A Tale of Two Views: Organization of California's COVID-19 Vaccination Data

TIMOTHY LO

JASMINE DROZE

Topic

California COVID-19 vaccination data structure from an epidemiologist perspective

Challenges, results, and uses of COVID-19 vaccination data

- Reasons for resulting set-up, uses, pros and cons

Historical Context

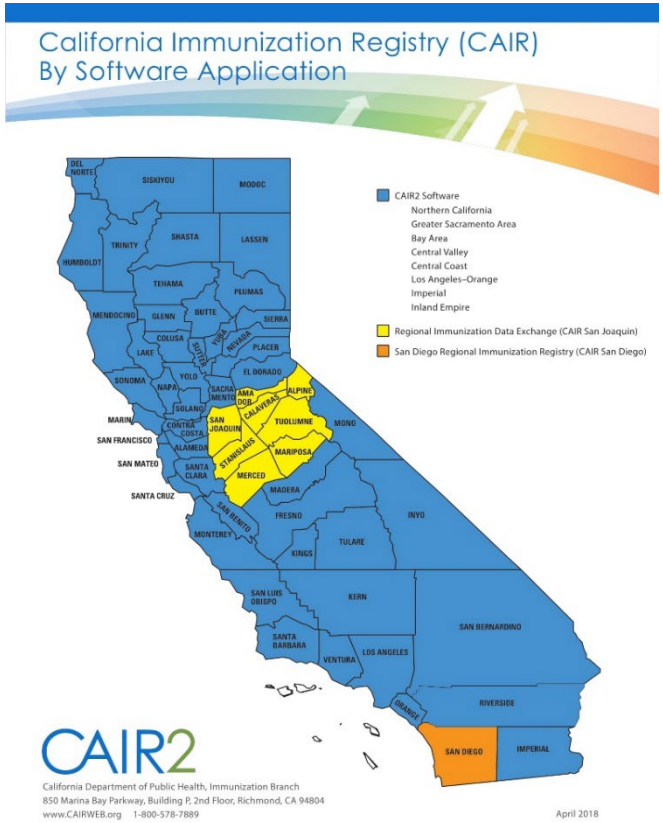
California had 3 separate, independent immunization registries

- CAIR, SDIR, RIDE

No state mandate for reporting of vaccinations (except pharmacies)

- Limitations for any statewide vaccination data reports or analyses

And then…



Some Challenges…

Statewide reports were now expected and required

Multiple asks, requests, workgroups, and needs

“Can we get the number of persons fully vaccinated by county?”

- What counts as a person?
- What counts as fully vaccinated?
- What counts as a county?

“Why don’t our numbers match?”

Some Challenges… (cont.)

Real decisions were being made off what came out

- Provider performance, vaccine allocation, vaccine equity
- Urgency of requests
  - Delays in reporting
  - Delays in processing

Number of dose in series

- First dose? Second dose? According to who/which source?
- How to calculate a person who's fully vaccinated?

"Other data issues"

The Set-up Today

Data is housed on an online cloud-based data repository, Snowflake

CDC required data elements were the starting point

- Required a standardized set of core data elements

Two views: dose-level and recipient level

- View = stored SQL query; only exists when someone queries it. Essentially a table
- Dose level: one row per vaccination
- Recipient level: one row per recipient

https://cairweb.org/snowflake/

View Visual Aid

DOSE LEVEL VIEW

VAX_EVENT_ID	RECIP_ID	ADMIN_DATE	VAX_LABEL
ID_1001	ID_JOHN	3/1/2020	PFIZER
ID_1002	ID_JOHN	3/29/2020	PFIZER
ID_1003	ID_JANE	3/4/2020	MODERNA
ID_1004	ID_JANE	4/2/2020	MODERNA
ID_1005	ID_JEFF	3/15/2020	PFIZER

RECIPIENT LEVEL VIEW

RECIP_ID	DS1_ADMIN_DATE	DS2_ADMIN_DATE	FULLY_VACCINATED
ID_JOHN	3/1/2020	3/29/2020	TRUE
ID_JANE	3/4/2020	4/2/2020	TRUE
ID_JEFF	3/15/2020		FALSE

*This is fake data for example purposes only

View Usage Example

Example: “Can we get the number of persons fully vaccinated by county?”

- Counting “persons”? Recipient-level view!
- “Fully vaccinated”? Make it a column!
  - And explain how the values are assigned
- “County”? Pick either recipient county or admin county!
  - And explain why we’re using that column
- Report the results somewhere (dashboard, report) and explain which columns we used and why

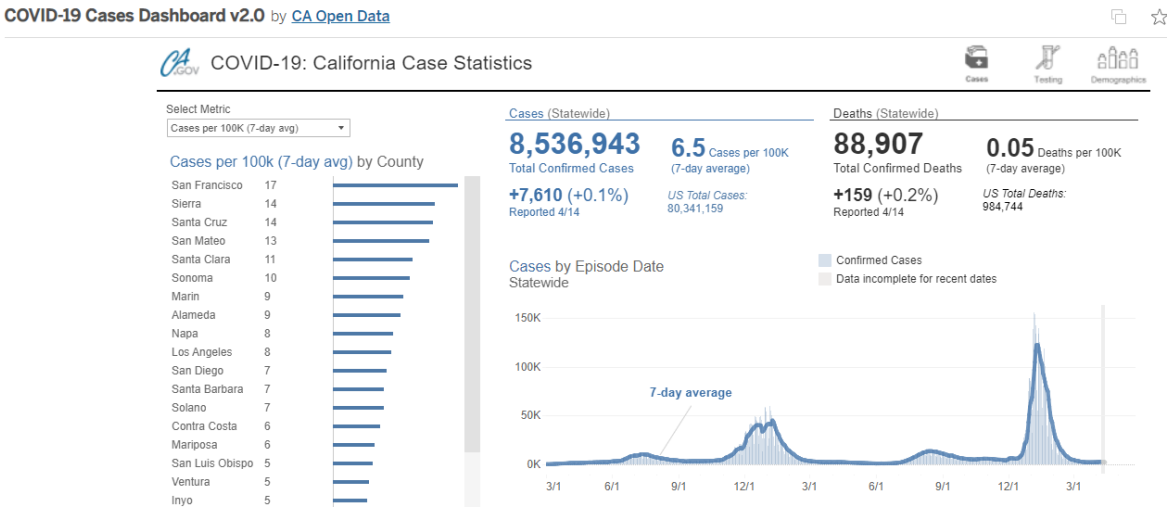
Advantages

Allows for consistency across different reports, counts, and analyses

- Counting doses? Use dose-level view
- Counting persons? Use recipient-level view

Allows sharing of data and reports

- Open Data Portal, public dashboards



Advantages (cont.)

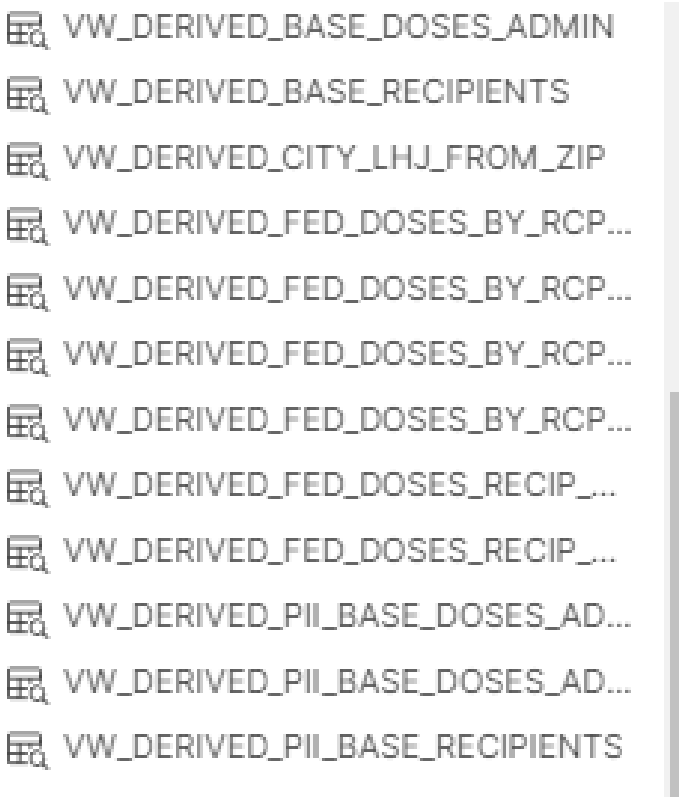
Ease of adding new columns, fields, and methodologies
- "Fully_vaccinated" column

Allows for documentation and data dictionaries

Allows for subsequent views to be derived from base views

Snowflake and other data resources

- Snowflake COVID-19 Vaccination Data Dictionary (updated 4/8/2022)
- Snowflake COVID-19 Vaccination Data Dictionary – Recipient Level View (updated 4/8/2022)
- Snowflake Job Aid – How to Access (updated 10/19/2021)
- Snowflake Job Aid – Frequently Used SQL Queries (updated 2/22/2022)
- Snowflake COVID-19 Dose De-duplication Logic
- CAIR2 Site List (as of 7/12/2021)



Limitations

Standardizing

- Over 70mil records, decisions have to be made (for better or worse) about how to count/assign values
- May not capture everyone's needs
  - E.g.: smaller jurisdictions do more post processing and clean-up

Very clear and consistent communication required

- Processing time, overnight updates
- Two views = "two raw data sources"
- Trainings, emails, data dictionary updates

Conclusion

Two base level views work well for our needs

Allows for consistency across multiple reports

Clear communication required

https://covid19.ca.gov/state-dashboard/

Questions?

Contact

Timothy Lo

Timothy.lo@cdph.ca.gov

Jasmine Droze

Jasmine.Droze@cdph.ca.gov