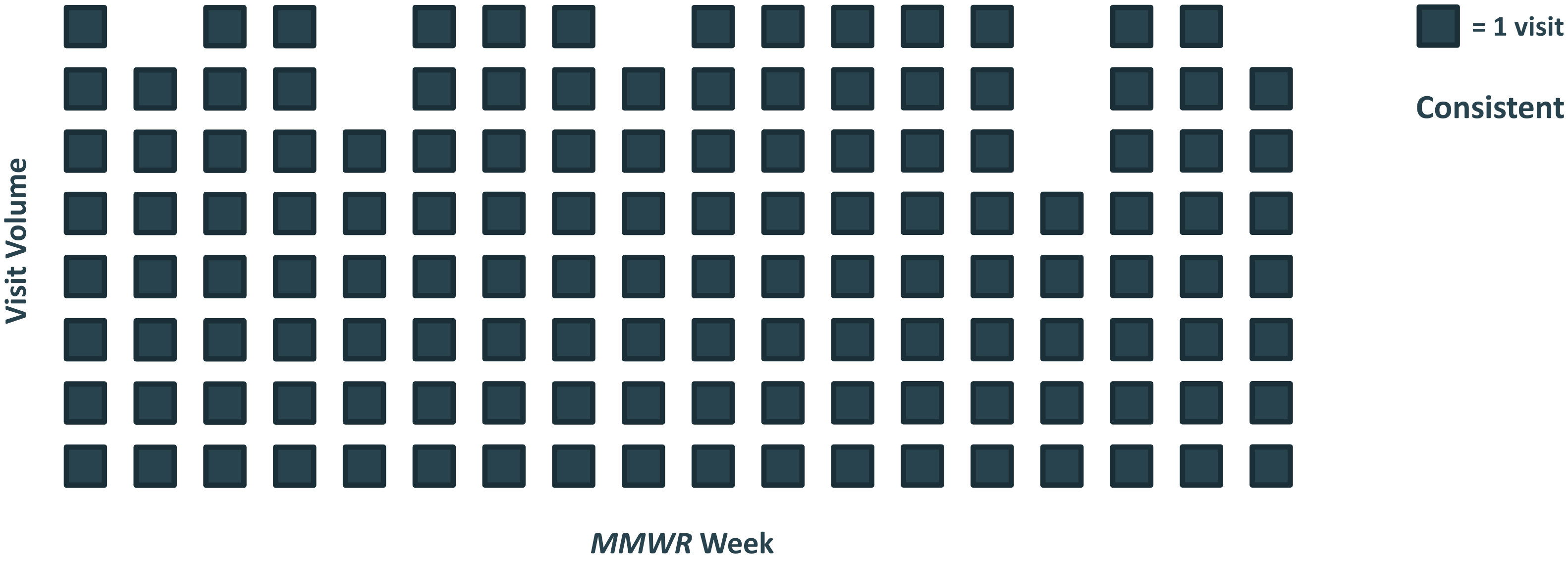
Coefficient of variation (CoV) controls for gaps in facility reporting

$$CoV = \frac{standard\ deviation\ (\sigma)}{mean\ (\mu)}$$



Coefficient of variation (CoV) controls for gaps in facility reporting

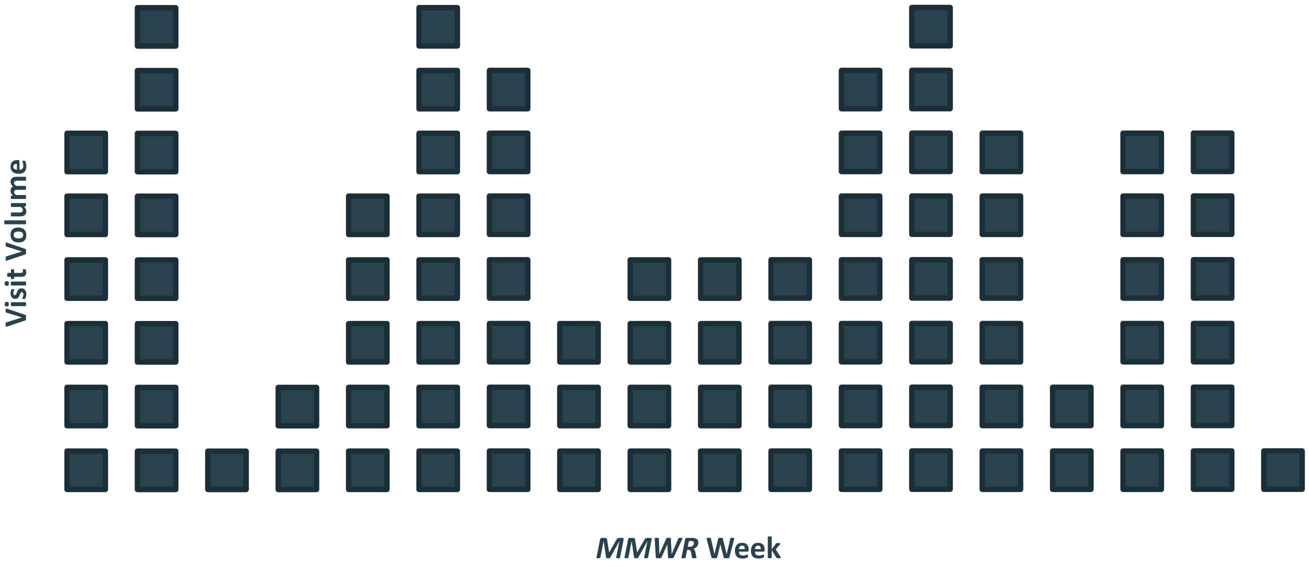
$$CoV = \frac{standard\ deviation\ (\sigma)}{mean\ (\mu)}$$



Consistent reporting

Coefficient of variation (CoV) controls for gaps in facility reporting

$$CoV = \frac{standard\ deviation\ (\sigma)}{mean\ (\mu)}$$



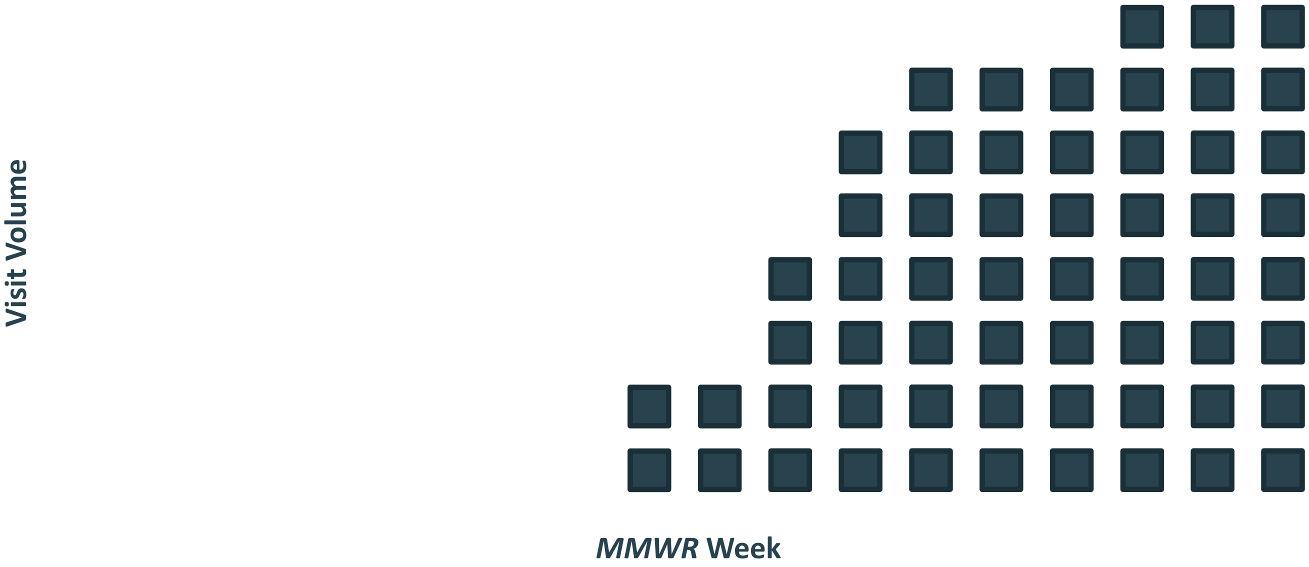
= 1 visit

Consistent reporting

Highly variable reporting

Coefficient of variation (CoV) controls for gaps in facility reporting

$$CoV = \frac{standard\ deviation\ (\sigma)}{mean\ (\mu)}$$



= 1 visit

Consistent reporting

Highly variable reporting

Late onboarding

Coefficient of variation (CoV) controls for gaps in facility reporting

$$CoV = \frac{standard\ deviation\ (\sigma)}{mean\ (\mu)}$$

= 1 visit

Consistent reporting

Highly variable reporting

Late onboarding

Data feed dropout

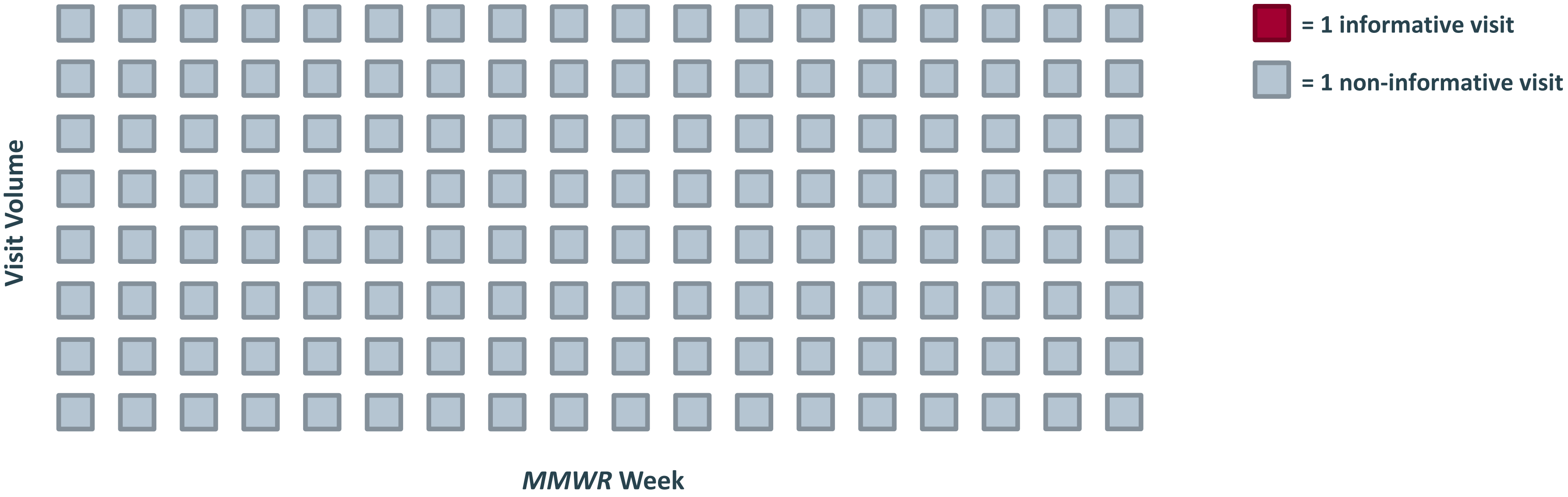
Average weekly discharge diagnosis informative (DDI) percent controls for the availability and quality of discharge diagnosis

$$DDI = {\left(\frac{Informative\,Visits}{Total\,Visits}\right)}\Big/{Number\ of\ Weeks}$$

- Measures how informative the information in the discharge diagnosis fields are over time
- Controls for diagnosis code availability and quality
- Recommend values ≥ 70%

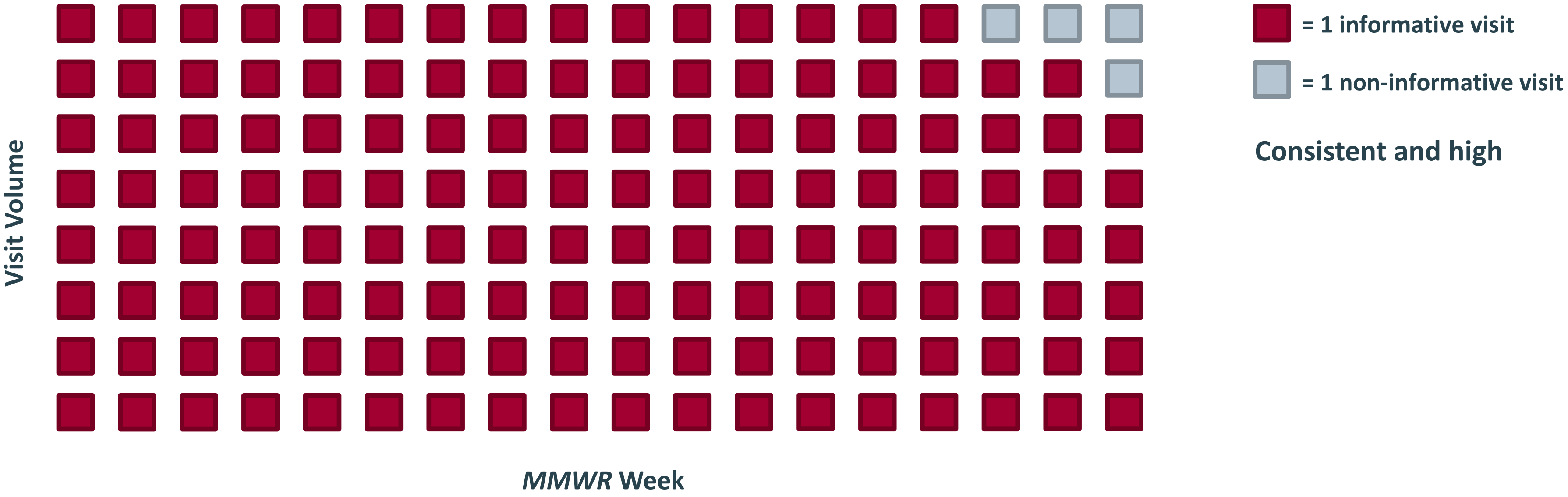
Average weekly discharge diagnosis informative (DDI) percent controls for the availability and quality of discharge diagnosis

$$DDI = \left(\frac{Informative\ Visits}{Total\ Visits}\right) \Big/ Number\ of\ Weeks$$



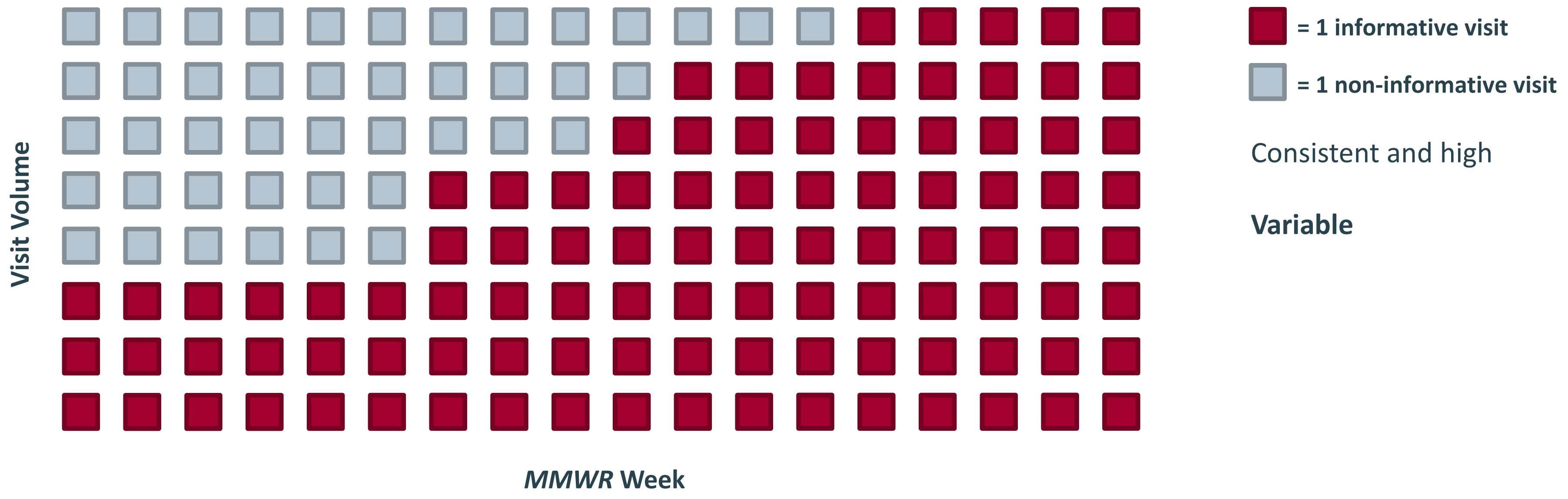
Average weekly discharge diagnosis informative (DDI) percent controls for the availability and quality of discharge diagnosis

$$DDI = \left(\frac{Informative\,Visits}{Total\,Visits}\right) \Big/ Number\ of\ Weeks$$



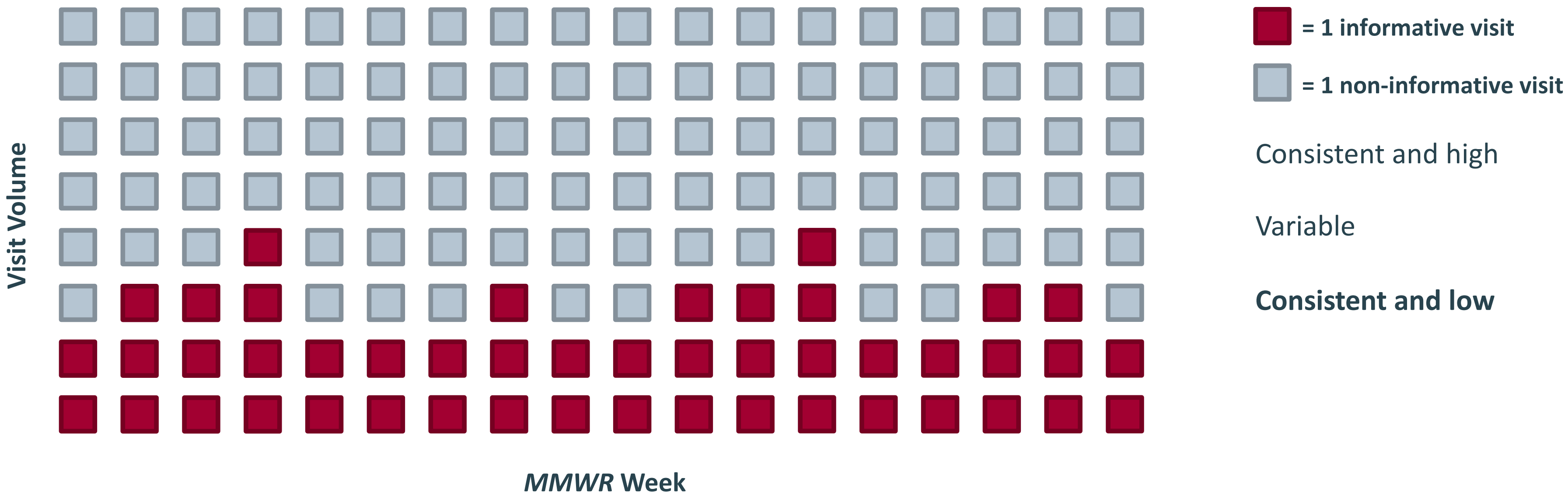
Average weekly discharge diagnosis informative (DDI) percent controls for the availability and quality of discharge diagnosis

$$DDI = \left(\frac{Informative\,Visits}{Total\,Visits}\right) \Big/ Number\ of\ Weeks$$



Average weekly discharge diagnosis informative (DDI) percent controls for the availability and quality of discharge diagnosis

$$DDI = \left(\frac{Informative\,Visits}{Total\,Visits}\right) \Big/ Number\;of\;Weeks$$

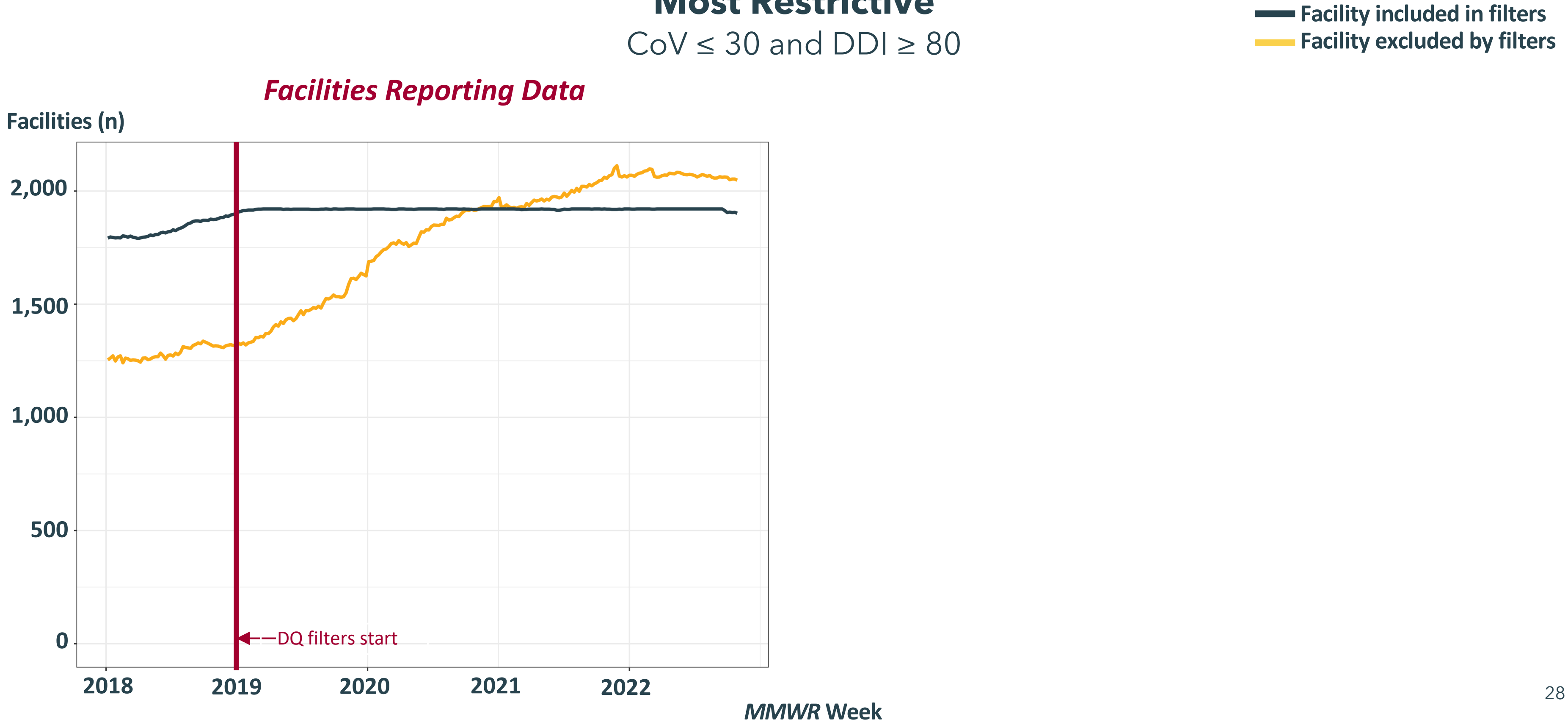


USE OF DATA QUALITY FILTERS CAN REDUCE BIAS IN RESULTS

Application of the 3-year data quality filters stabilizes the number of facilities reporting data and diagnosis codes from 2019–2022

Most Restrictive

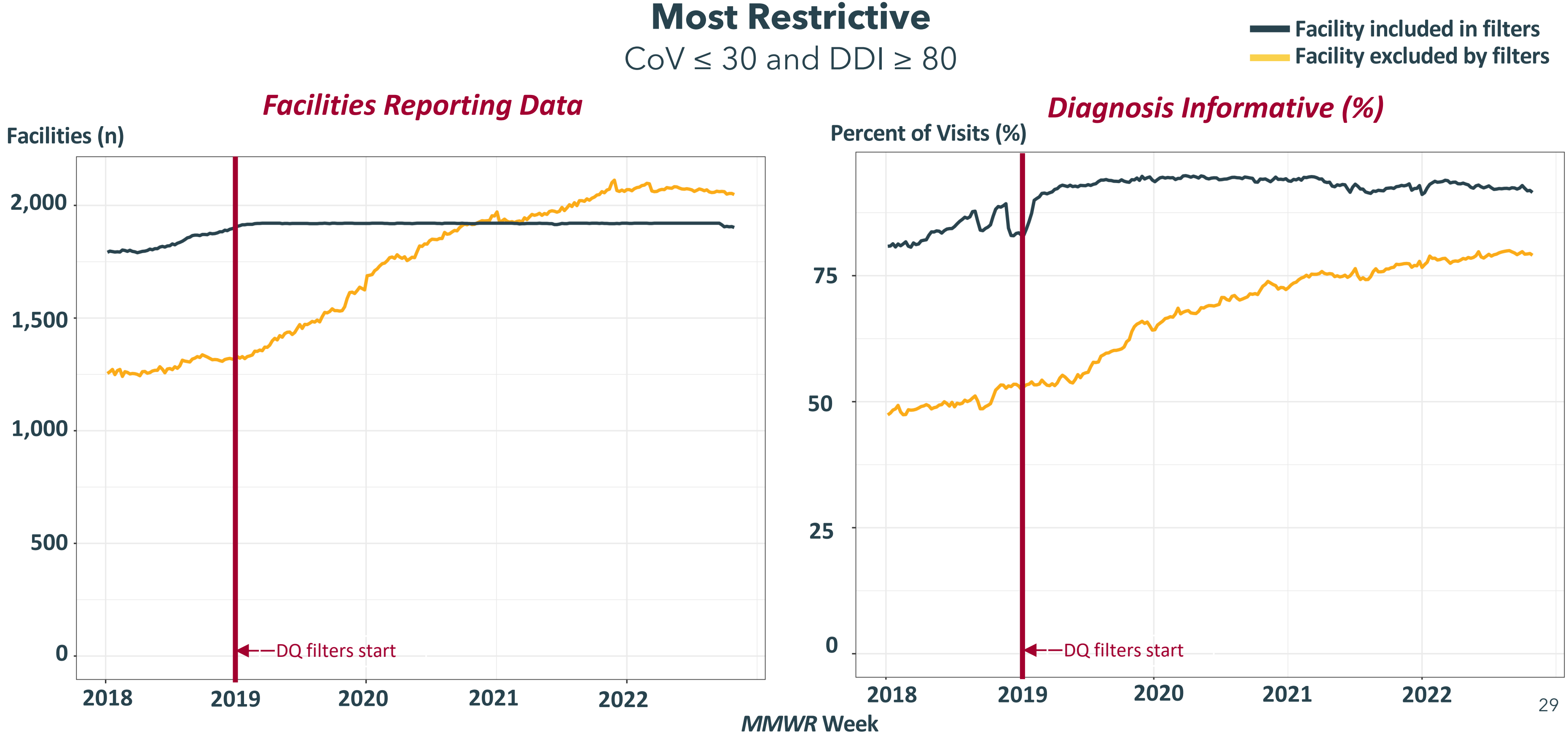
CoV ≤ 30 and DDI ≥ 80



Application of the 3-year data quality filters stabilizes the number of facilities reporting data and diagnosis codes from 2019–2022

Most Restrictive

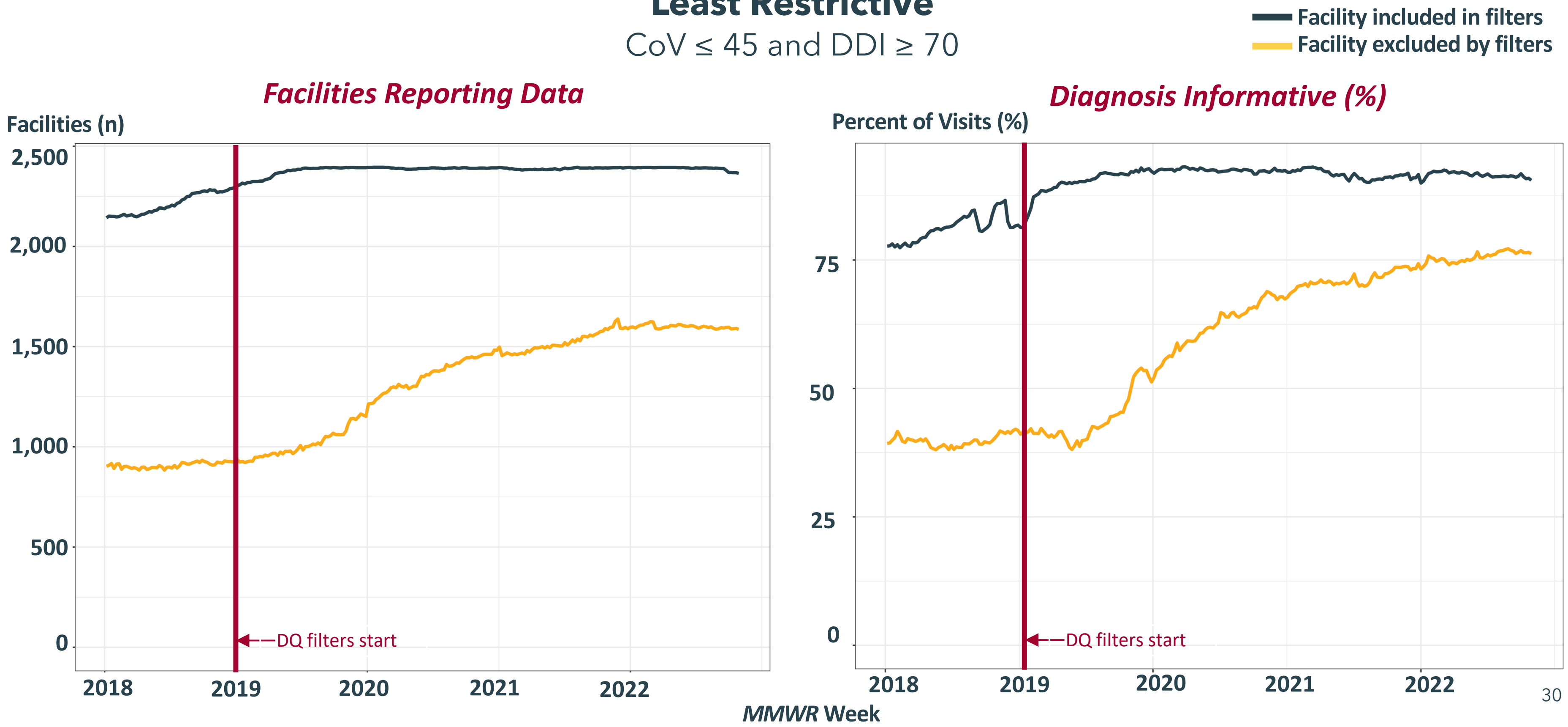
CoV ≤ 30 and DDI ≥ 80



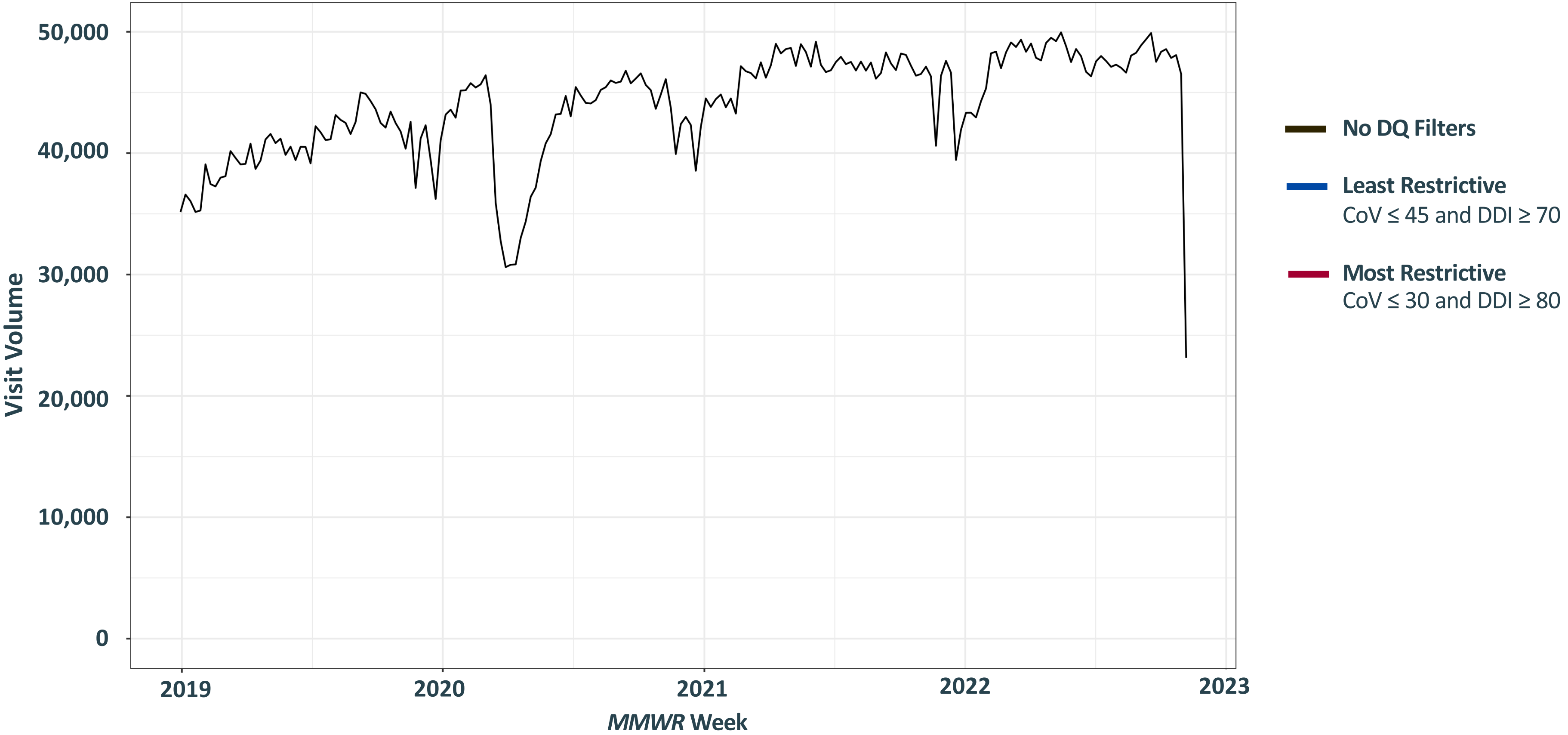
Stabilization also occurs with less restrictive filter values

Least Restrictive

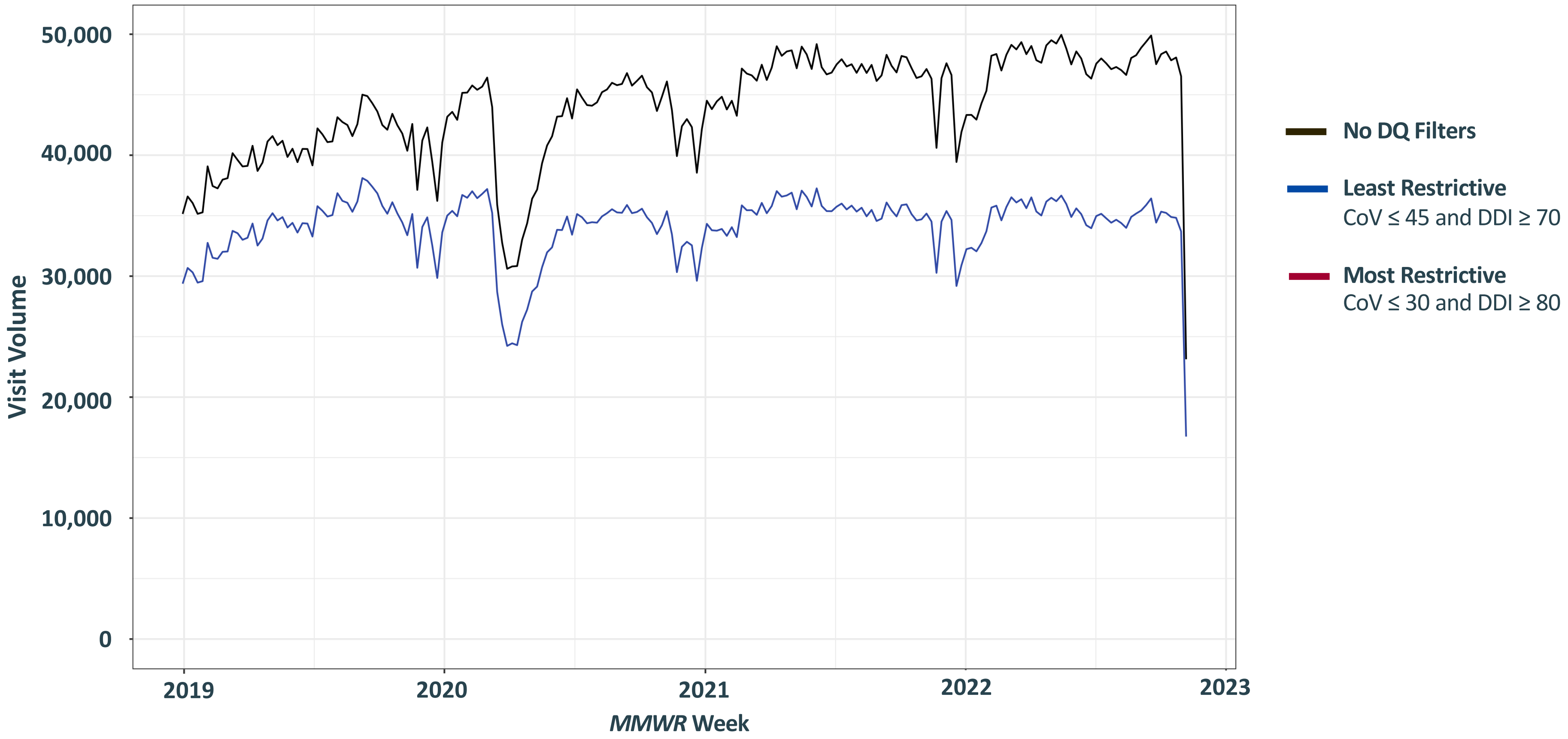
CoV ≤ 45 and DDI ≥ 70



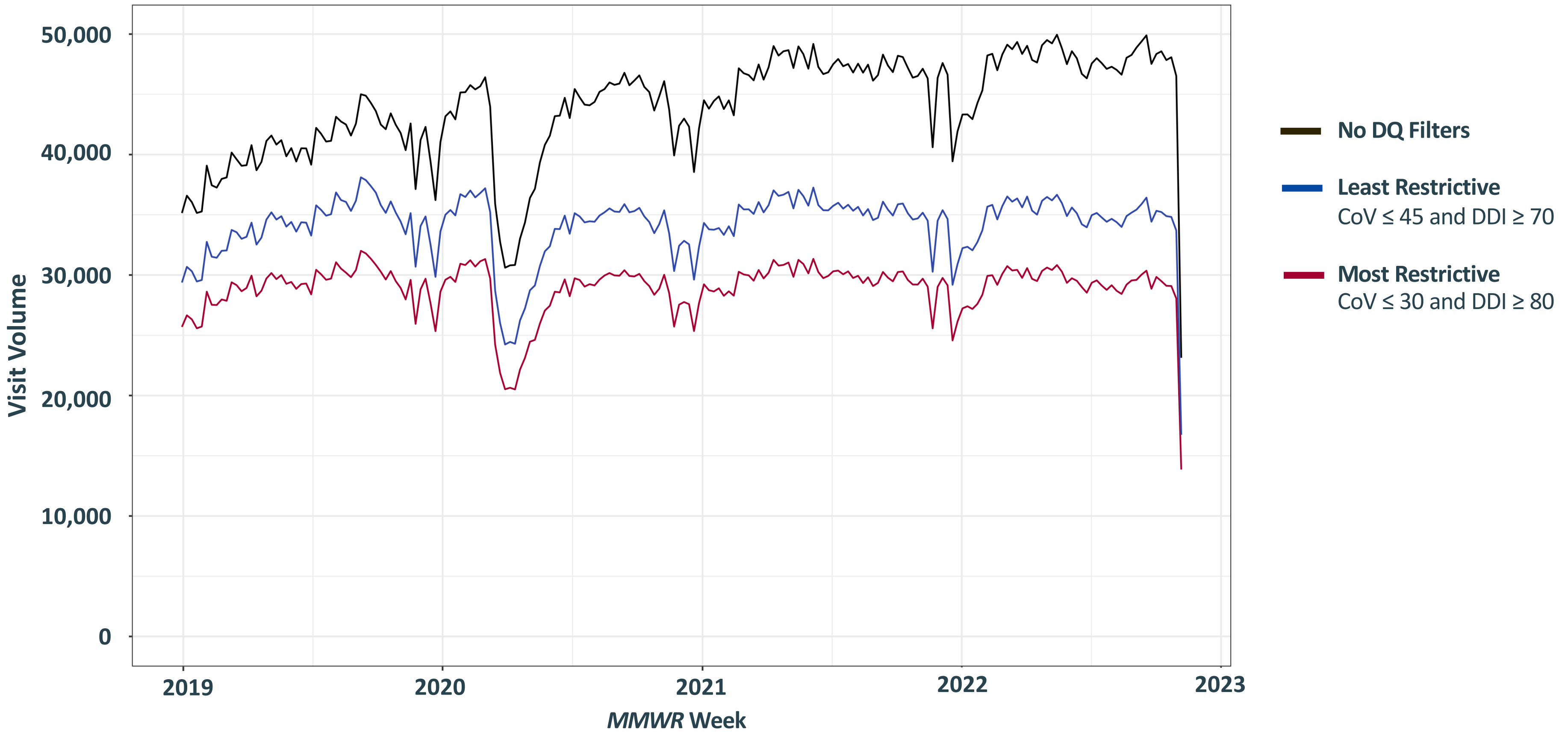
Use case: Syndrome Definition Committee (SDC) Disaster-related Mental Health v1 trends in the United States, 2019–2022



Use case: Syndrome Definition Committee (SDC) Disaster Related Mental Health v1 trends in the United States, 2019–2022

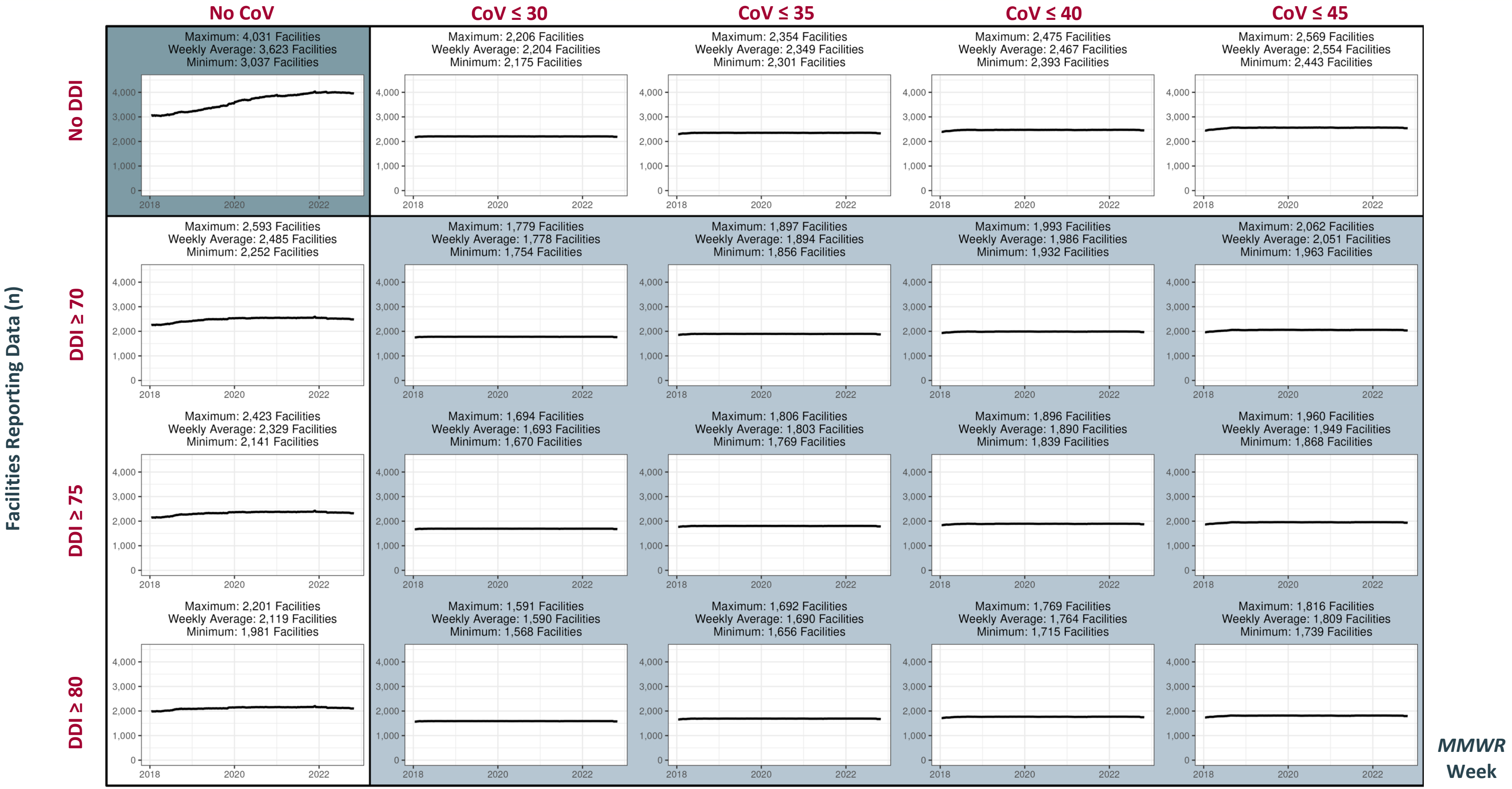


Use case: Syndrome Definition Committee (SDC) Disaster Related Mental Health v1 trends in the United States, 2019–2022

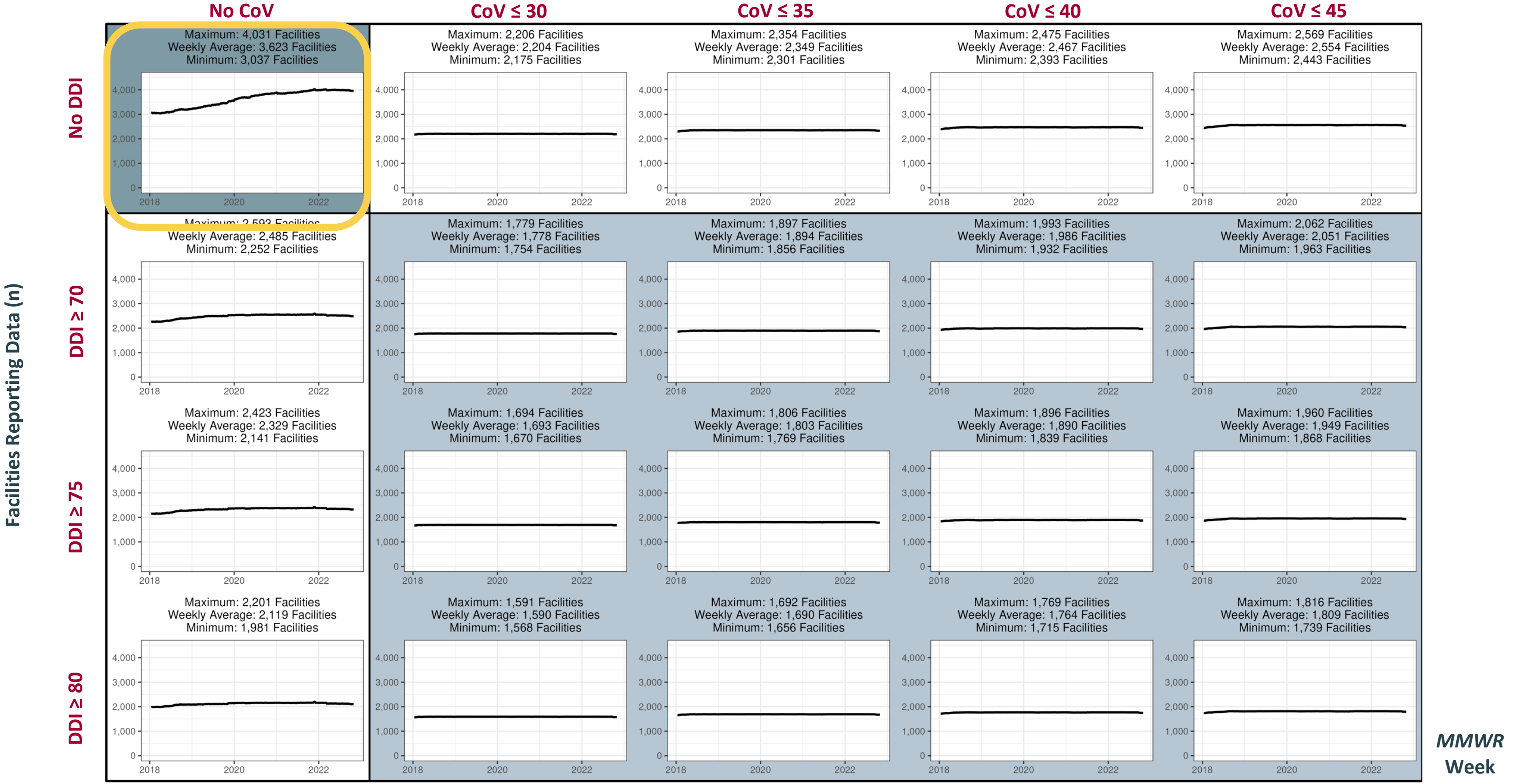


HOW TO USE DATA QUALITY FILTERS

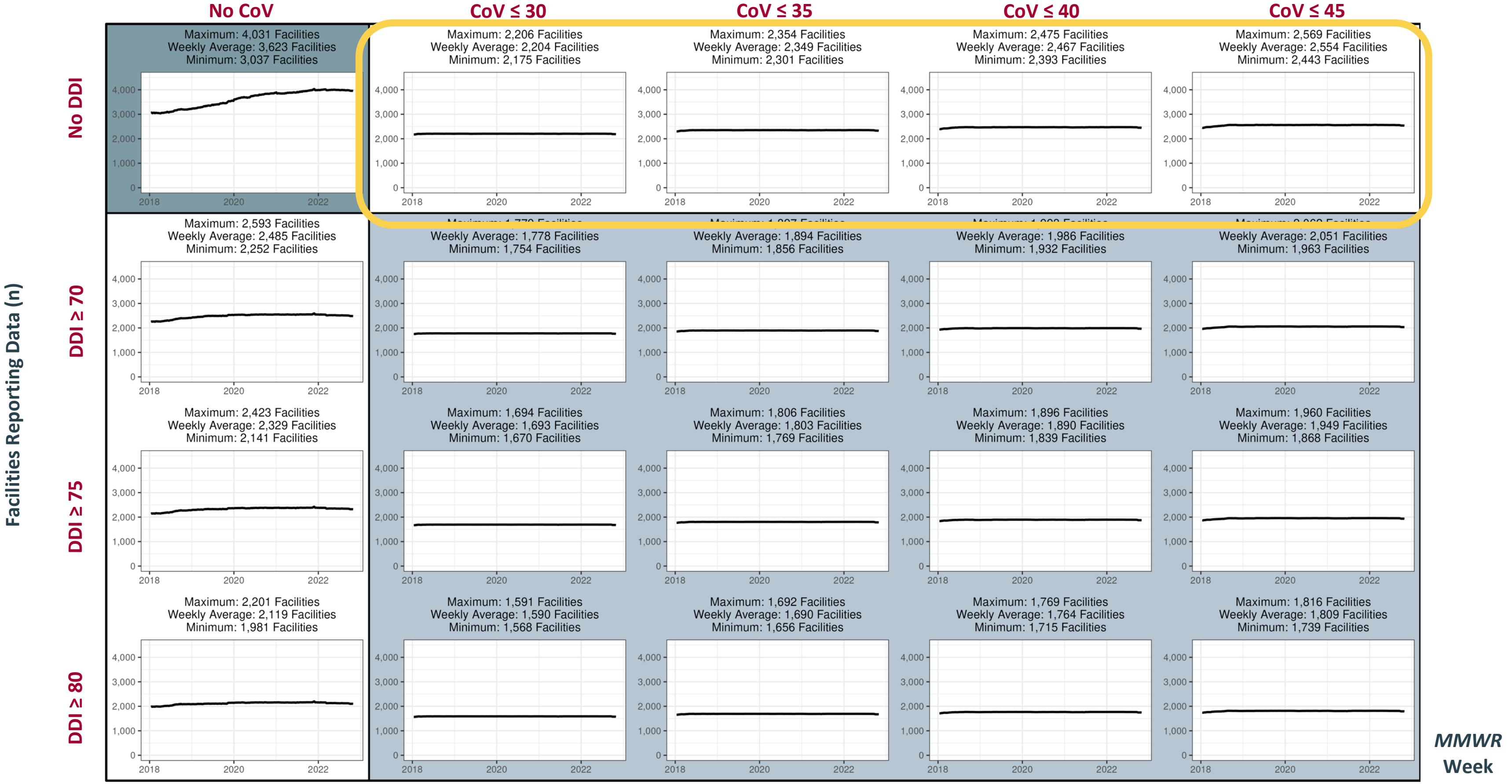
Data quality matrix visually summarizes how many facilities meet specified CoV and DDI cut-points for an analysis period



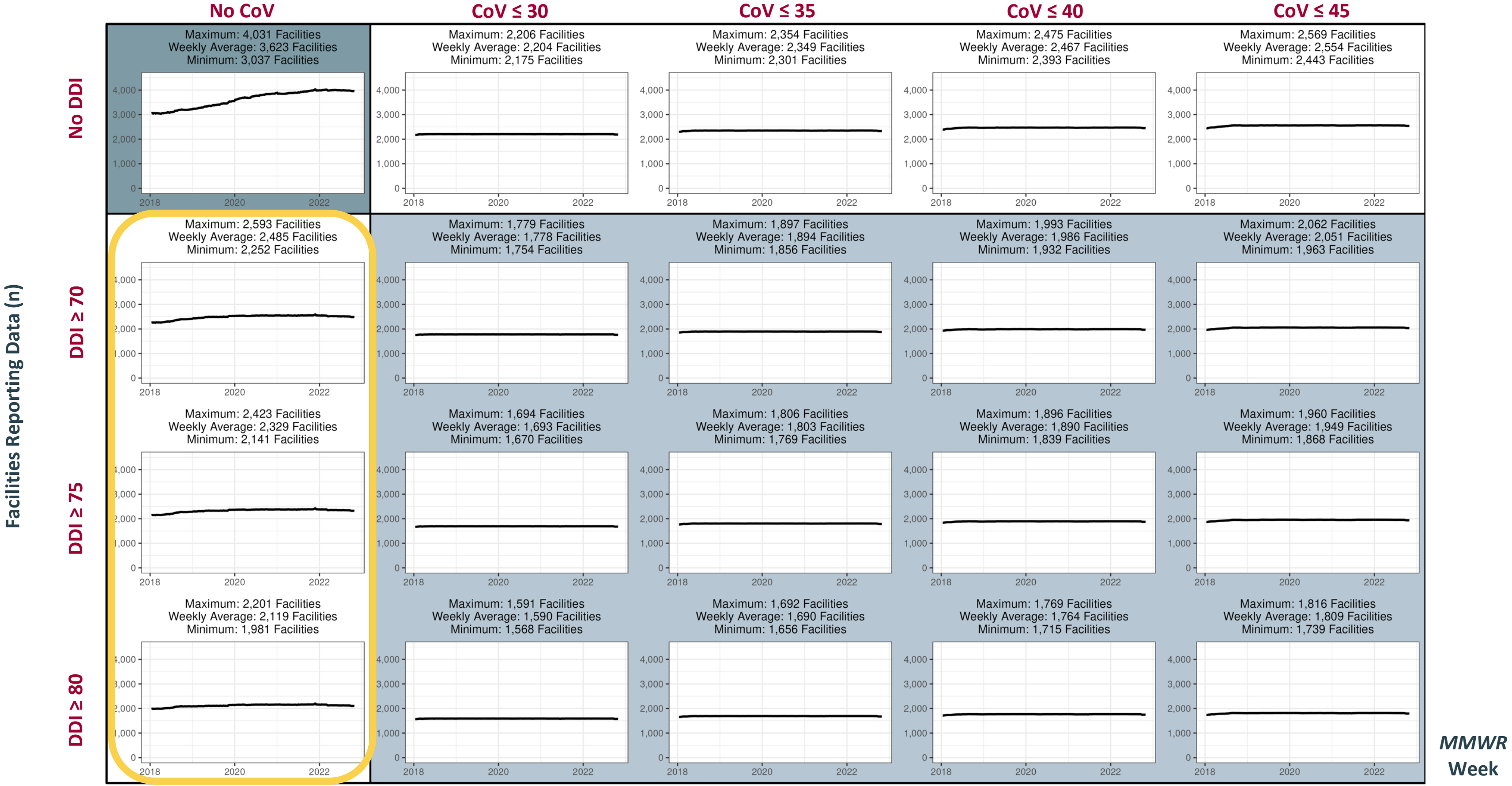
Data quality matrix visually summarizes how many facilities meet specified CoV and DDI cut-points for an analysis period



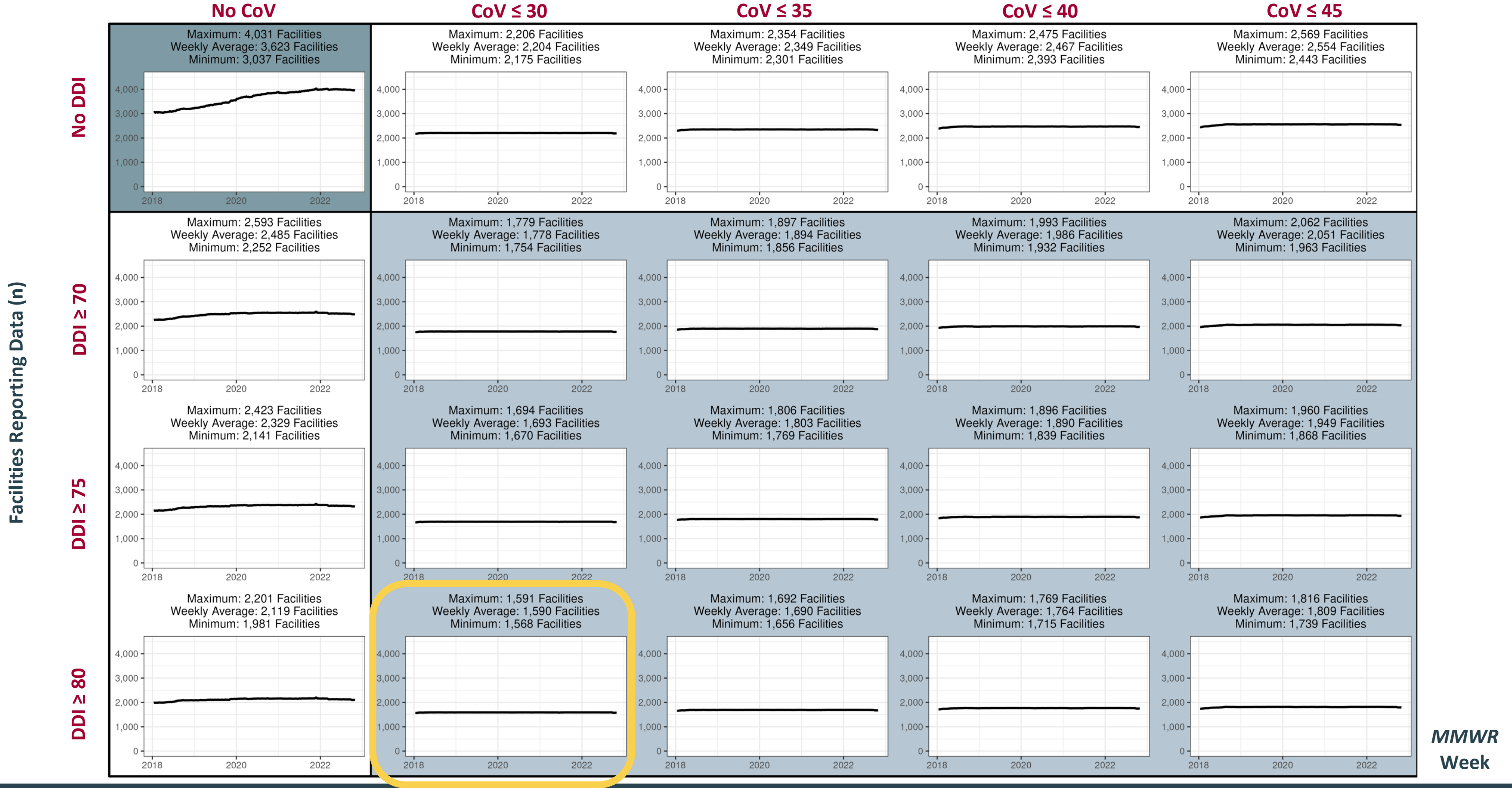
Data quality matrix visually summarizes how many facilities meet specified CoV and DDI cut-points for an analysis period



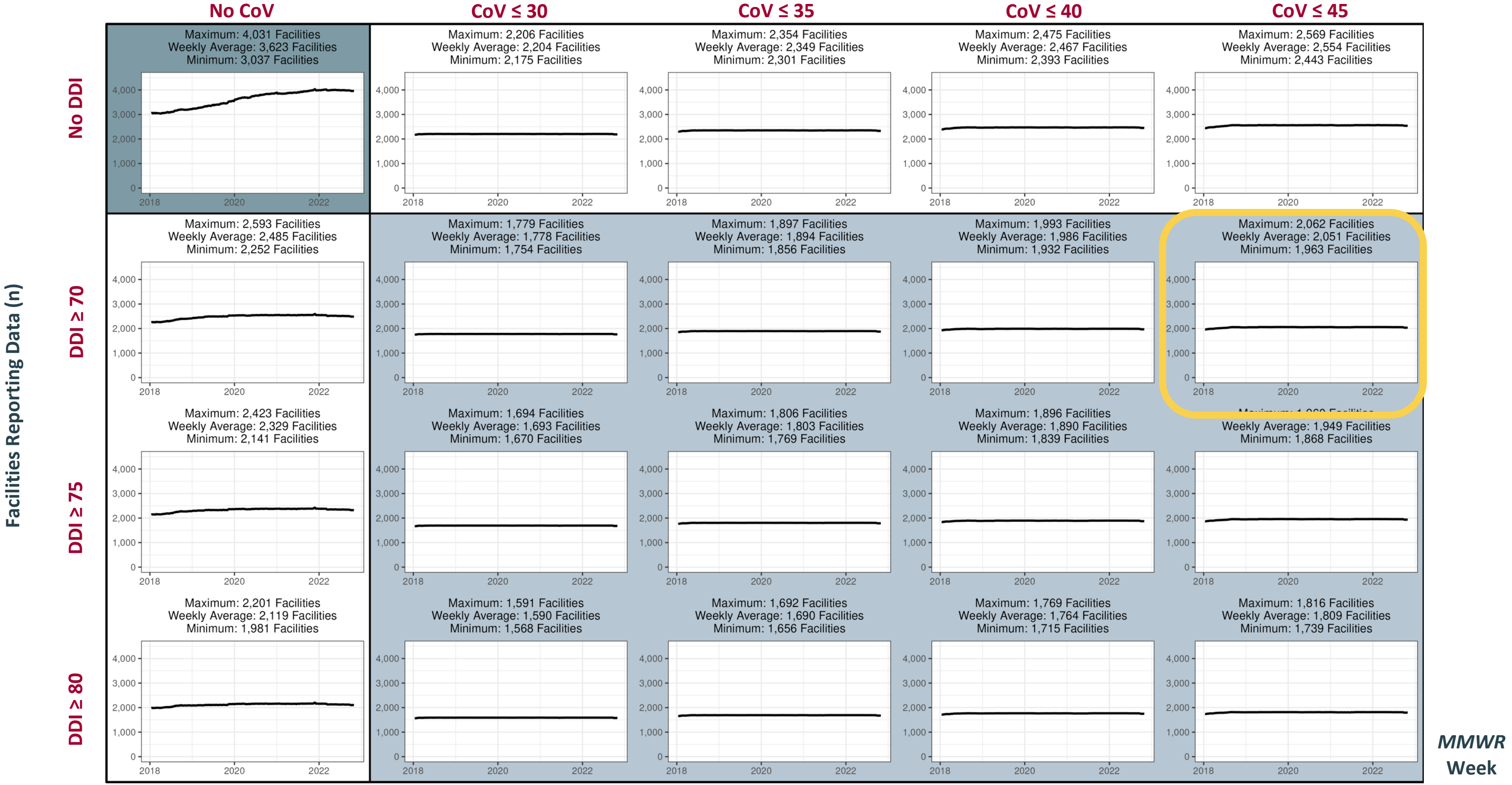
Data quality matrix visually summarizes how many facilities meet specified CoV and DDI cut-points for an analysis period



Data quality matrix visually summarizes how many facilities meet specified CoV and DDI cut-points for an analysis period



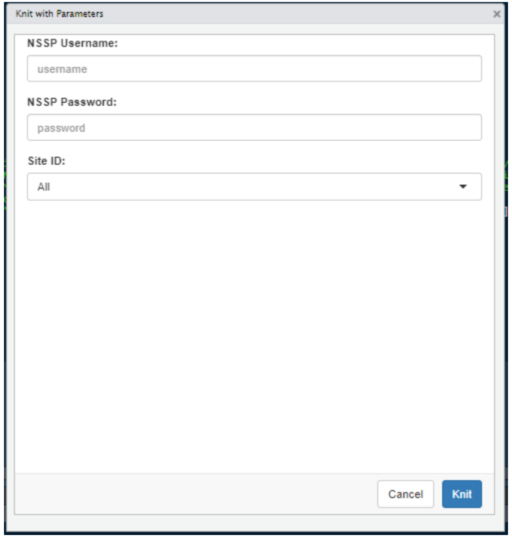
Data quality matrix visually summarizes how many facilities meet specified CoV and DDI cut-points for an analysis period



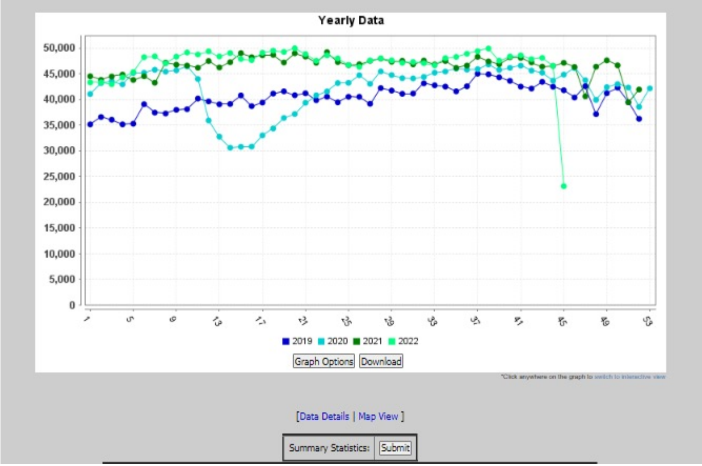
Accessing the data quality matrix template

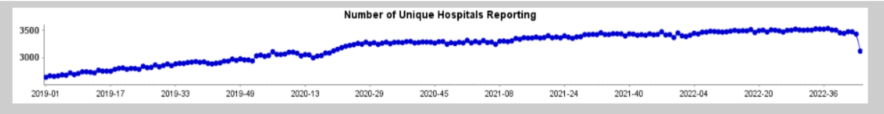
Rnssp R package and Rnssp templates

```r
# Add `dq_filters` to my existing Rnssp installation
Rnssp::add_rmd_template("dq_filters")
```



ESSENCE time series summary statistics





Summary

Data quality filters can be used to restrict analyses to facilities with consistent visit reporting and availability of DD information across years

- CoV standardizes the number of facilities included *(denominator)*
- DDI standardizes the likelihood of being categorized into a surveillance category *(numerator)*

When comparing trends across multiple years, this may reduce bias due to when facilities were onboarded and inconsistent reporting and data quality

- Not necessary for short-term surveillance comparisons

Population of facilities that meet the filters may differ from the population of facilities that do not meet the filter, and it is important to understand these differences

Acknowledgments

Centers for Disease Control and Prevention

Michael Sheppard

Aaron Kite-Powell

Karl Soetebier

Jennifer Adjemian

Johns Hopkins University Applied Physics Laboratory

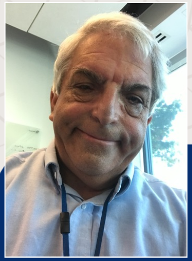
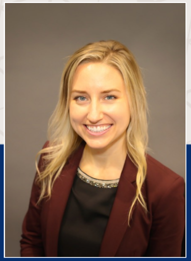
Wayne Loschen

Leif Powers

Thank you!

For more information, questions, or to walk through the data quality matrix for your site's data, please email Kelly at **qlv5@cdc.gov**.

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.

Howard Burkom, PhD
Breanna Swan Hovey, PhD
Johns Hopkins Applied Physics Laboratory (JHUAPL)

CSTE

KAHUINA
CONSULTING

Enhanced Cluster Detection Algorithm in Maryland ESSENCE for Targeted Public Health Interventions

Howard Burkom, Breanna Swan Hovey, Wayne Loschen (JHU/APL),

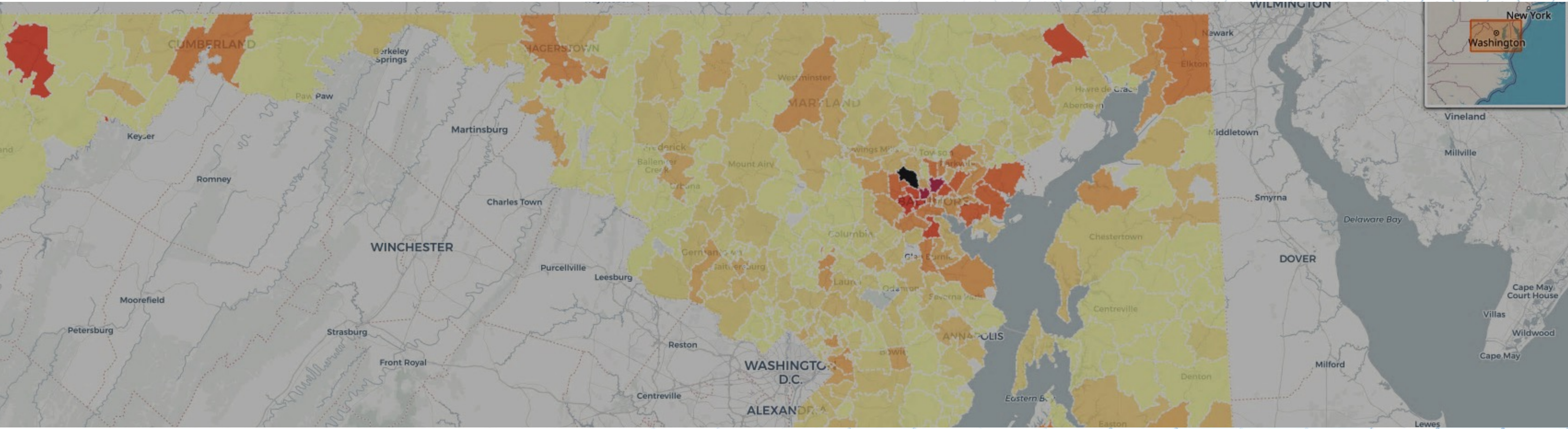
Tim Dupree, Sharmin Hossain (Maryland Dept. of Health)

Virtual 2022 Syndromic Surveillance Symposium
Session: Ensuring High-Quality Data for Public Health Use December 6, 2022

This project was made possible by Cooperative Agreement Number, 6NU17CE924961, from the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.

Outline

- Spatiotemporal Cluster Detection
- The Kulldorff Paradigm
- Comparative Methodology (space-time perm., sliding baseline)
- Empirical Study Design
- Results
- Integrated Visualization in Maryland ESSENCE
- Conclusions



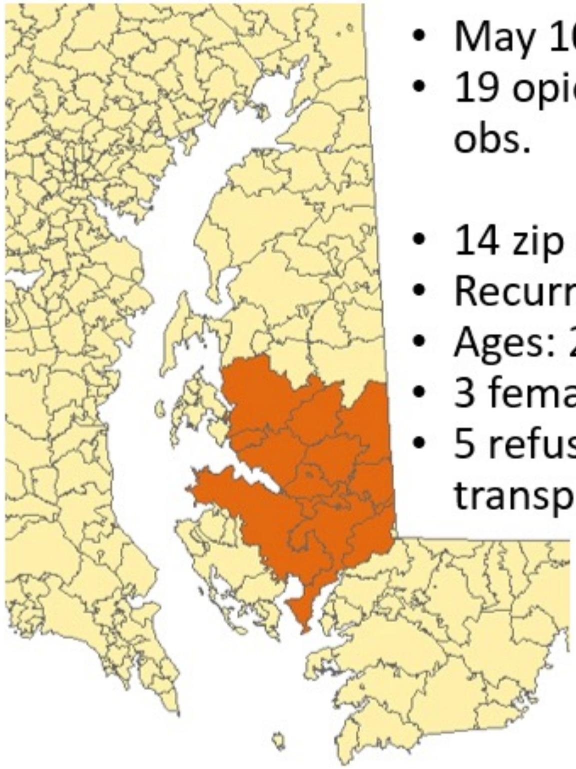
Objective:

To enable spatiotemporal cluster detection and visualization within Maryland ESSENCE for timely identification of overdose hotspots to assist coordinated public health interventions.

Importance of spatiotemporal cluster detection in biosurveillance

- Purely temporal methods can find anomalies, IF you know where to monitor case counts
  - Location of outbreak?
  - Spatial extent?
- Advantages of space-time cluster investigation:
  - Potential for early alerting
  - Tracking progression of outbreak
  - Identifying population at risk

Example from Maryland ESSENCE EMS Data, Opioid Overdose Query:

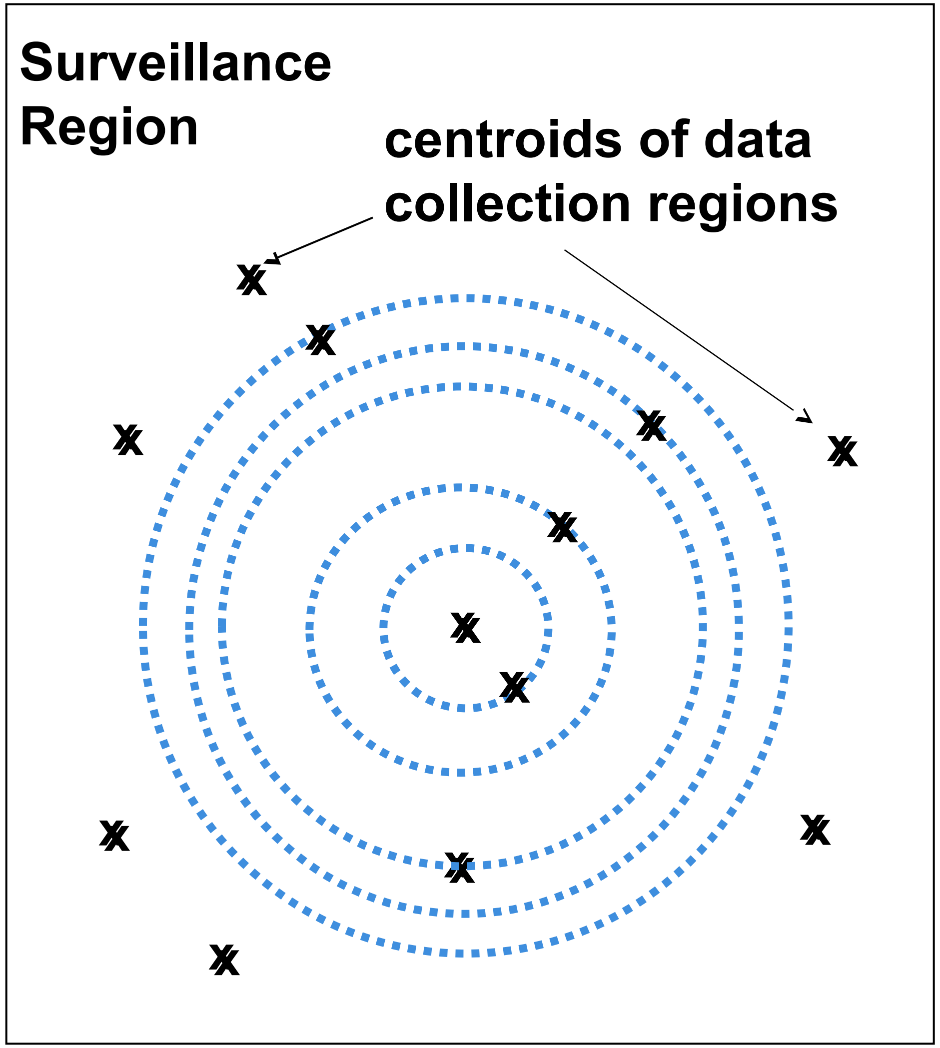


- May 10 - 16
- 19 opioid overdose related obs.

- 14 zip codes
- Recurrence Int: 2,662 days
- Ages: 23 – 59 (1 unknown)
- 3 female, 16 male
- 5 refused transport, 13 transported to hospitals

The Kulldorff Paradigm

Methodology of M. Kulldorff available in SaTScan software at http://www.satscan.org



- <u>form cylinders</u>: Bases are circles about each centroid in region A, height is time
- <u>calculate statistic:</u> for event count in each cylinder relative to entire region, within space & time limits
- <u>most significant clusters</u>: regions whose centroids form base of cylinder with maximum statistic
- <u>but how unusual is it?</u> Repeat procedure with Monte Carlo trials, compare max statistic to maxima of each of these
- How to calculate expected number of cases for a location? What is the background spatial distribution?
- Deciding significance: Once a cluster with a maximum LR is found, how to choose representative data sets for random trials

Approach and Data

The Maryland Department of Health collaborated with the Johns Hopkins Applied Physics Lab to develop a spatiotemporal cluster detection capability within ESSENCE

- Built into the MDH-ESSENCE platform for identification of significant clusters spanning multiple days.
- Also added a novel Cluster Explorer feature that summarizes the demographic and clinical features for any cluster alert of potential interest.

<u>Outcome Datasets used for Testing</u>

- COVID-like illness
  - Used COVID-like illness (CLI) query ‘CDC CC with CLI DD and Coronavirus DD V2’, which does not require a COVID-19 diagnosis and includes many ED visits from before the pandemic.
- Opioid overdose
  - Used ‘CDC Opioid Overdose V3’ query to pull cases from MDH ESSENCE
- For both outcomes
  - Selected data for interval 7/1/2019 – 12/31/2019
  - Input data for cluster determination:
    - Number of ED Cases for each zip code, each day

Testing Candidate Clusters: Alternate Methods to Form Simulated Datasets for Monte Carlo Hypothesis Testing

Direct Adjustment
- For each case in baseline, make random location draw based on marginal location totals
- Adds buffer between current cases and baseline
- Omitting current cases from baseline means current locations may not be represented in baseline – use practical adjustment
- Programmed in ESSENCE

Direct Adjustment Method **Space-time Permutation Method**

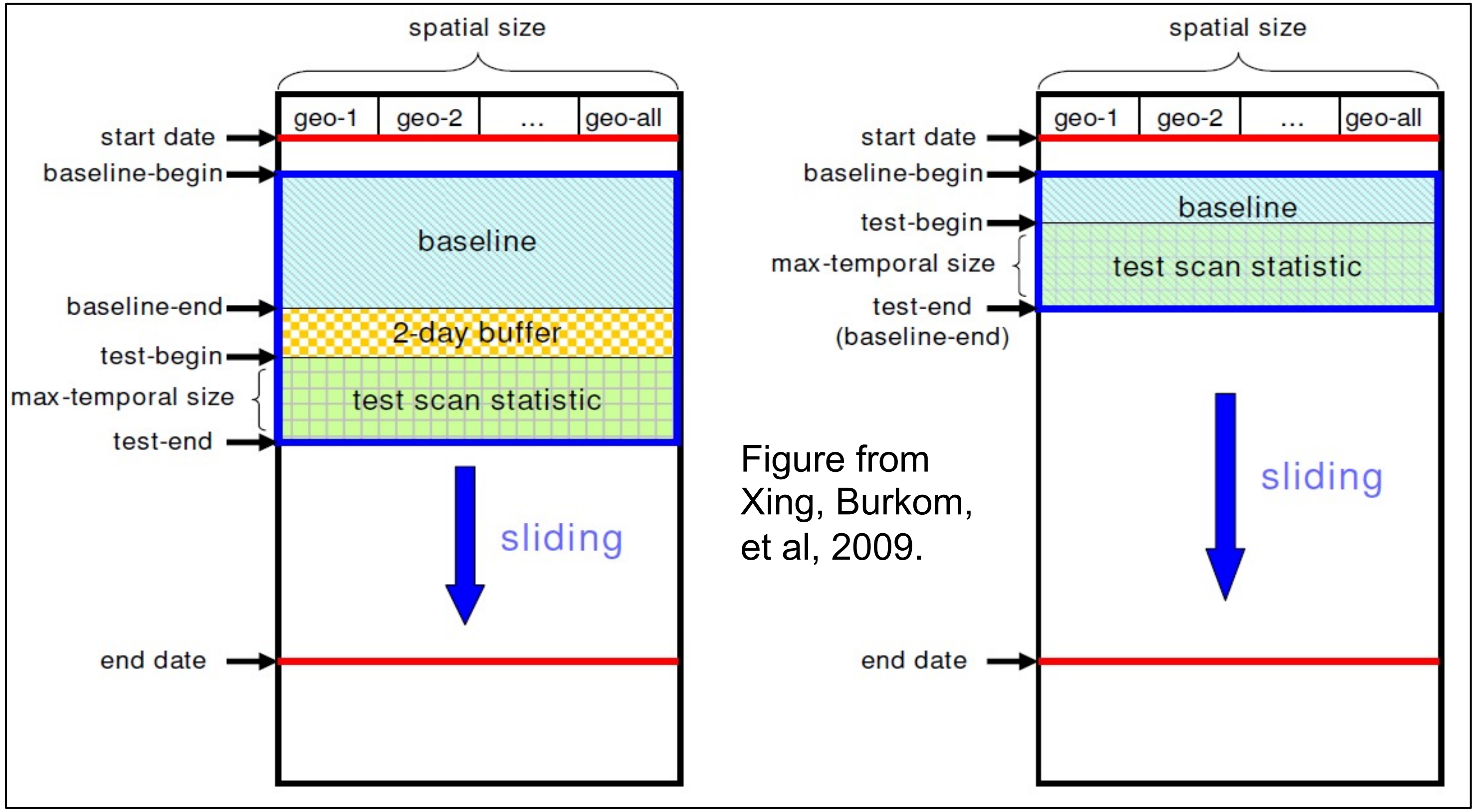


Figure from Xing, Burkom, et al, 2009.

Space-time Permutation Method
- Most commonly used method for prospective syndromic cluster monitoring in SaTScan
- “shuffle the dates/times and assign them to the original set of case locations, ensuring that both the spatial and temporal marginals are unchanged” (from Kulldorff, Heffernan, et al., 2004)
- Includes the current day’s data in shuffling

After both methods, “the most likely cluster is calculated for each simulated dataset in exactly the same way as for the real data. Statistical significance is evaluated using Monte Carlo hypothesis testing”

Validation of Enhanced ESSENCE Algorithm

- The ESSENCE algorithm using direct adjustment was validated by comparing its performance to the SaTScan space-time permutation (S-TP) method, using plausible, simulated scenarios.
- Scenarios were formed with synthetic clusters by inserting case counts into authentic background data.
- For each replication of each scenario,
  - ran both ESSENCE algorithm and SaTScan space-time permutation methods
  - identified which reported clusters (if any) contained zip codes with inserted counts
  - recorded minimum p-value among those clusters

- Inserted cases to represent clusters of varying parameters:
  - Number of cases
  - Geographical width of the cluster (miles)
  - Duration (days)
  - Background data frequency
- Conducted 10 replications for each scenario (next slide) with center zip codes chosen randomly in each replication, for a total of 1530 algorithm tests.
- Compared results at significance thresholds of 0.001, 0.01, 0.05, 0.10, 0.20.

Study Design

Selected scenarios within expected detection capability

Sparse Data: Opioid Overdose Outcome Scenarios

		Duration (days)		
Case Count	Spatial Extent (mi)	1	4	7
5 Cases	0	x	x	x
	0.5	x	x	x
	2	x	x	x
	5			
10 Cases	0	x	x	x
	0.5	x	x	x
	2	x	x	x
	5	x	x	x
20 Cases	0			
	0.5	x	x	x
	2	x	x	x
	5	x	x	x
50 Cases	0			
	0.5			
	2	x	x	x
	5	x	x	x

Rich Data: COVID-like Illness Outcome Scenarios

		Urban			**Suburban**			**Rural**		
		Duration (days)			Duration (days)			Duration (days)		
Case Count	Spatial Extent (mi)	1	4	7	1	4	7	1	4	7
3 Cases	0							x	x	x
	0.5							x	x	x
	2							x	x	x
	5							x	x	x
5 Cases	0	x	x		x	x	x	x	x	x
	0.5	x	x		x	x	x	x	x	x
	2	x	x		x	x	x	x	x	x
	5	x	x		x	x	x	x	x	x
10 Cases	0	x	x	x	x	x	x			
	0.5	x	x	x	x	x	x	x	x	x
	2	x	x	x	x	x	x	x	x	x
	5	x	x	x	x	x	x	x	x	x
20 Cases	0	x	x	x		x	x			
	0.5	x	x	x	x	x	x		x	x
	2	x	x	x	x	x	x	x	x	x
	5	x	x	x	x	x	x	x	x	x
50 Cases	0	x	x	x	x	x	x			
	0.5	x	x	x	x	x	x			
	2	x	x	x	x	x	x			
	5	x	x	x	x	x	x			

Results using Opioid Query returns as Outcome

How do the simulation methods compare at detecting injected clusters?

				Number of Reps Detected at P-Value <= 0.001			Number of Reps Detected at P-Value <= 0.005			Number of Reps Detected at P-Value <= 0.01			Number of Reps Detected at P-Value <= 0.05			Number of Reps Detected at P-Value <= 0.1			Number of Reps Detected at P-Value <= 0.2		
	Scenario	Radius	Cases	ESSENCE Direct Adj.	SaTScan S-T Perm.	Δ	ESSENCE Direct Adj.	SaTScan S-T Perm.	Δ	ESSENCE Direct Adj.	SaTScan S-T Perm.	Δ	ESSENCE Direct Adj.	SaTScan S-T Perm.	Δ	ESSENCE Direct Adj.	SaTScan S-T Perm.	Δ	ESSENCE Direct Adj.	SaTScan S-T Perm.	Δ
1-Day Injects	1	0	5	2	2	0	5	4	1	6	4	2	6	6	0	7	6	1	8	6	2
	2	0.5	5	6	6	0	6	7	-1	6	7	-1	9	7	2	10	9	1	10	10	0
	3	2	5	3	3	0	3	3	0	3	3	0	4	5	-1	4	5	-1	5	6	-1
	4	0	10	9	9	0	9	9	0	9	10	-1	9	10	-1	9	10	-1	10	10	0
	5	0.5	10	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0
	6	2	10	6	6	0	7	6	1	7	6	1	7	6	1	8	7	1	8	8	0
	7	5	10	5	4	1	5	5	0	5	5	0	7	6	1	7	6	1	7	7	0
	8	0.5	20	10	10	0	10	10	0	10	10	0	10	10 Δ	0	10	10	0	10	10	0
	9	2	20	8	8	0	9	9	0	10	10	0	10	10	0	10	10	0	10	10	0
	10	5	20	7	6	1	7	8	-1	7	8	-1	9	9	0	9	9	0	10	9	1
	11	2	50	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0
	12	5	50	10	9	1	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0
4-Day Injects	13	0	5	3	4	-1	5	5	0	5	6	-1	5	7	-2	6	7	-1	6	8	-2
	14	0.5	5	3	0	3	3	3	0	3	4	-1	3	5	-2	4	6	-2	5	7	-2
	15	2	5	1	1	0	1	1	0	1	1	0	3	1	2	3	2	1	4	2	2
	16	0	10	8	8	0	8	8	0	9	8	1	9	8	1	10	9	1	10	10	0
	17	0.5	10	6	6	0	7	6	1	7	6	1	10	10	0	10	10	0	10	10	0
	18	2	10	5	4	1	5	6	-1	6	6	0	7	7	0	8	9	-1	9	9	0
	19	5	10	4	3	1	5	3	2	5	3	2	6	4	2	7	5	2	7	5	2
	20	0.5	20	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0
	21	2	20	9	9	0	9	9	0	9	9	0	10	10	0	10	10	0	10	10	0
	22	5	20	5	5	0	5	5	0	5	5	0	5	5	0	7	5	2	7	5	2
	23	2	50	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0
	24	5	50	5	4	1	8	5	3	8	5	3	10	8	2	10	9	1	10	9	1
7-Day Injects	25	0	5	2	0	2	3	2	1	4	3	1	6	5	1	8	5	3	8	6	2
	26	0.5	5	3	2	1	6	4	2	6	5	1	6	7	-1	8	7	1	9	7	2
	27	2	5	0	0	0	0	0	0	1	0	1	3	1	2	3	2	1	3	2	1
	28	0	10	4	3	1	6	6	0	7	7	0	8	7	1	9	8	1	9	9	0
	29	0.5	10	4	3	1	6	6	0	7	7	0	8	7	1	9	8	1	9	8	1
	30	2	10	4	3	1	4	4	0	4	4	0	4	4	0	6	4	2	6	4	2
	31	5	10	0	0	0	0	0	0	0	0	0	1	0	1	2	0	2	2	2	0
	32	0.5	20	8	9	-1	8	9	-1	8	9	-1	9	9	0	9	9	0	9	9	0
	33	2	20	6	7	-1	7	8	-1	7	8	-1	8	8	0	8	9	-1	8	9	-1
	34	5	20	2	2	0	3	3	0	4	3	1	4	4	0	4	5	-1	4	8	-4
	35	2	50	8	10	-2	10	10	0	10	10	0	10	10	0	10	10	0	10	10	0
	36	5	50	5	6	-1	6	6	0	6	6	0	7	9	-2	9	9	0	9	9	0

Table

- Displays the number of replications in which the algorithm detected the injected cluster at the set p-value (top row).
- Colored columns compare ESSENCE to SaTScan (# ESSENCE detected - # SaTScan).
  - Green = ESSENCE outperformed SaTScan
  - Orange/Red = SaTScan outperformed ESSENCE

Results vary

- ESSENCE as good or better than SatScan in Rows 5-12 (10 had issue at .01) and Rows 15 – 31 (18 had issue at .1, 26 had issue at .05)
- ESSENCE underperformed SatScan:
  - Low cases across 4 days, small radius (R13, 14)
  - High cases across 7 days and larger radius (R32, 33, 35, 36)
- Effect of p-value
  - ESSENCE underperforms with less sensitive p-value (p value increases) Rows 3, 33, 34

ESSENCE Algorithm Results

Number of Replications (of 10) that the inserted cluster was detected with a p-value <=0.5

Color scale: 0-3 | 4-7 | 8-10

Opioid Data

		Duration (days)		
Case Count	Spatial Extent	1	4	7
5 Cases	0	6	5	6
	0.5	9	3	6
	2	4	3	3
	5			
10 Cases	0	9	9	8
	0.5	10	10	8
	2	7	7	4
	5	7	6	1
20 Cases	0			
	0.5	10	10	9
	2	10	10	8
	5	9	5	4
50 Cases	0			
	0.5			
	2	10	10	10
	5	10	10	7

CLI Data

		Urban			Suburban			Rural		
		Duration			Duration			Duration		
Case Count	Spatial Extent (mi)	1	4	7	1	4	7	1	4	7
3 Cases	0							9	3	4
	0.5							8	5	5
	2							8	4	2
	5							3	4	5
5 Cases	0	5	5		10	6	5	9	8	7
	0.5	6	1		10	6	7	10	8	5
	2	1	3		9	9	4	10	8	6
	5	0	3		10	7	3	8	5	2
10 Cases	0	10	5	8	10	10	9			
	0.5	8	7	5	10	10	10	10	10	10
	2	7	5	2	10	10	9	10	10	10
	5	6	3	1	10	9	9	10	10	8
20 Cases	0	10	10	9		10	10			
	0.5	10	10	10	10	10	10		10	10
	2	9	7	8	10	10	10	10	10	10
	5	9	7	3	10	10	10	10	10	10
50 Cases	0	10	10	10	10	10	10			
	0.5	10	10	10	10	10	10			
	2	10	10	10	10	10	10			
	5	10	8	9	10	10	10			

Analysis In Progress to Determine Best Practice for Staggered Application

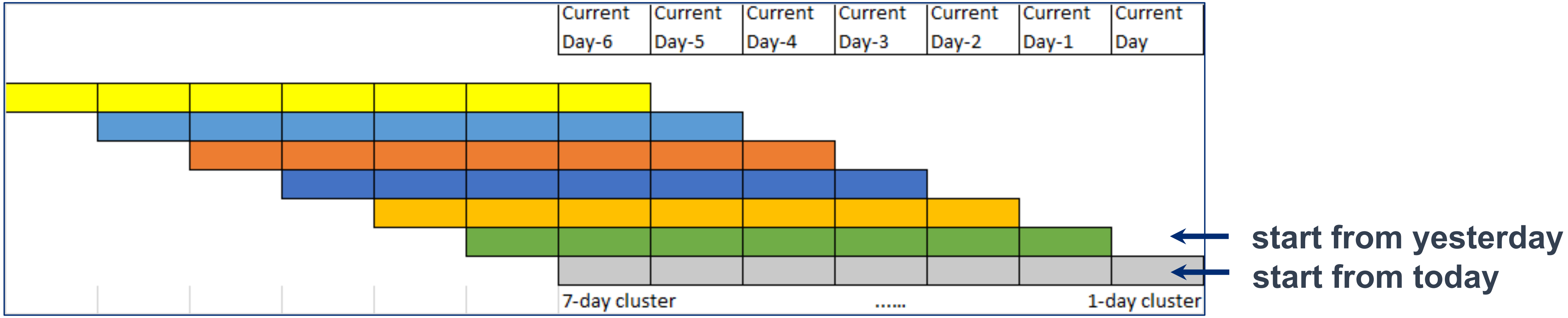
Issue of variable reporting timeliness among hospitals that provide daily data to ESSENCE

- Even within individual hospitals
- Conventional procedure in ESSENCE for alerting algorithms
  - Run univariate algorithms for <u>each of the last 7 days</u>
  - Spike occurring 2-3 days ago may be completely missed – data late!
    - A temporary system outage, or outbreak itself may cause late reporting!

Analysis In Progress to Determine Best Practice for Staggered Application

Issue of variable reporting timeliness among hospitals that provide daily data to ESSENCE

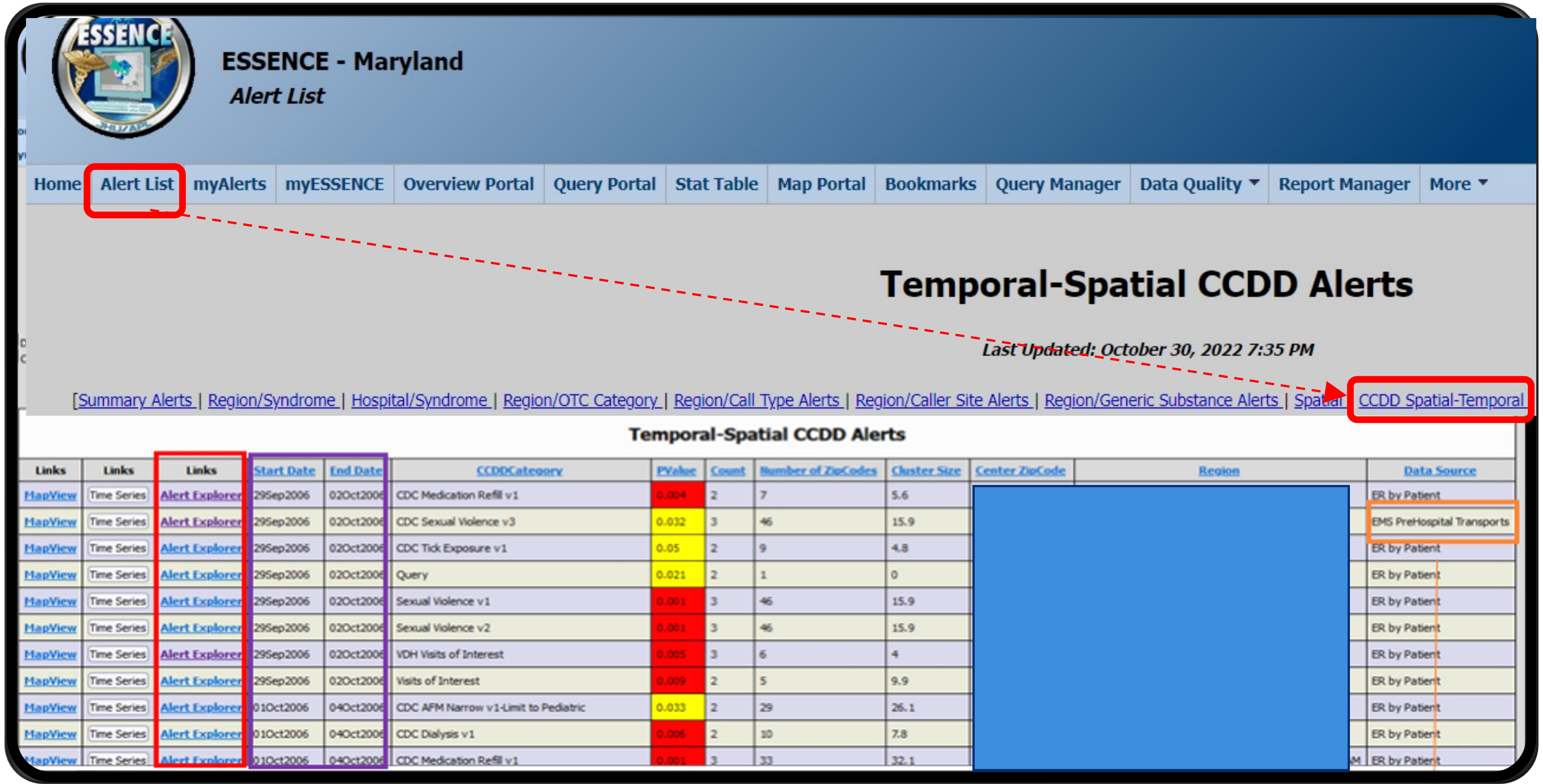
- Even within individual hospitals
- Conventional procedure in ESSENCE for alerting algorithms
  - Run univariate algorithms for <u>each of the last 7 days</u>
  - Spike occurring 2-3 days ago may be completely missed – data late!
    - A temporary system outage, or outbreak itself may cause late reporting!



Applying this procedure to spatiotemporal clustering → excessive repeated cluster alerts

- In-depth discussions with MD Health Dept. epidemiologists to discard repeated clusters
- Testing criteria based on: maximal case counts, minimal time span to finalize cluster alert list
- <u>Effective cluster alert rate will depend on the final criteria</u>

Inspecting Current Clusters in Maryland ESSENCE Alert List



Links	Links	Links	Start Date	End Date	CCDDCategory	PValue	Count	Number of ZipCodes	Cluster Size	Center ZipCode	Region	Data Source
MapView	Time Series	Alert Explorer	29Sep2006	02Oct2006	CDC Medication Refill v1	0.004	2	7	5.6			ER by Patient
MapView	Time Series	Alert Explorer	29Sep2006	02Oct2006	CDC Sexual Violence v3	0.032	3	46	15.9			EMS PreHospital Transports
MapView	Time Series	Alert Explorer	29Sep2006	02Oct2006	CDC Tick Exposure v1	0.05	2	9	4.8			ER by Patient
MapView	Time Series	Alert Explorer	29Sep2006	02Oct2006	Query	0.021	2	1	0			ER by Patient
MapView	Time Series	Alert Explorer	29Sep2006	02Oct2006	Sexual Violence v1	0.001	3	46	15.9			ER by Patient
MapView	Time Series	Alert Explorer	29Sep2006	02Oct2006	Sexual Violence v2	0.001	3	46	15.9			ER by Patient
MapView	Time Series	Alert Explorer	29Sep2006	02Oct2006	VDH Visits of Interest	0.005	3	6	4			ER by Patient
MapView	Time Series	Alert Explorer	29Sep2006	02Oct2006	Visits of Interest	0.009	2	5	9.9			ER by Patient
MapView	Time Series	Alert Explorer	01Oct2006	04Oct2006	CDC AFM Narrow v1-Limit to Pediatric	0.033	2	29	26.1			ER by Patient
MapView	Time Series	Alert Explorer	01Oct2006	04Oct2006	CDC Dialysis v1	0.006	2	10	7.8			ER by Patient
MapView	Time Series	Alert Explorer	01Oct2006	04Oct2006	CDC Medication Refill v1	0.001	3	33	32.1			ER by Patient

Alerts for a date range (start / end) vs single date

Links to new Alert Explorer Visualization

Ability to combine ED & EMS data

- Developed a new view in ESSENCE: an Alert Explorer.
- Enhanced alert list with several new columns
  - *Alert Explorer* column enabling the user to drill down for summarized details of each alerted cluster.
  - Columns for the start and end date of each cluster.
  - Field added to indicate data source.

Draft details screen for Alert Explorer for alerted spatiotemporal clusters

- Coding and implementation are complete.
- Evaluation of the visualization components by MDH staff and users is underway.
- Also included in the visualization (not shown):
  - Full data details for cluster cases
  - Counts of selected terms in chief complaint (e.g. fentanyl, heroin…)

Conclusions

- Adapted cluster detection capability in Maryland ESSENCE to spatiotemporal version set for anomalous clusters of 1-7 days using ED CCDD categories and EMS queries
- Validated method with an extensive study design varying number of cases, number of days, spatial extent, outcome type
- Added an Anomaly Explorer capability to show details of cluster cases for investigation
  - Allowing extended drill-down within the ESSENCE application
- Next steps:
  - Finalize alert list reduction criteria
  - Apply and test on EMS queries → investigate combined ED and EMS clustering
  - Refine algorithm and visualization based on user experience
  - Accelerate algorithm to provide on-demand cluster capability for ad hoc queries
  - Provide for other ESSENCE installations?

References

1. Xing J, Burkom H, Moniz L, Edgerton J, Leuze M, Tokars J: **Evaluation of sliding baseline methods for spatial estimation for cluster detection in the biosurveillance system.** *Int J Health Geogr* 2009, **8:45.:**10.1186/1476-1072X-1188-1145.
2. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F: **A space-time permutation scan statistic for disease outbreak detection.** *Plos Medicine* 2005, **2:**216-224.

National Syndromic Surveillance Program

Community of Practice

Resources that Advance the Science and Practice of Syndromic Surveillance

2022

Syndromic Surveillance Symposium

December 6–8, 2022

KAHUINA

CONSULTING