Using large healthcare databases in the USA to assess vaccine safety

Jennifer Clark Nelson, PhD
Biostatistics Unit, Group Health Research Institute
Department of Biostatistics, University of Washington

October 29, 2014
Vaccine Safety Datalink’s (VSD’s) general approach to proactive vaccine safety monitoring

Example VSD study
  → Safety of Pentacel combination vaccine

Methodological challenges and opportunities

Comparison of methods
  → Measles-mumps-rubella-varicella (MMRV) vaccine safety

Summary
CDC-sponsored Vaccine Safety Datalink

- Large population: >9.6 million (>3% of U.S. population)
- Population-based: members from 10 (now 6) health plans
Real-Time Vaccine Safety Surveillance for the Early Detection of Adverse Events

Tracy A. Lieu, MD, MPH,*† Martin Kulldorff, PhD,* Robert L. Davis, MD, MPH,*†
Edwin M. Lewis, MPH,§ Eric Weintraub, MPH,‡ Katherine Yih, PhD, MPH,* Ruihua Yin, MS,*
Jeffrey S. Brown, PhD,* and Richard Platt, MD, MSc,* for the Vaccine Safety Datalink Rapid Cycle Analysis Team

Background: Rare but serious adverse events associated with vaccines or drugs are often nearly impossible to detect in prelicensure studies and require monitoring after introduction of the agent in large populations. Sequential testing procedures are needed to detect vaccine or drug safety problems as soon as possible after introduction.

Objective: To develop and evaluate a new real-time surveillance system that uses dynamic data files and sequential analysis for early detection of adverse events after the introduction of new vaccines.

Research Design: The Centers for Disease Control and Prevention (CDC)-sponsored Vaccine Safety Datalink Project developed a real-time surveillance system and initiated its use in an ongoing study of a new meningococcal vaccine for adolescents. Dynamic data files from 8 health plans were updated and aggregated for analysis every week. The analysis used maximized sequential probability ratio testing (maxSPRT), a new signal detection method that supports continuous or time-period analysis of data as they are collected.

Conclusions: Real-time surveillance combining dynamic data files, aggregation of data, and sequential analysis methods offers a useful and highly adaptable approach to early detection of adverse events after the introduction of new vaccines.

Key Words: vaccine safety, active surveillance, sequential analysis, meningococcal vaccine, drug safety

(Med Care 2007;45: S89–S95)

Concerns about the safety of vaccines and drugs introduced in recent years have highlighted the need to enhance systems for early detection of potential adverse events. Uncommon but serious adverse events have led to the withdrawal of both biologic and pharmacologic agents from the market. Examples include the discontinuation of rotavirus vaccine after reports of intussusception (a rare bowel obstruction), and modification of the influenza vaccine formulation after two deaths were attributed to influenza vaccine-induced encephalitis.
VSD’s general approach

- Identify primary exposure and comparator populations
  - Historical comparators (compute expected counts)
  - Concurrent comparators (e.g., new user design)
  - Within-person comparisons (self-controlled designs)

- Pre-specify adverse event (AE) outcomes of interest
  - 5-10 AEs (AERs, pre-licensure, use outside USA)

- Identify potential confounders
  - Age, gender, site, comorbidities, concomitant medications

- Routine dataset construction and sequential analysis
  - Update de-identified data regularly (e.g. weekly, monthly,…)
  - Repeatedly test hypotheses of interest over time
  - Signaling threshold controls overall false positive error rate
Specify sequential design
- Monitoring frequency: daily, weekly, monthly, quarterly
- Boundary shape: flat (Pocock), decreasing (O’Brien-Fleming)
- Effect measure: relative risk or risk difference
  - $H_0: \text{RR}=1$ vs $H_A: \text{RR}>1$
  - Large value implies increased risk among exposed

After each new observation or group accrues...
- Count events and person-time among exposed & unexposed
- Compute test statistic, $Z$, to compare risk between groups
- If $Z > B$: stop, signal $H_A$; else continue
- If end of study and no signal, fail to reject $H_0$
- $B$ chosen to maintain a preset false positive (FP) error
Example sequential monitoring design (LRT statistic & flat stopping boundary)

**Diagram:**
- **log(LRT)** vs **# of events**
- **STOP, reject H₀**
- **STOP, fail to reject H₀**
- **Continuation region**
- **Upper time limit**

**Note:** B is chosen to maintain a pre-set false positive (FP) error
Hypothetical observed data and resulting test statistic trajectory

H_A: RR > 1
H_0: RR = 1

log(LRT)
Example VSD study (Pentacel vaccine)

- **A combination DTaP-IPV-Hib vaccine** (diphtheria & tetanus toxoids & acellular pertussis adsorbed, inactivated poliovirus, & Haemophilus influenza b)
- **Licensed in 2008** for children at 2, 4, 6 & 15-18 mos. of age to replace separate component injections
- **Aim:** To sequentially monitor the safety of the new Pentacel® vaccine among the VSD cohort of children 6 weeks through 2 years of age.
Study measurements and design

- **Population:** Children 6 weeks – 2 years from 7 VSD sites
  - Primary exposure: Pentacel vaccine, received 2008-11
  - Historical control: Other DTaP vaccines, 2-4 years prior

- **Pre-specified AEs:**

<table>
<thead>
<tr>
<th>AE</th>
<th>ICD-9 Codes</th>
<th>Interval (days)</th>
<th>Visit type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seizure</td>
<td>345, 780.3</td>
<td>0-7</td>
<td>Inpatient ED</td>
</tr>
<tr>
<td>Fever</td>
<td>780.6</td>
<td>1-5</td>
<td>All</td>
</tr>
<tr>
<td>Serious non-anaphylactic allergic reaction</td>
<td>995.1-2, 708.0-1, 708.9</td>
<td>1-2</td>
<td>Inpatient ED</td>
</tr>
</tbody>
</table>

- **Potential confounders:** age, gender, site
Sequential design and analysis

- **Frequency of testing** (data accruing weekly)
  - 1st test at 1 year (~33,000 doses as of Sept 2009)
  - 11 subsequent tests, equally-spaced based on dose

- **Test statistic**
  - Likelihood ratio test ($H_0: \text{RR}=1$ versus $H_A: \text{RR}>1$)
  - Confounder control: used site, gender, & age-based historical event rates to compute expected counts

- **Stopping boundary**: flat over time (Pocock)

- **Sample size**: to detect RR=2.0 with >80% power
  - Fever/Seizure: ~73,000 (test every 3500 doses)
  - Meningitis/Allergic Rxn: ~150,000 (test every 10,500)
Results

- No significant increased risk detected for any outcome
- Example: Medically-attended fever

- 72,651 doses from Sept ‘08 – Feb ’10
- 348 events (48/10K doses) versus 318 events expected
- RR=1.09, p=0.06 (LLR=1.40 vs. 1.79 critical threshold)
Methodological challenges

1. Defining the scope of surveillance
   - Primary aim as well as scale

2. High priority on rapid detection
   - Motivates use of sequential monitoring
   - How to best select a sequential design?

3. Lack of a controlled research setting
   - Unpredictable uptake, population, and data
   - Heterogeneity across sites

4. Safety outcomes are rare

5. Data come from a distributed environment
1a. Defining primary aim of surveillance

**Signal Generation**
To identify potential associations between *any* medical product & adverse event (often 100’s, 1000’s of pairs).

*Example:* Data mining of passive adverse event reports

**Signal Refinement**
To *rapidly* evaluate the magnitude of suspected associations between several target product-event pairs.

Prospective surveillance intends to fill this gap – but exactly what should this look like?

**Signal Evaluation**
To establish or refute causality between a particular product and adverse event of interest.

*Example:* Phase IV trial or rigorous epidemiological cohort study
1b. Scaling up

- Routine prospective surveillance is desirable for...
  - Many products, many events, many populations
  - FDA in 2011: 30 new drugs; ~30-50 adverse events of interest
- But resources are limited
  - Can one (or a few) design and analysis strategies be re-used across many product-event pairs?
  - Or, should strategies be customized (i.e., a full protocol) for each pair to ensure validity?
  - Balance: do many things versus do a few (well)
- What about multiple comparisons?
  - Increased potential for false positive signals (costly)
  - Simple corrections possible, but largely unaddressed
Ongoing or completed prospective surveillance within VSD since 2005

Menactra – Guillain-Barre Syndrome (GBS), others (Harvard)
Rotateq/Rotarix – intussusception, others (Marshfield)
MMRV – seizures, fever, others (N California Kaiser)
Tdap -- seizures, other outcomes (Health Partners)
HPV– seizures, syncope, stroke, VTE (Kaiser NW, CDC)
Seasonal & H1N1 Influenza – conducted annually (Harvard, CDC)
Kinrix – seizures, stroke, GBS, others (Kaiser Colorado)
Pentacel – fever, seizure, allergic reactions, others (Group Health)
PCV13 – seizures, Kawasaki disease, others (S California Kaiser)
2. High priority on rapid surveillance

- Introduces need for sequential testing
  - Allows detection as soon as possible, as data accrue
  - But, increases complexity
    - Additional design decisions (e.g., how often to test?)
    - Statistical analysis complications (multiple testing with stopping rules, maintaining overall false positive rate)

- Established approach in randomized trials

- A challenge to adapt to observational safety setting
  - Research question is different (safety vs efficacy)
  - Setting is different (not a randomized experiment)
How to select a sequential design?

- Designs geared for efficacy in trials are well-studied
  - What test statistic is best to compare risks, make decisions?
  - How frequently should testing be performed?
  - At what level should stopping boundaries be set?

- Post-market safety implies different scientific/practical issues
  - Implications of stopping, costs of false positive/negative signals
  - Higher frequency monitoring? Conservative (lower) boundaries?
  - Delaying the first test until pre-licensure N’s are achieved?

- Optimal design(s) for rare safety endpoint not known
  - More dialogue between clinicians and statisticians needed
  - More statistical information needed to inform choices
  - Magnitude of trade-offs between power & timeliness not known
Sequential boundary examples

Typical efficacy trial
- **Frequency**: quarterly
- **Boundary**: decreasing
- **Test statistic**: varies (LRT, RR, risk difference)

Vaccine Safety Datalink
- **Frequency**: weekly
- **Boundary**: flat
- **Test statistic**: LRT
Pros (+) & cons (-) of VSD initial approach

+ Highly frequent testing → shorter average time-to-detection
+ Highly frequent testing is feasible, has been successful
- Highly frequent testing
  - Resource intensive weekly analysis & review
  - May sacrifice data quality
  - May not be necessary
    - Pre-licensure testing (when safety is less certain) isn’t as frequent
    - Rare AE’s not found pre-licensure won’t accumulate quickly
  - Is less powerful than less frequent testing
- Flat boundary (vs. higher early lower later like O’Brien-Fleming)
  - Powerful at early time points, but less powerful overall
  - More false positives early when greater uncertainty in data
3a. Lack of a controlled setting: Unpredictable uptake rate

- Slow uptake implies instability in early tests
- Lack of control over uptake → harder to plan
  - Want to detect RR=2 for seizure risk
  - Specify 80% power
  - Requires sample size of N=73,000 doses
  - How long to accrue? (1.5 years)
3b. Lack of a controlled setting: Unpredictable population over time

- Confounder composition changes over time
- Impacts ability to adjust for confounding due to lack of overlap in exposed/unexposed
- Standard adjustment methods (propensity scores) harder to estimate at early surveillance time points when exposure is rare
- Impacts sequential boundary plans and formulation
3c. Lack of a controlled setting: Heterogeneity across sites

- Site differences can be big (e.g., N, uptake rate)

- Stratifying or conducting subgroup analyses by site is critical
- When to begin to include information from a contributing site?
- When is it reasonable to combine data across sites?
- How to best account for site variability in overall estimates?
3d. Lack of a controlled setting: Unpredictable data

- **Electronic data accessed in real-time are dynamic**
  - Exposure and adverse event data ‘arrive late’
  - People (exposures and events) can disappear

- **Results vary depending on how you deal with this**
  - Analysis approach #1 (o+n): freeze old and add new data
  - Analysis approach #2 (cum): cumulatively refresh all data

<table>
<thead>
<tr>
<th>Adverse event (AE) outcome</th>
<th>Total # of doses</th>
<th># of AEs</th>
<th>Expected # of AEs</th>
<th>AE rate per 10K</th>
<th>RR</th>
<th>LLR</th>
<th>LLR Critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever: o+n</td>
<td>66,400</td>
<td>343</td>
<td>303.5</td>
<td>51.7</td>
<td>1.13</td>
<td>2.47</td>
<td>1.93</td>
</tr>
<tr>
<td>Fever: cum</td>
<td>68,826</td>
<td>335</td>
<td>302.5</td>
<td>48.7</td>
<td>1.11</td>
<td>1.69</td>
<td>1.85</td>
</tr>
<tr>
<td>Seizure: o+n</td>
<td>66,400</td>
<td>8</td>
<td>7.9</td>
<td>1.2</td>
<td>1.01</td>
<td>&lt;0.01</td>
<td>1.91</td>
</tr>
<tr>
<td>Seizure: cum</td>
<td>68,826</td>
<td>9</td>
<td>8.1</td>
<td>1.3</td>
<td>1.11</td>
<td>0.04</td>
<td>1.46</td>
</tr>
</tbody>
</table>
4. Adverse events are rare

Seizure risk over time

- Large sample sizes needed for adequate power
  - 2 seizure events/10K doses
  - N=73K needed for RR=2
- Large sample assumptions of standard methods may fail
  - Use exact testing methods
  - Handle zero cell counts
- Variability at early test points
  - Each (possibly misclassified) event is highly influential
  - Must guard against early false positives by chance
Errors can be more influential

- Outcome misclassification result in false signals
- Need to assess accuracy in advance to determine ‘fitness-for-purpose’
- Conduct and use more valid endpoints
- Conduct sensitivity analyses to assess potential impact
5. Data are distributed

- For privacy and proprietary reasons
- No combined individual-level dataset across sites
- De-identified combined central datasets
  - Aggregated counts of events and person-time by categorical exposure and confounder strata
  - Limited individual-level data: binary event & exposure and a summary confounder score (propensity score)
  - Combined central dataset with summary statistics
- Inherently constrains analytic options
- Can present a barrier for signal follow-up
Measles-mumps-rubella-varicella (MMRV) combination vaccine safety comparisons

- **Original VSD active surveillance using historical controls:**
  - Signaled after 43,353 MMRV doses; Adjusted RR ~2 (exact RR not reported) using Poisson MaxSPRT (continuous testing method)

- **VSD follow-up study using prospective controls + chart review**
  - Adjusted RR of 1.98
  - Adjusted RD of 4.3 per 10K vaccinated
  - 83,107 MMRV and 376,354 MMR+V with chart reviewed outcomes

- **Mini-Sentinel Group Sequential GEE regression application:**
  - Signaled after 48,233 MMRV doses
  - Adjusted OR of 2.37 (metric upon which signal was based)
  - Adjusted RD of 5.3 per 10K vaccinated

- **Mini-Sentinel Group Sequential IPTW regression application:**
  - Signaled after 17,321 MMRV doses
  - Adjusted RR of 2.86
  - Adjusted RD of 5.2 per 10K vaccinated (metric upon which signal based)
Summary

- New systems and methods are promising
- Many methodological challenges are faced
  - Defining scope of surveillance
  - Applying sequential testing in observational settings
  - Lack of a controlled setting
  - Safety outcomes are rare
  - Data are from a distributed setting
- Need careful planning (choice of method matters)
- There are many existing and emerging design and analysis strategies that should be considered
Acknowledgements

Group Health Research Institute / University of Washington
Andrea Cook, Rob Wellman, Onchee Yu, Lisa Jackson, James Floyd, Susan Heckbert, Patrick Heagerty, Bruce Psaty

VSD Investigators, including CDC
Sharon Greene, Katherine Yih, Lingling Li, Matthew Daley, Steven Jacobsen, Nicola Klein, Eric Weintraub, Karen Broder

Mini-Sentinel Investigators, including FDA
Rich Platt, Jeff Brown, Grace Lee, Darren Toh, Azadeh Shoabi

Funding
This work was supported by a subcontract with America's Health Insurance Plans (AHIP) under contract 200-2002-00732 from CDC as well as the FDA Mini-Sentinel contract.
Rotashield vaccine and rotavirus
- Licensed in 1998, 1st effective vaccine to prevent rotavirus
- Most common cause of diarrhea among children
- 55K hosp/year in US; 600K deaths/year worldwide

Pre-licensure intussusception (rare bowel obstruction) data
- 27 pre-licensure trials of rotavirus vaccines
- 5 cases/10K vaccinees (0.05%) vs 1/4633 for placebo (0.02%)
- Not statistically difference, but noted on package insert

Post-licensure data
- 15 reports of intussusception to VAERS
- Rotashield withdrawn in July 1999 (9 months after licensure)
- Risk after 1st dose: 20-30 times that expected
A. PATIENT INFORMATION
1. Patient Identifier
2. Age at Time of Event, or Date of Birth:
   Female
   Male
3. Sex
4. Weight
   lb
   kg

B. ADVERSE EVENT, PRODUCT PROBLEM OR ERROR
Check all that apply:
1. Adverse Event
   Product Problem (e.g., defects/malfunctions)
   Product Use Error
   Problem with Different Manufacturer of Same Medicine
2. Outcomes Attributed to Adverse Event
   (Check all that apply)
   • Death: (mm/dd/yyyy)
   • Life-threatening
   • Disability or Permanent Damage
   • Congenital Anomaly/Birth Defect
   • Hospitalization - initial or prolonged
   • Other Serious (Important Medical Events)
   • Required Intervention to Prevent Permanent Impairment/Damage (Devices)
3. Date of Event (mm/dd/yyyy)
4. Date of this Report (mm/dd/yyyy)
5. Describe Event, Problem or Product Use Error

D. SUSPECT PRODUCT(S)
1. Name, Strength, Manufacturer (from product label)
2. Dose or Amount
3. Dates of Use (if unknown, give duration) from/to (or best estimate)
4. Diagnosis or Reason for Use (Indication)
5. Event Abated After Use Stopped or Dose Reduced?
   Yes
   No
   Doesn't Apply

E. SUSPECT MEDICAL DEVICE
1. Brand Name
2. Common Device Name
3. Manufacturer Name, City and State
4. Model #
   Lot #
   Catalog #
   Expiration Date (mm/dd/yyyy)
   Serial #
   Other #
5. Operator of Device
   Health Professional
   Lay User/Patient
   Other:
6. If Implanted, Give Date (mm/dd/yyyy)
7. If Explanted, Give Date (mm/dd/yyyy)
8. Is this a Single-use Device that was Reprocessed and Reused on a Patient?
   Yes
   No
9. If Yes to Item No. 8, Enter Name and Address of Reprocessor

6. Relevant Tests/Laboratory Data, Including Dates
The bottom line

- Existing gaps in safety evidence are problematic
- Some unsafe products are on the market too long
  - Rotashield: removed after 9 months due to VAERS
  - Vioxx: removed after 5 years due to Phase IV trial
- No single existing system is simultaneously
  - Large, generalizable (gap for pre-licensure RCTs)
  - Proactive, prospective, and population-based (gap for passive reports)
  - Available rapidly enough (gap for Phase IV studies)
Supplemental data elements

- Laboratory results
- Electronic medical records
  - Vital signs (e.g., BP, weight, height)
  - Procedure results (e.g., EF for heart failure, radiology images)
  - Indication for prescription
  - Flow sheets (PHQ-9, Medicare wellness survey)
- Disease registries (e.g., cancer, diabetes)
- Mortality and cause of death
- Census data
- Provider characteristics (e.g., gender, years practice)
- Patient characteristics (e.g., race, education, smoking status, family history, health behaviors)
Medication data

- Drug name, strength, date dispensed, quantity, form, route, days supply, provider, out pocket cost, NDC, therapeutic class, for each medication dispensed or claim submitted
  - Supplemental data may include directions for use, indication, intended duration, completion of therapy (e.g., chemotherapy)
- Record for each dispensing
VSD Population

- Collects medical care and vaccination data on more than 9.6 million members annually (3.1% of the US population)
- As of 12/31/2010:
  - 2,238,537 children (<18) enrolled
    - 3.0% of US population
  - 7,357,734 adults (≥18) enrolled
    - 3.1% of US population
  - Average yearly birth cohort ~ 96,000
VSD Centralized Data Model

- Northwest Kaiser
- Northern California Kaiser
- Southern California Kaiser
- Harvard Pilgrim
- Kaiser Colorado
- Marshfield
- Group Health
- Health Partners
- Medical Record
- Medical Record
- Medical Record
- Medical Record
- Medical Record
- Medical Record
The VSD Distributed Data Model

SAS Programs, Logs, Output, & Analytical Datasets
HMO Research Network

- Consortium of 16 health plans and their affiliated research centers
- **Mission**: To facilitate and promote the unique research capabilities of member organizations
  - Fostering collaborative studies
  - Sharing methodologies & best practices
  - Disseminating & translating research findings
- Combined population = ~15 million individuals
- Geographically & socio-culturally diverse
Interconnected Data Resources at GHRI and other HMORN partners
Virtual Data Warehouse (VDW)

VDW is populated by automated data from the following sources:

- Tumor registry
- Enrollment
- Demographics
- Pharmacy
- Utilization – includes diagnoses & procedures
- Some vitals such as BMI & blood pressure
- Geocoding
- Laboratory
- Chemotherapy
- Radiology
- Pathology
Example of longitudinal data within HMORN

GHC Data Availability

Source:
- Tumor
- Census
- Multiple Cause of Death
- Death
- Provider Specialty
- Vitals
- Outpatient Pharmacy
- Lab
- Procedures
- Diagnoses
- Utilization
- Enrollment
- Demographics

Years:
- 1950
- 1960
- 1970
- 1980
- 1990
- 2000
- 2010
Group sequential methods with confounder adjustment (concurrent controls)

<table>
<thead>
<tr>
<th>Method</th>
<th>Used in</th>
<th>Statistic</th>
<th>Boundary</th>
<th>Confounding</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS-LD</td>
<td>RCTs</td>
<td>Any</td>
<td>Error Spending, using normal approx</td>
<td>Any (Matching, Stratification, Regression)</td>
<td>Easy to apply; flexible confounding control</td>
<td>Assumptions ok? (normal approx, indep data increments)</td>
</tr>
<tr>
<td></td>
<td>Lan &amp; Demets 1983</td>
<td>standardized normal</td>
<td>(even spending, flat/Pocock-like)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS-LRT</td>
<td>RCTs, VSD</td>
<td>LRT</td>
<td>Unifying (or error spending) boundary, permutation derived (flat/Pocock-like)</td>
<td>Matching with fixed (1:M) ratio</td>
<td>Easy to apply; appealing interpretation</td>
<td>Information loss (restricted sample) too strict→lose cases; too loose→poor control</td>
</tr>
<tr>
<td></td>
<td>Lai 1991 (GLRs) Kulldorff 2011 (MaxSPRT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSSP</td>
<td>Drug safety</td>
<td>Number of AEs among exposed</td>
<td>Error spending, conditions on total # of AEs within strata (even spending)</td>
<td>Stratification (including on calendar time)</td>
<td>Works well for rare AEs</td>
<td>May not hold Type I error if strata small or testing highly frequent</td>
</tr>
<tr>
<td></td>
<td>Li 2009 &amp; 2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GS-EE</td>
<td>VSD</td>
<td>Score statistic</td>
<td>Unifying (or error spending), boundary permutation derived (flat/Pocock-like)</td>
<td>Regression</td>
<td>Flexible confounding control; few assumptions</td>
<td>Requires ‘enough’ AEs to estimate parameters; computer intensive</td>
</tr>
<tr>
<td></td>
<td>Cook 2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Outcome over time amongst MMRV recipients

<table>
<thead>
<tr>
<th></th>
<th>Look 1</th>
<th>Look 2</th>
<th>Look 3</th>
<th>Look 4</th>
<th>Look 5</th>
<th>Look 6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total, Outcome(%)</strong></td>
<td>5 (0.179)</td>
<td>10 (0.143)</td>
<td>14 (0.121)</td>
<td>18 (0.104)</td>
<td>21 (0.087)</td>
<td>29 (0.090)</td>
</tr>
<tr>
<td><strong>Age, Outcome(%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11m-12m</td>
<td>1 (0.089)</td>
<td>4 (0.135)</td>
<td>5 (0.100)</td>
<td>7 (0.091)</td>
<td>9 (0.079)</td>
<td>13 (0.084)</td>
</tr>
<tr>
<td>13m-14m</td>
<td>1 (0.144)</td>
<td>2 (0.129)</td>
<td>3 (0.118)</td>
<td>3 (0.079)</td>
<td>4 (0.077)</td>
<td>6 (0.087)</td>
</tr>
<tr>
<td>15m-16m</td>
<td>2 (0.312)</td>
<td>2 (0.124)</td>
<td>2 (0.075)</td>
<td>3 (0.080)</td>
<td>3 (0.062)</td>
<td>4 (0.065)</td>
</tr>
<tr>
<td>17m-19m</td>
<td>1 (0.417)</td>
<td>2 (0.338)</td>
<td>3 (0.295)</td>
<td>3 (0.194)</td>
<td>3 (0.152)</td>
<td>4 (0.160)</td>
</tr>
<tr>
<td>20m-23m</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
<td>1 (0.255)</td>
<td>2 (0.313)</td>
<td>2 (0.229)</td>
<td>2 (0.174)</td>
</tr>
<tr>
<td><strong>Sex, Outcome(%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3 (0.212)</td>
<td>5 (0.141)</td>
<td>7 (0.120)</td>
<td>9 (0.102)</td>
<td>10 (0.082)</td>
<td>13 (0.080)</td>
</tr>
<tr>
<td>Female</td>
<td>2 (0.145)</td>
<td>5 (0.146)</td>
<td>7 (0.122)</td>
<td>9 (0.105)</td>
<td>11 (0.092)</td>
<td>16 (0.101)</td>
</tr>
<tr>
<td><strong>Site, Outcome(%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0 (NaN)</td>
<td>0 (NaN)</td>
<td>0 (NaN)</td>
<td>0 (0.000)</td>
<td>1 (0.075)</td>
<td>2 (0.103)</td>
</tr>
<tr>
<td>4</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
</tr>
<tr>
<td>15</td>
<td>5 (0.184)</td>
<td>8 (0.131)</td>
<td>11 (0.113)</td>
<td>15 (0.106)</td>
<td>17 (0.091)</td>
<td>23 (0.091)</td>
</tr>
<tr>
<td>16</td>
<td>0 (0.000)</td>
<td>2 (0.240)</td>
<td>3 (0.166)</td>
<td>3 (0.099)</td>
<td>3 (0.075)</td>
<td>4 (0.086)</td>
</tr>
</tbody>
</table>

*Abbreviations: Outcome(%) = Number(Risk %) of outcome within look and covariate stratum.*
Outcome over time amongst MMRV recipients (cont)

Table A.2: (Continued) Demographics by Outcome across Analysis Times among MMRV

<table>
<thead>
<tr>
<th></th>
<th>Look 7</th>
<th>Look 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total, Outcome(%)</td>
<td>36 (0.089)</td>
<td>45 (0.093)</td>
</tr>
<tr>
<td>Age, Outcome(%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11m-12m</td>
<td>15 (0.076)</td>
<td>20 (0.084)</td>
</tr>
<tr>
<td>13m-14m</td>
<td>8 (0.091)</td>
<td>9 (0.084)</td>
</tr>
<tr>
<td>15m-16m</td>
<td>5 (0.068)</td>
<td>7 (0.085)</td>
</tr>
<tr>
<td>17m-19m</td>
<td>6 (0.194)</td>
<td>6 (0.163)</td>
</tr>
<tr>
<td>20m-23m</td>
<td>2 (0.139)</td>
<td>3 (0.169)</td>
</tr>
<tr>
<td>Sex, Outcome(%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16 (0.078)</td>
<td>21 (0.086)</td>
</tr>
<tr>
<td>Female</td>
<td>20 (0.101)</td>
<td>24 (0.101)</td>
</tr>
<tr>
<td>Site, Outcome(%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2 (0.092)</td>
<td>2 (0.088)</td>
</tr>
<tr>
<td>4</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
</tr>
<tr>
<td>15</td>
<td>30 (0.092)</td>
<td>35 (0.089)</td>
</tr>
<tr>
<td>16</td>
<td>4 (0.075)</td>
<td>8 (0.128)</td>
</tr>
</tbody>
</table>

*Abbreviations: Outcome(%) = Number(Risk %) of outcome within look and covariate stratum.
MMRV Demographics across Analysis Times

<table>
<thead>
<tr>
<th>Table A.4: Demographics across Analysis Times among MMRV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Look 1</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Total, N(Row%)</td>
</tr>
<tr>
<td>Age, N(Col%)</td>
</tr>
<tr>
<td>11m-12m</td>
</tr>
<tr>
<td>13m-14m</td>
</tr>
<tr>
<td>15m-16m</td>
</tr>
<tr>
<td>17m-19m</td>
</tr>
<tr>
<td>20m-23m</td>
</tr>
<tr>
<td>Sex, N(Col%)</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Site, N(Col%)</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
</tbody>
</table>
Table A.4: (Continued) Demographics across Analysis Times among MMRV

<table>
<thead>
<tr>
<th></th>
<th>Look 7</th>
<th>Look 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total, N(Row%)</td>
<td>40326(83.6)</td>
<td>48233(100.0)</td>
</tr>
<tr>
<td>Age, N(Column%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11m-12m</td>
<td>19677 (48.8)</td>
<td>23898 (49.5)</td>
</tr>
<tr>
<td>13m-14m</td>
<td>8785 (21.8)</td>
<td>10651 (22.1)</td>
</tr>
<tr>
<td>15m-16m</td>
<td>7341 (18.2)</td>
<td>8226 (17.1)</td>
</tr>
<tr>
<td>17m-19m</td>
<td>3086 (7.7)</td>
<td>3687 (7.6)</td>
</tr>
<tr>
<td>20m-23m</td>
<td>1437 (3.6)</td>
<td>1771 (3.7)</td>
</tr>
<tr>
<td>Sex, N(Column%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>20541 (50.9)</td>
<td>24548 (50.9)</td>
</tr>
<tr>
<td>Female</td>
<td>19785 (49.1)</td>
<td>23685 (49.1)</td>
</tr>
<tr>
<td>Site, N(Column%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2178 (5.4)</td>
<td>2262 (4.7)</td>
</tr>
<tr>
<td>4</td>
<td>394 (1.0)</td>
<td>402 (0.8)</td>
</tr>
<tr>
<td>15</td>
<td>32443 (80.5)</td>
<td>39334 (81.5)</td>
</tr>
<tr>
<td>16</td>
<td>5311 (13.2)</td>
<td>6235 (12.9)</td>
</tr>
</tbody>
</table>
Figure A.1: Uptake of MMR+V and MMRV over Time for Site 2
Figure A.2: Uptake of MMR+V and MMRV over Time for Site 4
Figure A.3: Uptake of MMR+V and MMRV over Time for Site 15
Uptake by Site

Figure A.4: Uptake of MMR+V and MMRV over Time for Site 16
Propensity Scores

- Propensity Score Definition
  - $P(\text{Exposed}|\text{Confounders}) = P(X|Z)$

- Why are they relevant to safety surveillance?
  - Allows for the adjustment of a large number of confounders even in the rare event setting given:
    - good proportion of exposed and unexposed
    - De-identifies data since numerous confounders combinations can give the same propensity score value

- How can they be used?
Confounding Control Using Propensity Scores

- **Design-based Approaches**
  - Exposure Matching
    - Match each exposed person to a set of controls with similar propensity scores
  - Stratification
    - Stratify the population with propensity score strata

- **Analysis-based Approaches**
  - Regression
    - Adjust directly for propensity scores
  - Inverse Probability of Treatment Weighting
    - Re-weights population based on propensity scores to average out the effect of confounding
**What is IPTW?**

What is an inverse probability of treatment weight?

For those exposed
- Weight = \(1 / P(\text{Exposed} | \text{Confounders}) = 1 / P(X | Z)\)

For those unexposed
- Weight = \(1 / P(\text{Unexposed} | \text{Confounders}) = 1 / (1 - P(X | Z))\)

<table>
<thead>
<tr>
<th></th>
<th>Unlikely to be exposed</th>
<th>Equally likely to be exposed or unexposed</th>
<th>Likely to be exposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>↑ weight</td>
<td>Neutral-weight</td>
<td>↓ weight</td>
</tr>
<tr>
<td>Unexposed</td>
<td>↓ weight</td>
<td>Neutral-weight</td>
<td>↑ weight</td>
</tr>
</tbody>
</table>

Use these weights to estimate marginal RR, OR, Risk Difference
IPTW in Distributed Data Settings: Currently Available

IPTW

- Use site-specific propensity scores to create IPTW weights
- Estimates: marginal OR, RR or Risk Difference
- Issues:
  - Large Weights: Observations become too informative and variance inflates (Standardize weights, Trimming or Restriction)
  - Incorporate site-specific propensity score models: use bootstrapping or permutation to handle differential variability of propensity score across site, but not very computationally feasible.
  - Can we do this better?
What is a pragmatic trial??

- **Just an effectiveness Study??** (Patsopoulos, N.A. 2011)
  - Research Question:
    - Does an intervention actually work in a real life setting??
    - Generalizes to a larger population (no/limited inclusion or exclusion criteria)
    - No plans to improve or alter compliance for the experimental or the comparative treatment

- **Studies designed to provide information that can be directly adopted by healthcare providers (PCORI)**
  - Simpler than traditional RCTs
  - Conducted in routine clinical care settings
  - Relatively large

- **More cost-effective then standard RCTs**
  - Use EHR to collect outcomes and/or distribute intervention
  - More feasible/cheaper study designs (e.g. Cluster randomized)
Pragmatic Trials Initiatives

- **PCORI**
  - PCORnet
    - Large network of distributed healthcare data to conduct both pragmatic trials and observational studies
  - Spring 2014 Funding Announcement: Pragmatic Clinical Studies and Large Simple Trials to Evaluate Patient-Centered Outcomes

- **NIH Collaboratory**
  - “Rethinking Clinical Trials”
  - Health Care Systems Research Collaboratory
  - First round funded 7 Pragmatic Trials with 5 going to UH3 stage so far
  - Second round being announced soon.
Use of EHR Data in Pragmatic Trials for…

- **Study Population, Design, and Recruitment**
  - Use EHR to define eligible study population
  - Design/Randomization: Obtain information on clinic population to stratify cluster randomization on important variables (size, baseline outcomes, age of population, …)
  - Recruit participants through EHR systems

- **Implementing the intervention**

- **Data Collection**
  - Baseline Data
  - Confounders (patient, physician, or clinic level)
  - Primary and Secondary Outcomes (patient reported outcomes?)
Common Outcome Assumptions/Issues

- Health Outcomes: If someone is enrolled in the healthcare plan if there is no outcome coded then they don’t have it
  - Healthcare utilization varies for different participants is this assumption valid
  - Likely ok for serious health outcomes

- Patient Reported or other Process Outcomes: Only obtain measures on those who attend a health care visit (e.g. Depression measures, blood pressure, weight)
  - Bias due to having a measure or not
  - Bias when the measure occurred
  - Bias how many measures occurred over a follow-up window
Pragmatic Trials are a cross between RCTs and Full observational studies.

Do not assume randomization fixes all issues even individually randomized designs.

Consider potential bias from outcome ascertainment, design choices (e.g. cluster), consent, …

Apply statistical methods similar to full observational studies.

More considerations in conducting sensitivity analyses to assess assumptions.
Pragmatic Trial Conclusions

- Pragmatic Trials are becoming increasingly common
- Using EHR data has advantages, but causes statistical complications
- Large networks have formed or are being formed to conduct even larger multi-site pragmatic clinical trials
  - Distributed Data systems are likely to be used to protect patient privacy
  - Statistical methods that can be applied in distributed data networks will need to be developed to conduct such studies correctly.
Permutation Approach for Rare Events

- Under Ho: Adjusted Risk Difference = 0
- Outcome|Confounders independent of exposure
- Permutation Approach:
  - Fix outcomes and confounders and permute exposures within site
  - On permuted datasets calculate $\Delta_s^{(p)}$, $V(\Delta_s^{(p)})$
- Computational Advantages
  - For permuted datasets $P(X|Z)=P(X)$
  - Don’t need to refit propensity models on permuted data, but just assume propensities are fixed and calculate correct variance
- For analyses sites would also send these permuted summary data to a central location