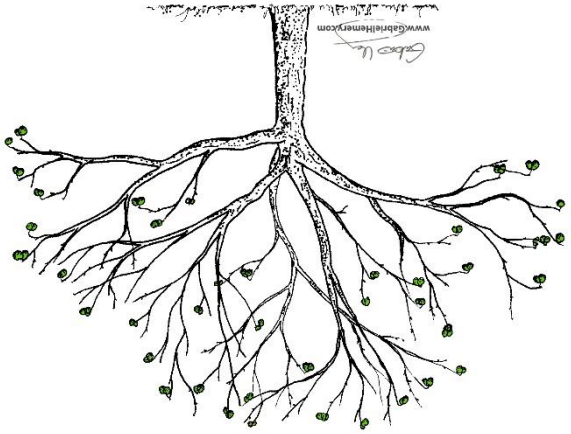


Data-Mining for Adverse Events Using the Self-Controlled Tree-Temporal Scan Statistic

W. Katherine Yih, PhD, MPH
Department of Population Medicine
Harvard Medical School, and
Harvard Pilgrim Health Care Institute

Disclosures

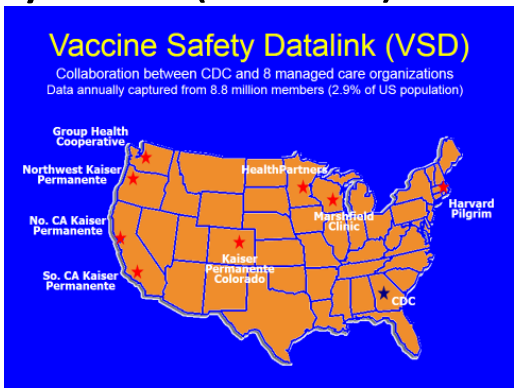
- Funding sources:
  - U.S. Food and Drug Administration
    - HHS Sentinel Contract number HHSF223201400030I
  - Harvard Pilgrim Health Care Institute
    - Robert H. Ebert Career Development Awards
- I receive some research funding from GlaxoSmithKline

- Many thanks to Martin Kulldorff for permission to use several of his slides

Background

Post-licensure vaccine safety monitoring systems in U.S.

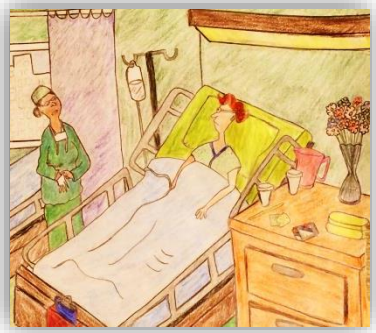
- Spontaneous reporting system:
  - Vaccine Adverse Event Reporting System (**VAERS**)
- Population-based systems:
  - CDC-sponsored Vaccine Safety Datalink (**VSD**) →
  - FDA-sponsored **Sentinel** system
    - 67 million people currently accruing new data



Bob's Story

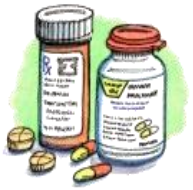
Demographic					
Database	**Member ID**	**Birth Date**	**Sex**	**Race/Ethnicity**	**ZIP Code**
Bob's insurance company	5291321	07/29/63	Male	Unknown	02119

Lives in Boston, MA



Has appendectomy

Diagnosed with hypertension



Visit at another delivery system

Encounter
- 1/1/11
- **Office Visit**
- **Dx:** Diabetes

Dispensings
- 1/1/11
- **Rx:** Anti-hyperglycemic drug

Encounters
- 3/15/12
- **Emergency Department**
- **Px:** appendectomy
- 3/15/2012-3/18/2012
- **Hospital stay**

Encounter
- 12/11/12
- **Office Visit**
- **Dx:** Hypertension

Dispensings
- 12/11/12
- **Rx:** Anti-hypertensive drug

Encounter
- 10/31/13
- **Office Visit**
- **Dx:** Depression

2011 2012 2013 2014

Stages of vaccine safety assessment

(hypothesis-generating) (hypothesis-testing)

VAERS can identify previously unsuspected AEs but has limitations

Past work of CDC's **VSD** and FDA's **Sentinel** systems has been concentrated here, addressing one or more suspected AEs

Stages of vaccine safety assessment

VAERS can identify previously unsuspected AEs but has limitations

Past work of CDC's **VSD** and FDA's **Sentinel** systems has been concentrated here, addressing one or more suspected AEs

"TreeScan" data-mining method is used with population-based data like **VSD**'s and **Sentinel**'s

Two signal detection systems, #1

System	VAERS	Sentinel and VSD
Data	Spontaneous reports	EHR or claims
Analysis method	Disproportionality analysis	Self-controlled tree-temporal scan statisticis
Year available	VAERS started in 1990, disproportionality analysis came later	Method used by FDA's Sentinel and CDC's VSD since ~2014 but not yet in routine use by public health agencies nor available to public
Geographic scope	National Also, VAERS transmits its vaccine adverse event (AE) reports to Uppsala Monitoring Center, contributing to global pharmacovigilance	Depends on dataset used—typically, geographically diverse subset of national population (Sentinel, VSD, Truven/Marketscan)
Speed	Relatively fast due to direct reporting capability and the speed at which reports and follow-up information can be processed and analyzed Less impacted by data lags and delayed access to health records than claims-based monitoring systems	Depends on source data system lags Claims-based surveillance datasets, e.g., Sentinel, Truven/Marketscan, have data lag
Track record	Has successfully detected safety signals, e.g., Rotashield and intussusception	Has detected known AEs, not yet any unexpected AEs, but not yet in routine use
Numerator	Subject to reporting bias, including underreporting of AEs (especially common, mild ones) and stimulated reporting (e.g., in response to intense media attention)	Events must be medically attended to be captured Less subject to reporting bias than spontaneous reporting systems But may be subject to "upcoding" or other coding idiosyncrasies in source data
Denominator	Vaccine doses distributed provides proxy measure of persons vaccinated	Vaccines administered
Attributable risk calculable?	No	Yes

Two signal detection systems, #1

System	VAERS	Sentinel and VSD
Data	Spontaneous reports	EHR or claims
Analysis method	Disproportionality analysis	Self-controlled tree-temporal scan statisticis
Year available	VAERS started in 1990, disproportionality analysis came later	Method used by FDA's Sentinel and CDC's VSD since ~2014 but not yet in routine use by public health agencies nor available to public
Geographic scope	National Also, VAERS transmits its vaccine adverse event (AE) reports to Uppsala Monitoring Center, contributing to global pharmacovigilance	Depends on dataset used—typically, geographically diverse subset of national population (Sentinel, VSD, Truven/Marketscan)
Speed	Relatively fast due to direct reporting capability and the speed at which reports and follow-up information can be processed and analyzed Less impacted by data lags and delayed access to health records than claims-based monitoring systems	Depends on source data system lags Claims-based surveillance datasets, e.g., Sentinel, Truven/Marketscan, have data lag
Track record	Has successfully detected safety signals, e.g., Rotashield and intussusception	Has detected known AEs, not yet any unexpected AEs, but not yet in routine use
Numerator	Subject to reporting bias, including underreporting of AEs (especially common, mild ones) and stimulated reporting (e.g., in response to intense media attention)	Events must be medically attended to be captured Less subject to reporting bias than spontaneous reporting systems But may be subject to "upcoding" or other coding idiosyncrasies in source data
Denominator	Vaccine doses distributed provides proxy measure of persons vaccinated	Vaccines administered
Attributable risk calculable?	No	Yes

Methods

TreeScan overview

For selected vaccine, use population-based longitudinal data to:

- Evaluate thousands of potential AEs via electronic diagnosis codes
- Evaluate multiple potential risk windows
- Adjust for the multiple testing

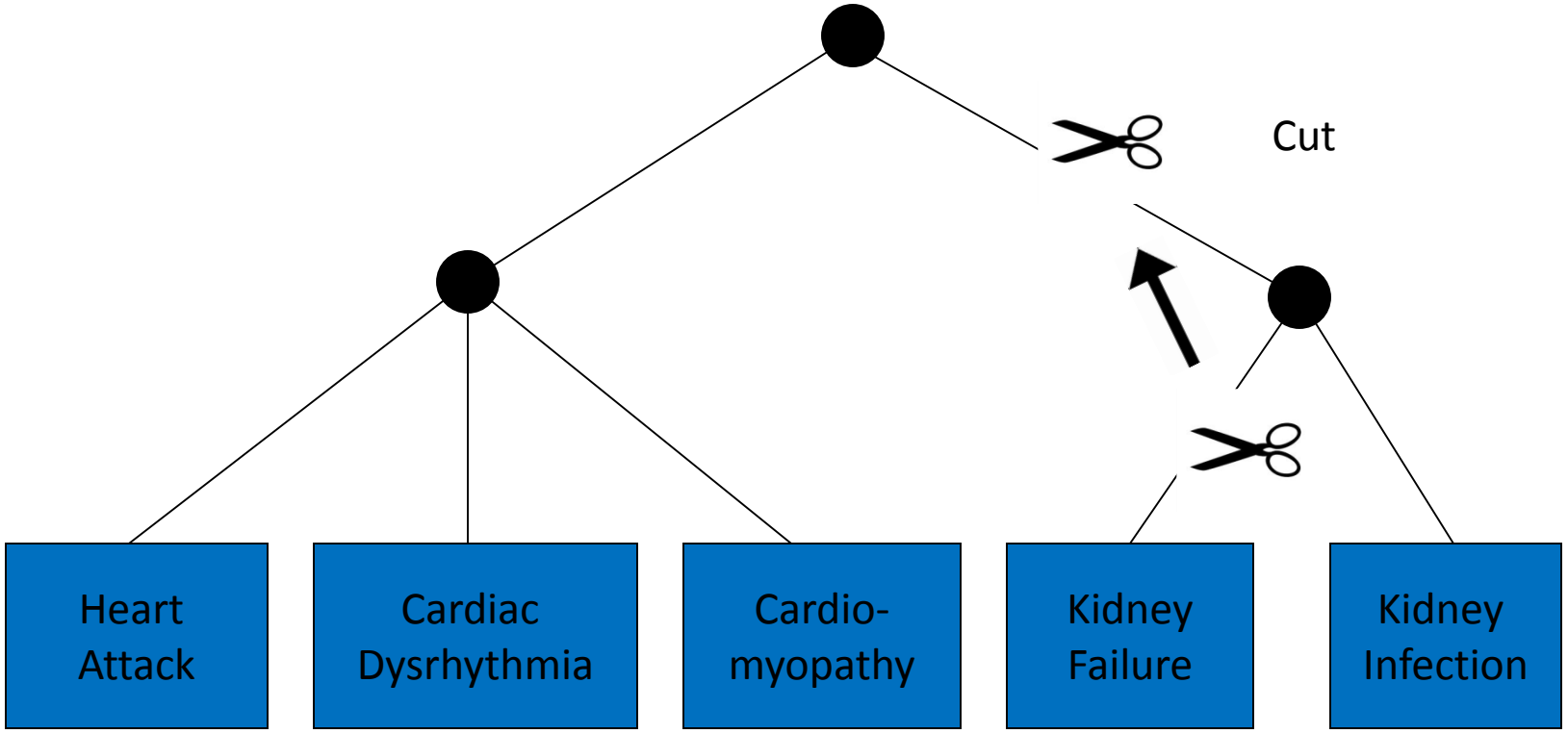
Goals:

- Find known and any previously unsuspected AEs in specified follow-up period after vaccination
- Minimize false positives
- Have enough sample size to detect very rare AEs

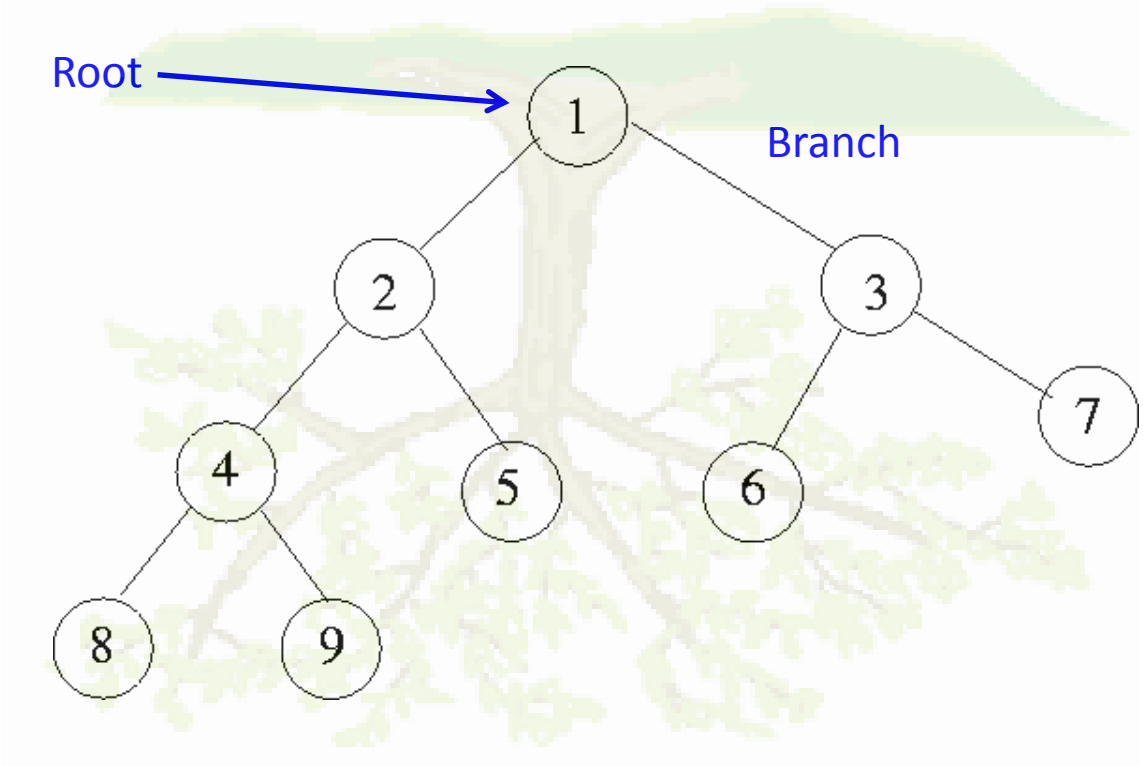
Relatedness of potential AEs

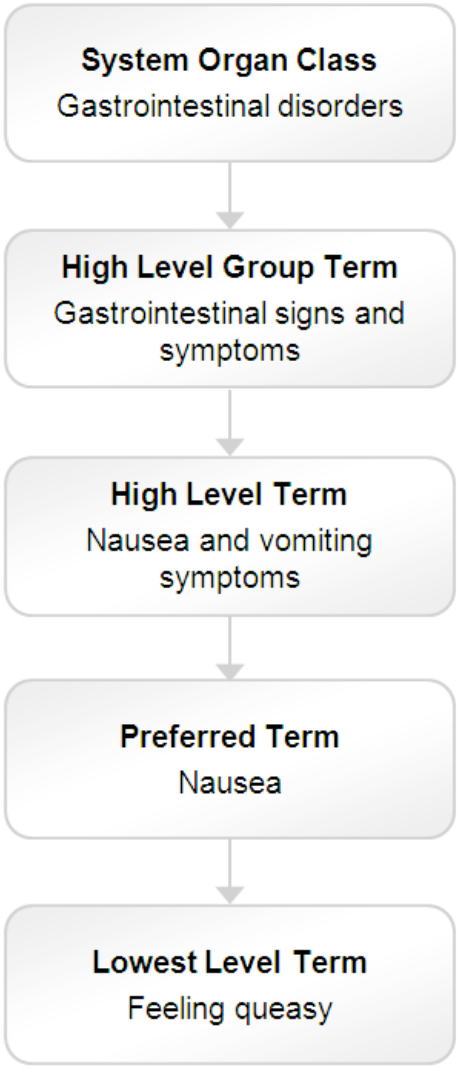
- The vaccine might cause a spectrum of disease rather than a highly specific condition corresponding to just one ICD code
- Clinicians might differ somewhat in how they code for a given condition
- So it's desirable to evaluate groups of related conditions in addition highly specific conditions

A small three-level tree



Examples of diagnosis trees





- MedDRA reporting terms
- Multi-Level Clinical Classification System

Multi-Level Clinical Classification System

- MLCCS—product of Agency for Healthcare Research and Quality (AHRQ)'s Healthcare Cost and Utilization Project (HCUP)
  - http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp
- Examples of first, broadest level of diagnosis
  - Diseases of the nervous system and sense organs
  - Diseases of the circulatory system
  - Diseases of the respiratory system
  - Diseases of the digestive system
  - Diseases of the genitourinary system
  - Diseases of the skin and subcutaneous tissue
  - Injury and poisoning

Lowest level: ~6,000 ICD-9-CM codes

Example of MLCCS hierarchical classification scheme

Level of tree and code	Description
1 07	Diseases of the circulatory system
2 07.05	Diseases of veins and lymphatics
3 07.05.01	Phlebitis; thrombophlebitis and thromboembolism
4 07.05.01.01	Phlebitis and thrombophlebitis
5 451.0	Of superficial vessels of lower extremities
5 451.11	Femoral vein phlebitis
5 451.19	Deep phlebitis leg not elsewhere classified
5 451.2	Thrombophlebitis leg not otherwise specified
5 451.81	Ilias thrombophlebitis
5 451.82	Superficial phlebitis arm
5 451.83	Deep phlebitis arm
5 451.84	Thrombophlebitis arm not otherwise specified
5 451.89	Thrombophlebitis not elsewhere classified
5 451.9	Thrombophlebitis not otherwise specified

Data & parameters for two vaccine studies

	HPV4 (Gardasil)	ZVL (Zostavax)
Data source	5 Sentinel Data Partners	Truven Health MarketScan Research Databases
Age range	9-26	≥ 60
Settings	Inpatient or ED	Inpatient or ED
Incidence criterion	First in 183 days	First in 400 days
Follow-up period	Days 1-56	Days 1-56
Risk intervals evaluated	Intervals 2-28 days long starting in Days 1-28 and ending in Days 2-42	Intervals 2-28 days long starting in Days 1-28 and ending in Days 2-42

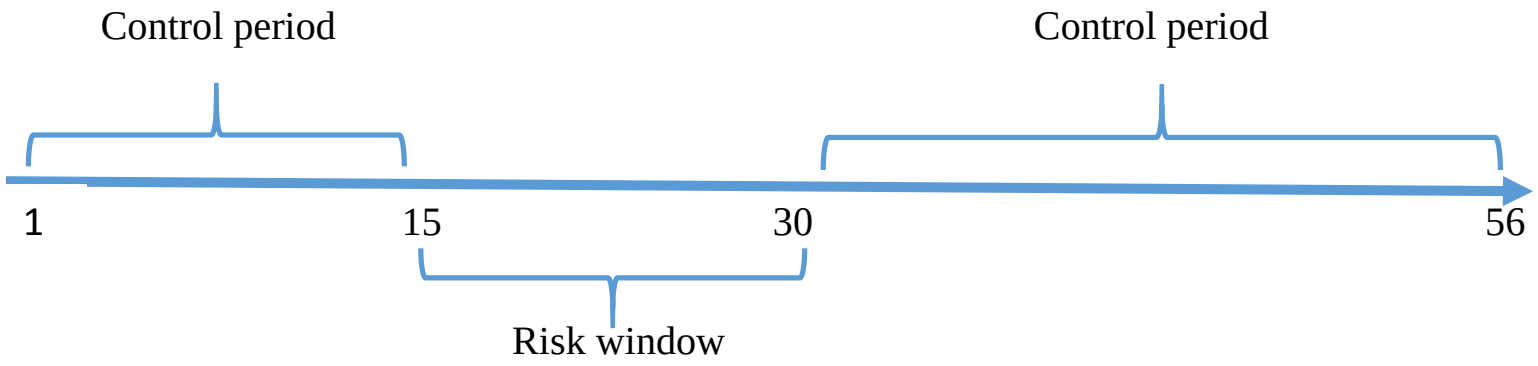
Potential risk windows scanned

Follow-up period: 1-56 days

Risk window start range: 1-28 days after vaccination

Risk window end range: 2-42 days after vaccination

Risk window length: 2-28 days



Results

Results for HPV4 (Gardasil), 1.9 M 1st doses

Ref #	Node code	Node text	RW	Cases in RW	AR	P
1	12	Diseases of the skin and subcutaneous tissue	2-4	214	3.8	0.0019
2	12.01	. Skin and subcutaneous tissue infections	2-4	111	2.3	0.042
3	12.01.01	. . Cellulitis and abscess	2-4	93	2.0	0.20
4	12.01.01.03	. . . Cellulitis and abscess of arm (only 682.3)	2-3	31	1.3	0.00001
5	**682.3**	**. . . . Cellulitis and abscess of upper arm and forearm**	**2-3**	**31**	**1.3**	**0.00001**
6	12.02	. Other inflammatory condition of skin				
7	695.9	. . . . Unspecified erythematous condition	2-3	13	0.5	0.25

Results for HPV4 (Gardasil), 1.9 M 1st doses

Ref #	Node code	Node text	RW	Cases in RW	AR	P
1	12	Diseases of the skin and subcutaneous tissue	2-4	214	3.8	0.0019
2	12.01	. Skin and subcutaneous tissue infections	2-4	111	2.3	0.042
3	12.01.01	. . Cellulitis and abscess	2-4	93	2.0	0.20
4	12.01.01.03	. . . Cellulitis and abscess of arm (only 682.3)	2-3	31	1.3	0.00001
5	**682.3**	**. . . . Cellulitis and abscess of upper arm and forearm**	**2-3**	**31**	**1.3**	**0.00001**
6	12.02	. Other inflammatory condition of skin				
7	695.9	. . . . Unspecified erythematous condition	2-3	13	0.5	0.25
8	16	Injury and poisoning	1-3	48	2.2	0.00001
9	16.10	. Complications	1-3	36	1.8	0.00001
10	16.10.02	. . Complications of surgical procedures or medical care	1-3	36	1.8	0.00001
11	**16.10.02.07**	**. . . Other complications of surgical and medical procedures**	**1-3**	**36**	**1.8**	**0.00001**
12	780.63	. . . . Post-vaccination fever	1-2	4	0.2	0.31
13	999.5	. . . . Other serum reaction not elsewhere classified	1-3	7	0.4	0.011
14	999.52	. . . . Other serum reaction due to vaccination	1-2	11	**0.6**	0.00001
15	999.9	. . . . Other and unspecified complications of medical care	1-6	12	0.6	0.0018

Follow-up of HPV4 “other complications”

- Generated claims reports for period 8 weeks before through 12 weeks after vaccination for patients contributing to signal
- Clinical review

Cases in HPV4 "other complications…"

Conditions	No.
Conditions identified in package insert as possible vaccine-associated adverse events*	29
No specified symptoms and no further medical visits within 60 days	3
Diverse symptoms, different in each case	4
Total	36

* e.g., headache, fever, nausea, and dizziness; local injection site reactions

31 (86%) of the 36 cases received ≥ 1 other vaccine along with HPV4

Results for ZVL (Zostavax), 1.2 M doses

Ref #	Node code	Node text	Risk window (days after vaccination)	Number of cases observed in risk window	Number of excess cases per 100,000 vaccinees	P
1	12.01.01	Cellulitis and abscess	1-4	113	5.2	0.001
2	12.01.01.03	. . Cellulitis and abscess of arm	1-3	61	4.6	0.001
3	682.3	. . . . Cellulitis and abscess of upper arm and forearm	1-3	61	4.6	0.001
4	12.02.00	Other inflammatory condition of skin	1-4	30	1.4	0.548
5	695.9	. . . . Erythematous condition NOS	2-4	16	1.1	0.001
6	16.10.02	Complications of surgical procedures or medical care	1-3	39	3.1	0.001
7	16.10.02.07	. . Other complications of surgical and medical procedures	1-3	39	3.1	0.001
8	999.52	. . . . Other serum reaction due to vaccination	1-3	20	1.6	0.001
9	999.0	. . . . Generalized vaccinia	1-3	7	0.6	0.001
10	999.9	. . . . Other and unspec complications of medical care	1-3	8	0.6	0.060
11	17.01.09	Allergic reactions	1-3	44	2.1	0.004
12	17.01.09.00	. . Allergic reactions	1-3	44	2.1	0.004
13	995.3	. . . . Allergy NOS	1-6	40	2.1	0.002

Conclusions

Conclusions, TreeScan data-mining

Thousands of potential adverse reactions and hundreds of potential risk windows evaluated, while adjusting for multiple testing

- Known adverse reactions found
- No false alerts
- High power to detect rare adverse reactions

Caveats

- When follow-up periods only a few weeks long, outcomes with long latency periods can be missed
- When follow-up periods longer, time-varying confounding can happen
- With a diagnosis tree organized by system, outcomes manifesting in diverse systems (neuro, GI, cardiovascular) might be missed

Stages of vaccine safety assessment

VAERS can identify previously unsuspected AEs but has limitations

Past work of CDC's **VSD** and FDA's **Sentinel** systems has been concentrated here, addressing one or more suspected AEs

"TreeScan" data-mining method is used with population-based data like **VSD**'s and **Sentinel**'s

Two signal detection systems, #2

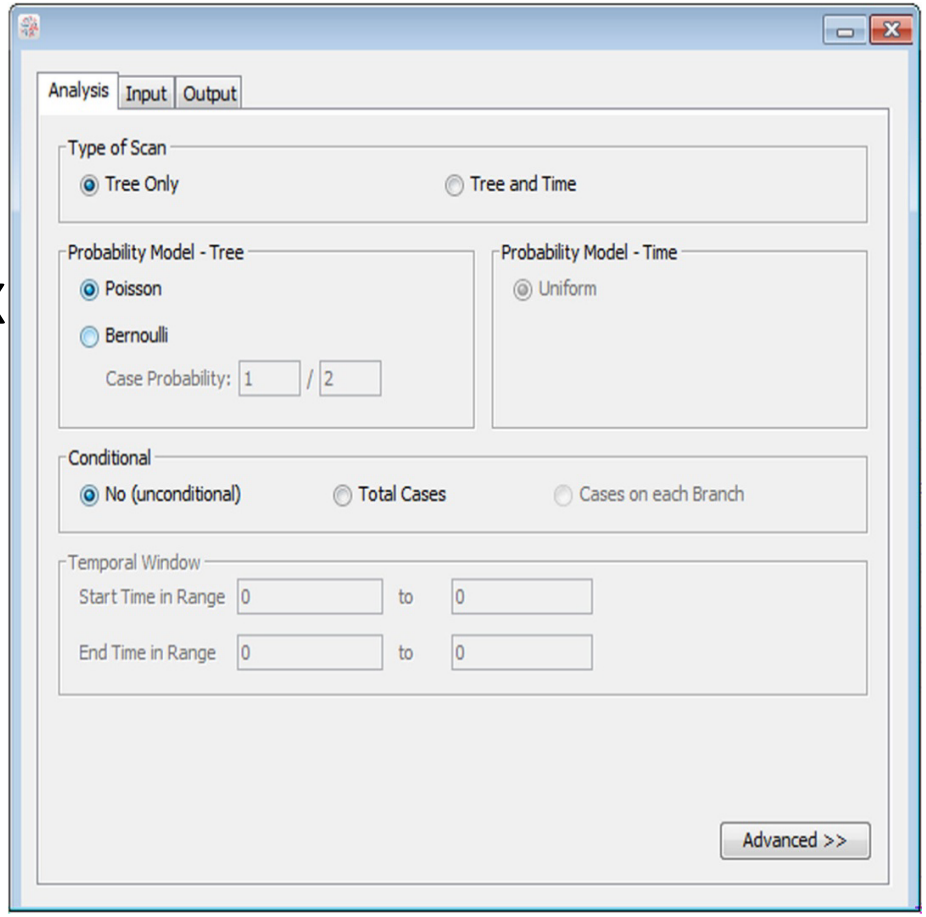
System	VAERS	Sentinel and VSD
Main limitations	• Subject to general and temporal reporting bias • No accurate denominators • Comparison group for analysis may not be similar	• Subject to time-varying confounding, especially if follow-up period several months long or more • Results may depend on specific tree structure employed
Main strengths	• Can detect unusual numbers of any reported AE • Fast	• Minimal reporting bias • Population based, allows AR to be calculated • Self-controlled • Detects unusual clustering of any medically attended specific AE or more general AE category • Detects temporal clustering of AEs • Formally controls for multiple testing

TreeScan updates

- ICD-10 code tree
- Enhancement to allow longer follow-up, with censoring
  - May detect AEs with longer latency
  - Can include subjects who disenroll during follow-up period
  - But may not work well for evaluating vaccine safety in older populations—censoring not wholly independent of the outcome
- Sequential version to assess vaccine safety repeatedly as data accumulate, adjusting for multiple testing (under development)

TreeScan software

- Free
- www.treescan.org
- Windows, Mac, Linux
- User Guide (47p)



Some references

- Kulldorff M, Fang Z, Walsh SJ. A tree-based scan statistic for database disease surveillance. *Biometrics*. 2003;59:323-31.
- Kulldorff M, Dashevsky I, Avery TR, et al. Drug safety data mining with a tree-based scan statistic. *Pharmacoepidemiol Drug Saf*. 2013;22:517-23.
- Shimabukuro TT, Nguyen M, Martin D, DeStefano F. Safety monitoring in the Vaccine Adverse Event Reporting System (VAERS). *Vaccine*. 2015;33:4398-405.
- Yih WK, Maro JC, Nguyen M, et al. Pilot of self-controlled tree-temporal scan analysis for Gardasil vaccine. *Am J Epidemiol.* 2018;187:1269-76.
- Li R, Weintraub E, McNeil MM, et al. Meningococcal conjugate vaccine safety surveillance in the Vaccine Safety Datalink using a tree-temporal scan data mining method. *Pharmacoepidemiol Drug Saf.* 2018;27:391–7.
- Yih WK, Kulldorff M, Dashevsky I, Maro JC. Using the self-controlled tree-temporal scan statistic to assess the safety of live attenuated herpes zoster vaccine. *Am J Epidemiol.* 2019;188:1383–8.

Extras

To bear in mind in using TreeScan

- We used the *Conditional* Self-Controlled Tree-Temporal Scan Statistic
  - Controls for phenomenon of follow-up visits after preventive care visits
- Age group matters
  - Based on pilot, method likely useful for other adolescent/young adult vaccines
  - Not suitable yet for infant vaccines—time-varying confounding

To bear in mind in using TreeScan *(cont'd)*

- Can prune tree to remove outcomes unlikely to be caused by vaccination, e.g.
  - Outcomes unlikely to be caused by vaccination, e.g., well-care visits, delivery of baby, vitamin deficiencies, or fractures
  - Conditions unlikely to appear within a few weeks, e.g., cancer
  - Most infectious diseases with identified organism, e.g., typhoid fever, tuberculosis, shigella
  - Congenital conditions, e.g., sickle cell disease, congenital heart disease
- Can map ICD-10 codes to ICD-9 codes in order to still use MLCCS tree; we used CMS General Equivalence Mappings (GEMs)
- https://www.sentinelinitiative.org/sentinel/surveillance-tools/software-toolkits/treeextraction-documentation