Sequential Statistical Analysis for Post-Market Vaccine and Drug Safety Surveillance

Martin Kulldorff
Department of Population Medicine
Harvard Medical School and
Harvard Pilgrim Health Care
The Need for Post-Marketing Safety Surveillance

- Due to insufficient sample size, rare but serious adverse events may not be detected in pre-licensure phase 3 clinical trials.
- Adverse events may be specific to population not included in the trial.
- Risk of adverse events may differ between drugs/vaccines with equal efficacy.
Post-Marketing Safety Data

- Phase IV randomized clinical trials
- Case-control studies
- Spontaneous adverse event reporting systems
- Routinely collected electronic health data
DO YOUR PART for Vaccine Safety — Report to VAERS.

If one of your patients becomes ill after being vaccinated, promptly report it to the Vaccine Adverse Event Reporting System (VAERS), even if you are not sure that the vaccine caused the illness.

As a healthcare provider, you can help to ensure the safety of vaccines given to patients in the United States by reporting adverse events to VAERS.

- You may report to VAERS online or download and print a VAERS form at www.vaers.hhs.gov.
- You may mail your VAERS form to:
  VAERS
  P.O. Box 1100
  Rockville, MD 20849-1100
  or fax to: (877) 721-0366
- For additional information on VAERS, call (800) 822-7967.

VAERS
Vaccine Adverse Event Reporting System

VAERS is a national vaccine safety surveillance program, co-sponsored by the Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA). VAERS is a post-licensure surveillance program that collects information about adverse events (possible side effects) that occur after the administration of vaccines licensed for use in the United States.
Drug / Vaccine Safety Surveillance

Two Options:

• Wait X months or years and evaluate if there are any adverse events, using standard statistical methods. How do we pick X?

• Rapid Cycle Analysis / Sequential Analysis

If there is a problem we want to detect it as soon as possible
Sequential Statistical Analysis

- Used when multiple analyses are done repeatedly over time as the data set grows
- Adjusts for the multiple testing due to the multiple looks at the data
- Useful for the early detection of problems
- Commonly applied in clinical trials
- Used by the Vaccine Safety Datalink since 2004 for weekly monitoring of vaccine safety
Part 1: A Maximized Sequential Probability Ratio Test

Collaborators:
Robert Davis, Kasier Permanente Georgia
Margarette Kolczak, Centers for Disease Control and Prevention (CDC)
Edwin Lewis, Kaiser Northern California
Tracy Lieu, Richard Platt, Harvard Medical School and Harvard Pilgrim Health Care
Data Structure

Weekly data and analyses, during weeks \( t=1,\ldots,T \), typically with \( T>100 \).

\( c_t = \# \) of observed adverse events in week \( t \).

\( n_t = \# \) of expected adverse events in week \( t \), typically with \( n_t <1 \).

Continuous sequential methods more appropriate than group sequential
Sequential Probability Ratio Test (SPRT, Wald 1945)

Log likelihood ratio test statistic at time $t$:

$$LLR(t) = \ln \left( \prod_{i=1}^{t} \frac{P(c_i \mid H_A)}{P(c_i \mid H_0)} \right)$$

where $c_i$ = the observed number of events at time $i$.

Reject $H_0$: $RR=1$, when $LLR(t) > B$
Reject $H_A$: $RR=r$, when $LLR(t) < A$

where $A = \ln(\beta/(1-\alpha)) = -1.56$ and $B = \ln((1-\beta)/\alpha) = 2.77$, when $\alpha=0.05$ and $\beta=0.2$.
Pediarix Vaccine Surveillance
Historical Data

- Adverse Events: Fever
- Adverse Event Window: 28 days
- Kaiser Permanent Northern California Data
- Time Period: Mar 30, 2003 to Jan 24, 2004
- Weekly Analyses
- Expected counts from historical DTaP data
Pediarix Vaccine Surveillance
Fever

Analysis Parameters:
- $\alpha = 0.05$
- $\beta = 0.20$
- $H_0: \text{RR}=1$
- $H_A: \text{RR}=2$ (two-fold excess risk)
Fever (0 – 28 DAYS Post PDXat NCK)

Risk Adjusted SPRT - Age Gender

Log LR

Vaccine Week Number
Pediarix Vaccine Surveillance

Fever

Analysis Parameters:

• \( \text{alpha} = 0.05 \)
• \( \text{beta} = 0.20 \)
• \( \text{H0: RR}=1 \)
• \( \text{HA: RR}=1.2 \) (20% excess risk)
Fever (0 – 28 DAYS Post PDXat NCK)

Risk Adjusted SPRT - Age Gender
Detect 20% increase from 1998 - 2003 to 2003/2004

Log LR

Vaccine Week Number
Fever (0 – 28 DAYS Post PDXat NCK)

Risk Adjusted SPRT - Age Gender

Log LR

Vaccine Week Number
What Is Going On?

Suppose that the true RR=1.2

- If the null is RR=1 and the alternative is RR=2, then there is more evidence for the null hypothesis than for the alternative hypothesis.

- If the null is RR=1 and the alternative is RR=1.2, then there is more evidence for the alternative hypothesis than for the null hypothesis.
Pediarix Vaccine Surveillance
Neurological Symptoms

Analysis Parameters:
• alpha = 0.05
• beta = 0.20
• H0: RR=1
• HA: RR=1.2 (20% excess risk)
Neurologic Symptoms (0–28 DAYS Post PDX at NCK)

Risk Adjusted SPRT - Age Gender
Detect 20% increase from DTaP to Pediarix

Log LR

Vaccine Week Number
Pediarix Vaccine Surveillance
Neurological Symptoms

Analysis Parameters:

- alpha = 0.05
- beta = 0.20
- H0: RR=1
- HA: RR=1.5 (50% excess risk)
Neurologic Symptoms (0–28 DAYS Post PDXat NCK)

Risk Adjusted SPRT - Age Gender
Detect 50% increase from DTaP to Pediarix
Neurologic Symptoms (0–28 DAYS Post PDXat NCK)

Risk Adjusted SPRT - Age Gender
Detect 20% increase from DTaP to Pediarix

Log LR

Vaccine Week Number
What Is Going On?

Suppose the true RR=2.

- If the alternative model is RR=1.2, then it is almost as bad as the null model with RR=1, so the log likelihoods are similar and the log likelihood ratio is close to one.
- If the alternative model is RR=1.5, then there is much more evidence for the alternative than for the null hypothesis.
Composite Alternative:

- HO: RR=1
- HA: RR>1
- \( \alpha = 0.05 \)
- \( \beta = 0.20 \)
Maximized SPRT

Log likelihood ratio test statistic at time $t$:

$$LLR(t) = \max_{RR>1} \ln \left( \prod_{i=1}^{t} \frac{P(c_i \mid H_A(RR))}{P(c_i \mid H_0)} \right)$$

where $c_i = \text{the observed number of events at time } i$.

Reject $H_0$ when $LLR(t) > B$

Reject $H_A$ when $t > T$

Specify $\alpha$ and $T$ in advance, calculate $B$. 
Critical Bounds

LLR

Reject H0

0

Accept H0

T

time
Length of Surveillance

The maximum length of surveillance, $T$, must be defined in terms of expected events under the null rather than calendar time.
Maximized SPRT

• Because of the non-constant rejection boundaries, it is a Generalized Sequential Probability Ratio Test (Weiss, Ann Math Stat, 1953, p273)

• Because of the composite alternative, it is a Sequential Generalized Likelihood Ratio Test (Siegmund and Gregory, Ann Stat, 1980 p1223)
Expected Counts

Expected counts may be based on for example

- Historical Data
- Estimates from the Literature
- Concurrent Matched Controls

> KEY ISSUE
Exact Critical Values

Basic Idea:

• While surveillance is done continuously, events happened at discrete time points

• For a specific critical value, we can calculate the expected needed to reject H0 with for each discrete value of the number of events

• Through iteration, we can get the critical value for a desired alpha
MaxSPRT: Critical Bounds

<table>
<thead>
<tr>
<th>T</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.001$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.84</td>
<td>4.57</td>
<td>6.88</td>
</tr>
<tr>
<td>3</td>
<td>3.10</td>
<td>4.75</td>
<td>6.96</td>
</tr>
<tr>
<td>5</td>
<td>3.24</td>
<td>4.93</td>
<td>7.06</td>
</tr>
<tr>
<td>10</td>
<td>3.45</td>
<td>5.12</td>
<td>7.51</td>
</tr>
<tr>
<td>50</td>
<td>3.85</td>
<td>5.39</td>
<td>7.74</td>
</tr>
<tr>
<td>100</td>
<td>3.98</td>
<td>5.49</td>
<td>7.83</td>
</tr>
<tr>
<td>1000</td>
<td>4.29</td>
<td>5.81</td>
<td>8.15</td>
</tr>
</tbody>
</table>
Pediarix Surveillance: Fever
Number of Weeks until Rejection

<table>
<thead>
<tr>
<th></th>
<th>$\alpha=0.05$</th>
<th>$\alpha=0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T~2 years</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>T~1 year</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>T~3 months</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>Classical SPRT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR=1.05</td>
<td>36</td>
<td>73</td>
</tr>
<tr>
<td>=1.1</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>=1.2</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>=1.5</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>=2</td>
<td>never</td>
<td>never</td>
</tr>
<tr>
<td>=5</td>
<td>never</td>
<td>never</td>
</tr>
</tbody>
</table>
Pediarix Surveillance: Neurological Number of Weeks until Rejection

<table>
<thead>
<tr>
<th>MaxSPRT</th>
<th>$\alpha=0.05$</th>
<th>$\alpha=0.01$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T~2 years</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>T~1 year</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>T~3 months</td>
<td>32</td>
<td>42</td>
</tr>
<tr>
<td>Classical SPRT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR=1.05</td>
<td>&gt;82</td>
<td>&gt;82</td>
</tr>
<tr>
<td>=1.1</td>
<td>&gt;82</td>
<td>&gt;82</td>
</tr>
<tr>
<td>=1.2</td>
<td>65</td>
<td>&gt;82</td>
</tr>
<tr>
<td>=1.5</td>
<td>42</td>
<td>52</td>
</tr>
<tr>
<td>=2</td>
<td>32</td>
<td>42</td>
</tr>
<tr>
<td>=5</td>
<td>never</td>
<td>never</td>
</tr>
</tbody>
</table>
MaxSPRT, Power ($\alpha=0.05$)

<table>
<thead>
<tr>
<th>T</th>
<th>RR = 1.2</th>
<th>1.5</th>
<th>2.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.07</td>
<td>0.11</td>
<td>0.18</td>
<td>0.73</td>
</tr>
<tr>
<td>5</td>
<td>0.09</td>
<td>0.19</td>
<td>0.44</td>
<td>0.999</td>
</tr>
<tr>
<td>10</td>
<td>0.11</td>
<td>0.28</td>
<td>0.68</td>
<td>1.00</td>
</tr>
<tr>
<td>50</td>
<td>0.22</td>
<td>0.80</td>
<td>0.999</td>
<td>1.00</td>
</tr>
<tr>
<td>100</td>
<td>0.37</td>
<td>0.98</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>500</td>
<td>0.96</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1000</td>
<td>0.999</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
MaxSPRT, Expected Time Until H0 is Rejected, when Rejected

<table>
<thead>
<tr>
<th>T</th>
<th>RR= 1.2</th>
<th>1.5</th>
<th>2.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.26</td>
<td>0.30</td>
<td>0.35</td>
<td>0.37</td>
</tr>
<tr>
<td>5</td>
<td>1.38</td>
<td>1.85</td>
<td>2.10</td>
<td>0.83</td>
</tr>
<tr>
<td>10</td>
<td>3.02</td>
<td>4.05</td>
<td>4.15</td>
<td>0.87</td>
</tr>
<tr>
<td>50</td>
<td>19.87</td>
<td>20.41</td>
<td>8.99</td>
<td>0.96</td>
</tr>
<tr>
<td>100</td>
<td>43.73</td>
<td>29.76</td>
<td>9.35</td>
<td>0.99</td>
</tr>
<tr>
<td>500</td>
<td>172.55</td>
<td>34.30</td>
<td>10.05</td>
<td>1.06</td>
</tr>
<tr>
<td>1000</td>
<td>197.26</td>
<td>35.23</td>
<td>10.33</td>
<td>1.08</td>
</tr>
</tbody>
</table>
### Power vs. Timeliness

<table>
<thead>
<tr>
<th>T</th>
<th>Power</th>
<th>Time</th>
<th>Power</th>
<th>Time</th>
<th>Power</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.11</td>
<td>0.3</td>
<td>0.19</td>
<td>0.4</td>
<td>0.73</td>
<td>0.4</td>
</tr>
<tr>
<td>5</td>
<td>0.19</td>
<td>1.8</td>
<td>0.45</td>
<td>2.1</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>10</td>
<td>0.28</td>
<td>4.0</td>
<td>0.69</td>
<td>4.2</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>50</td>
<td>0.80</td>
<td>20.4</td>
<td>1</td>
<td>9.0</td>
<td>1</td>
<td>1.0</td>
</tr>
<tr>
<td>100</td>
<td>0.98</td>
<td>29.8</td>
<td>1</td>
<td>9.3</td>
<td>1</td>
<td>1.0</td>
</tr>
<tr>
<td>500</td>
<td>1</td>
<td>34.3</td>
<td>1</td>
<td>10.0</td>
<td>1</td>
<td>1.1</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>35.2</td>
<td>1</td>
<td>10.3</td>
<td>1</td>
<td>1.1</td>
</tr>
</tbody>
</table>
Power vs. Timeliness

- A longer maximum surveillance period (T) gives higher power.
- The cost is a slight delay in detection when the null is rejected.
Self Controls

• For each vaccination, use self control, from a pre-exposure or post-post-exposure period
• Compare number of adverse events that are exposed versus unexposed
• Uses a binomial rather than Poisson likelihood
• Define the maximum length of surveillance $T$ in terms of the number of adverse events seen
• Exact critical values can be calculated
Part 2: Vaccine Safety Data Link

Collaborators – partial list
- James Baggs, CDC
- Roger Baxter, NCK
- Bob Davis, CDC
- Bruce Fireman, NCK
- Rich Fox, HAR
- Paul Gargiullo, CDC
- Julianne Gee, CDC
- Jason Glanz, CDC
- Sharon Greene, HAR
- Nicky Klein, NCK
- Margarette Kolczak, CDC
- Tracy Lieu (PI), HAR
- Ned Lewis, NCK
- Renny Li, HAR
- Dave McClure, KPC
- Jennifer Nelson, GHC
- Rich Platt, HAR
- Irene Shui, HAR
- Eric Weintraub, CDC
- Katherine Yih, HAR
- Ruihua Yin, HAR

GHC, Group Health Cooperative; HAR, Harvard; KPC, Kaiser Permanente Colorado; NCK, Northern California Kaiser
Vaccine Safety Datalink (VSD) Project

- Sequential Analysis for Early Detection of Vaccine Adverse Events
- Sponsored by the Centers for Disease Control and Prevention (CDC)
- Collaborative effort of CDC and 8 Health Insurance Plans
- Data on >5.5 million persons annually, ~1.9% of U.S. population
- At the end of 2005: 2.3 million children, 3.2 million adults
VSD Data

Linked by Study IDs
Data are linked and kept at each site, not at CDC

Vaccination Records
Health Outcomes
  (Hospital)
  (Emergency Dept)
  (Outpatient)
Patient Characteristics
  (Birth Certificate)
  (Census / Geocode)
Sequential Analysis

- Used by VSD since 2004 for near real-time monitoring of vaccine safety
- VSD updates data on all vaccines and all outcomes and conducts analyses every week
- All newly approved vaccines are or have been monitored, such as meningococcal, rotavirus, measles-mumps-rubella-varicella, HPV and influenza vaccines
Basic Study Design

- For each vaccine, choose specific outcomes to monitor
- Hypothesis testing, not data mining
- Each week, evaluate the number of outcomes in vaccinated persons
- Compare it to the expected number of outcomes based on a comparison group
Choosing Outcomes

Select outcomes based on:

- Pre-licensure data
- Known biologic properties of the vaccine
- Adverse events for similar vaccines
- Clearly defined, e.g., Guillain-Barre syndrome rather than “neurologic problems”
- Acute-onset
- Relatively uncommon and serious
# Example Comparison Groups

<table>
<thead>
<tr>
<th>Vaccine</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menactra®</td>
<td>Teens making preventive visits</td>
</tr>
<tr>
<td>Rotateq®</td>
<td>Infants who received any other vaccine</td>
</tr>
<tr>
<td>MMRV</td>
<td>Toddlers who received MMR or MMR+V</td>
</tr>
<tr>
<td>Tdap</td>
<td>Teens who received Td</td>
</tr>
<tr>
<td>HPV</td>
<td>Female teens and 18-26 yr old females with preventive visits</td>
</tr>
</tbody>
</table>
Results: Menactra

- No MaxSPRT signals
- That’s good!
When a Statistical Signal Occurs

- In a surveillance setting for multiple vaccines and outcomes, it is not possible to adjust for all possible confounders.
- Not all signals represent a true increase in risk.
- When a signal occurs, we conduct a series of evaluations using traditional epidemiologic methods.
Evaluation of Statistical Signals

1. Check data quality

2. Check whether comparison groups are defined appropriately

3. Conduct the analysis using a different control group (e.g., concurrent vs. historical) or different vaccine
Evaluation of Statistical Signals

4. Conduct a temporal scan to see if outcomes cluster during a post-vaccination time window

5. Conduct a definitive study using logistic regression analysis

6. Review charts to confirm or exclude cases as true cases
Vaccine Safety Datalink Project: Evaluation of MMRV and Febrile Seizures

**Northern California Kaiser Permanente**
- Roger Baxter, MD
- Ned Lewis, MPH
- Bruce Fireman, MS
- Nicola Klein (PI), MD, PhD
- Paula Ray, MPH
- Liisa Lyons
- Pat Ross

**Harvard Pilgrim**
- Tracy Lieu, MD, MPH
- Katherine Yih, PhD, MPH
- Ruihua Yin, MS
- Sharon Greene, PhD, MPH
- Martin Kulldorff, PhD

**CDC**
- Eric Weintraub
- James Baggs, PhD
- Julianne Gee, MPH
- John Iskander, MD, MPH
- Karen Broder, MD
Combination Measles, Mumps, Rubella and Varicella Virus Vaccine (MMRV)

- FDA licensed combined MMRV in 2005 for use in children 12 months to 12 years of age.

- The Advisory Committee on Immunization Practices recommended use of MMRV in 2006.
Overview of MMRV RCA study

- Age: 12-23 months

- Outcomes monitored:
  - Ataxia
  - Seizures
  - Meningitis and encephalitis
  - Thrombocytopenia
  - Arthritis
  - Allergic reactions

- Post vaccination observation for 42 days.

- Expected rates of seizures, ataxia, and allergic reactions were calculated based on historical rates among MMR recipients (with or without varicella vaccine).

MaxSPRT Seizure Signal

- The number of observed seizures in the 42 day post-vaccination time window first exceeded the number expected by enough to justify a signal in the week of 2/11/07.

- Cumulative doses at that time: 25,779

<table>
<thead>
<tr>
<th>Number Seizures</th>
<th>Observed</th>
<th>Expected</th>
<th>Relative Risk</th>
<th>LLR (critical value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>59</td>
<td>38</td>
<td>1.57</td>
<td>5.17 (4.12)</td>
</tr>
</tbody>
</table>
Temporal Scan Statistics Results on Seizures in 42 Days after Vaccination

<table>
<thead>
<tr>
<th></th>
<th>MMRV</th>
<th>MMR+V*</th>
<th>MMR w/o V*</th>
<th>V* w/o MMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total cases</td>
<td>93</td>
<td>164</td>
<td>32</td>
<td>101</td>
</tr>
<tr>
<td>Most likely cluster</td>
<td>Days 7-10</td>
<td>Days 7-10</td>
<td>Days 6-10</td>
<td>Days 21-24</td>
</tr>
<tr>
<td>Cases in cluster</td>
<td>45</td>
<td>44</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>RR</td>
<td>8.9</td>
<td>3.5</td>
<td>3.9</td>
<td>2.5</td>
</tr>
<tr>
<td>P-value</td>
<td>0.00001</td>
<td>0.00001</td>
<td>0.063</td>
<td>0.047</td>
</tr>
</tbody>
</table>

* V= varicella vaccine
Temporal distribution of seizures after MMRV vaccination

Days Post-MMRV Vaccine (2/06-9/07, after 47,137 vaccine visits) vs. Number of Seizures
Temporal distribution of seizures after simultaneous MMR and varicella vaccination

(2004-2005, ~90,000 vaccine visits)
Temporal distribution of seizures after MMR vaccination without varicella vaccination

(2004-05, ~23,500 vaccine visits)
Temporal distribution of seizures after varicella vaccination without MMR

(77,875 vaccine visits 2000-2006)
Outpatient Visits for Fever by Day after Vaccine at Northern California Kaiser Permanente: 1995-2008

6241 total fever visits after 302,670 MMR+V, 147,762 MMR, 46,390 MMRV, 38,251 VZV
Unadjusted Rates of Seizures 7-10 Days Post-Vaccination

Unadjusted rates:
- MMRV: 9.6/10,000
- MMR + V: 4.9/10,000
- MMR alone: 3.5/10,000
- Varicella alone: 1.5/10,000

* V = varicella vaccine
Logistic Regression Analysis: Risk of Seizure 7-10 days after MMRV Compared to MMR + Varicella Vaccines

<table>
<thead>
<tr>
<th></th>
<th>Odds ratio*</th>
<th>95% Confidence Interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMRV versus MMR + V</td>
<td>2.0</td>
<td>1.4, 2.8</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

*Adjusted for age and influenza season.

None of the following influenced the association between MMRV and seizures: Sex, VSD site, concomitant vaccines and seizure temporal trends.

N for MMRV = 43,356, MMR + V = 314,625
Majority of Charts Confirmed Seizures as Febrile

<table>
<thead>
<tr>
<th></th>
<th>MMRV (n=45)</th>
<th>MMR + V* (n=132)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Febrile seizure</td>
<td>42 (93%)</td>
<td>124 (94%)</td>
</tr>
<tr>
<td>Afebrile</td>
<td>3 (7%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>5 (4%)</td>
</tr>
</tbody>
</table>

*varicella vaccine
Logistic Regression Analyses:
Risk of seizure 7-10 days Post-Vaccination using Chart Verified Febrile Seizures

<table>
<thead>
<tr>
<th>Odds ratio*</th>
<th>95% Confidence Interval</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMRV versus MMR + V</td>
<td>2.3</td>
<td>1.6, 3.2</td>
</tr>
</tbody>
</table>

*Adjusted for age and influenza season.

N for MMRV = 43,353, MMR + V = 314,599
Risk Difference during 7-10 Day Post-Vaccination Window

- Attributable Risk for MMRV compared to MMR + varicella vaccines.

\[
\frac{5.2}{10,000} \text{ (95\% CI 2.2, 8.1)}
\]

For every 10,000 children who receive MMRV instead of separate MMR + varicella vaccines, there will be approximately 5 additional seizures 7-10 days after vaccination.

- Inverse of the above risk difference for MMRV compared to MMR + varicella vaccines in the 7-10 day window (Number Needed to Harm):

\[
1,939 \text{ (95\% CI 1,234, 4,516)}
\]

There will be approximately 1 additional seizure 7-10 days post-vaccination for every 2000 children vaccinated with MMRV instead of MMR + varicella vaccine.
Conclusions

• On February 27, 2008, the Advisory Committee on Immunization Practices (ACIP) revised its recommendation by a vote of 10-2, no longer recommending the MMRV vaccine over separate MMR and Varicella vaccines.

• On the same day, Merck and FDA revised the product label for MMRV, including information about the increased risk of seizures 7-10 days after vaccination.

• In 2009, after additional analyses, ACIP reaffirmed its decision not to recommend either MMRV or MMR+V over the other, leaving the decision to physicians and parents.

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.
Part 3: Continuous versus Group Sequential Analysis for Safety Surveillance

Ivair Silva
Martin Kulldorff
Harvard Medical School and
Harvard Pilgrim Health Care

Funded by: FDA – MiniSentinel – PRISM – Activity 10
Continuous Sequential Analysis

- Used when the data is monitored in a continuous or near continuous fashion
- Developed by Wald in the 1940s
- Used for industrial applications, but not much for clinical trials
- Actual analyses are only needed when a new event occurs
- Allows for any number of looks and analyses of the data
- May set requirement that at least 3 or 4 events occur before generating a signal
Group Sequential Analysis

- Used when there is a finite number of analyses performed, usually in the range 2-10
- Large statistical literature
- Method of choice for sequential analyses of clinical trial
Statistical Analysis
When to Use Which Approach?

• For data that arrives in a few large chunks, use group sequential methods
• For data that arrives continuously, use continuous sequential methods
• For data that arrives frequently, in 50 or 100+ batches, continuous sequential methods are often computationally more convenient, providing better statistical approximations
Study Design
When to Use Which Approach?

• Can we increase statistical power by looking less frequently, using group sequential analysis?

• If it is more costly to get more frequent data, what are the benefits?
Methods to be Compared

- A single, non-sequential analysis
- Group sequential, with e.g. \( G = 2, 5, 10 \) or 100 equally spaced analyses
- Continuous sequential
- Continuous sequential that requires at least \( M = 3, 6 \) or 10 events to signal
Continuous versus Group
Sequential Analysis

- Poisson based model
- Same rejection boundary shapes
Sequential Analysis Boundaries

Reject H0

0

Accept H0

T

time

LLR
Continuous versus Group Sequential Analysis

For Comparison:
- Poisson based model
- Same rejection boundary shapes
- For group sequential, fixed LLR does not give exact alpha level
- To make a fair comparison with exactly the same alpha levels, one must choose a random LLR boundary
Standard Non-Sequential Analysis

- Type 1 error (alpha)
- Statistical power
- Sample size
Standard Non-Sequential Analysis

- Type 1 error (alpha)
- Statistical power
- Sample size

If this is all that matters, don’t do a sequential analysis!!
Sequential Analysis

- Type 1 error (alpha)
- Statistical power
- Final sample size when the null is not rejected, i.e. maximum length of surveillance
- Expected time to signal when the null is rejected

There is a trade-off between these four metrics
Keeping Maximum Sample Size Fixed

• Power is by definition highest for (i) a single non-sequential analysis, followed by (ii) a group sequential analysis, and then (iii) a continuous sequential analyses.

• Trade-off between power and expected time to signal when the null hypothesis is rejected
Keeping Maximum Sample Size Fixed

• But, we can always get the statistical power we want by continuing to collect and analyze more data

• In observational safety surveillance, continued surveillance is cheap

(exception: flu vaccine)
Keep Statistical Power Fixed

• Hold both alpha and the power fixed
• Trade-off between time to signal when the null hypothesis is rejected and the length of surveillance when it is not rejected.
Results: Expected Time to Signal as a Function of Power

- Defined in terms of expected counts under the null
- Think of it as the number of vaccinations
Time to Signal as a Function of the Power RR=2

<table>
<thead>
<tr>
<th>Power</th>
<th>Continuous M</th>
<th>Group G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>0.50</td>
<td>2.5</td>
<td>2.1</td>
</tr>
<tr>
<td>0.60</td>
<td>3.2</td>
<td>2.9</td>
</tr>
<tr>
<td>0.70</td>
<td>4.3</td>
<td>3.8</td>
</tr>
<tr>
<td>0.80</td>
<td>5.3</td>
<td>4.8</td>
</tr>
<tr>
<td>0.85</td>
<td>5.9</td>
<td>5.4</td>
</tr>
<tr>
<td>0.90</td>
<td>6.5</td>
<td>6.0</td>
</tr>
<tr>
<td>0.95</td>
<td>7.5</td>
<td>6.9</td>
</tr>
<tr>
<td>0.98</td>
<td>8.1</td>
<td>7.6</td>
</tr>
<tr>
<td>0.99</td>
<td>8.4</td>
<td>7.9</td>
</tr>
</tbody>
</table>
Relative Risk

- Clinically important relative risk, for which the analysis is powered (e.g. RR=2)
- True relative risk, for which we evaluate the expected time to signal (e.g. RR=4)
Results: Maximum Length of Surveillance as a Function of Power

- Defined in terms of expected counts under the null
- Think of it as the number of vaccinations
<table>
<thead>
<tr>
<th>Power</th>
<th>0.50</th>
<th>0.60</th>
<th>0.70</th>
<th>0.80</th>
<th>0.85</th>
<th>0.90</th>
<th>0.95</th>
<th>0.98</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.0</td>
<td>7.7</td>
<td>10.3</td>
<td>13.5</td>
<td>15.6</td>
<td>18.2</td>
<td>22.9</td>
<td>28.8</td>
<td>32.8</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.8</td>
<td>6.7</td>
<td>9.2</td>
<td>12.1</td>
<td>14.2</td>
<td>17.1</td>
<td>21.6</td>
<td>27.1</td>
<td>31.1</td>
</tr>
<tr>
<td>6</td>
<td>3.9</td>
<td>5.8</td>
<td>8.0</td>
<td>11.0</td>
<td>13.1</td>
<td>15.8</td>
<td>20.3</td>
<td>25.9</td>
<td>29.9</td>
</tr>
<tr>
<td>10</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>9.9</td>
<td>n/a</td>
<td>11.9</td>
<td>19.0</td>
<td>24.6</td>
<td>28.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.8</td>
<td>5.2</td>
<td>7.1</td>
<td>9.4</td>
<td>11.0</td>
<td>13.3</td>
<td>16.7</td>
<td>21.6</td>
<td>25.4</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>6.0</td>
<td>8.1</td>
<td>10.3</td>
<td>11.9</td>
<td>14.3</td>
<td>17.7</td>
<td>23.6</td>
<td>27.1</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>6.3</td>
<td>8.0</td>
<td>10.8</td>
<td>12.8</td>
<td>15.2</td>
<td>19.2</td>
<td>25.1</td>
<td>27.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.3</td>
<td>8.8</td>
<td>11.2</td>
<td>12.8</td>
<td>15.7</td>
<td>19.6</td>
<td>25.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9.3</td>
<td>12.1</td>
<td>14.6</td>
<td>16.7</td>
<td>21.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>4.9</td>
<td>6.3</td>
<td>8.8</td>
<td>11.2</td>
<td>12.8</td>
<td>15.7</td>
<td>19.6</td>
<td>25.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.3</td>
<td>7.1</td>
<td>9.3</td>
<td>12.1</td>
<td>14.6</td>
<td>16.7</td>
<td>21.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Theoretical Result

Theorem: For any group sequential design, there always exist a continuous design for which all of the four metrics below are at least as good and at least one is better:

- Alpha level
- 1. Statistical power
- 2. Expected time to signal when the null is rejected
- 3. Length of surveillance when null is not rejected
Conclusions

• More frequent data is always better.
• Do not deliberately delay analyses. Look as soon as the data arrives.
• Data from different health plans do not have to be synchronized.
• If data arrives in frequent batches, continuous sequential analysis is the most natural choice.
• For some study designs with covariates, only group sequential methods exist.
• Some study designs with propensity scores may need a large batch for the first analyses, but can be analyzed near-continuously thereafter.
Thank You!