Data Needs for Signal Refinement:
Experience from the HMO Research Network and Vaccine Safety Datalink Project

Tracy Lieu, MD, MPH

Center for Child Health Care Studies
Department of Population Medicine
Harvard Pilgrim Health Care Institute and
Harvard Medical School

September 21, 2010

The HMO Research Network

- 16 HMOs with formal research capabilities
- Current combined cohort is ~ 14.5 million persons, ~ 4.5% of U.S. population
- Mini-Sentinel includes 11 HMORN sites: Kaiser Permanentes, plus 5 others

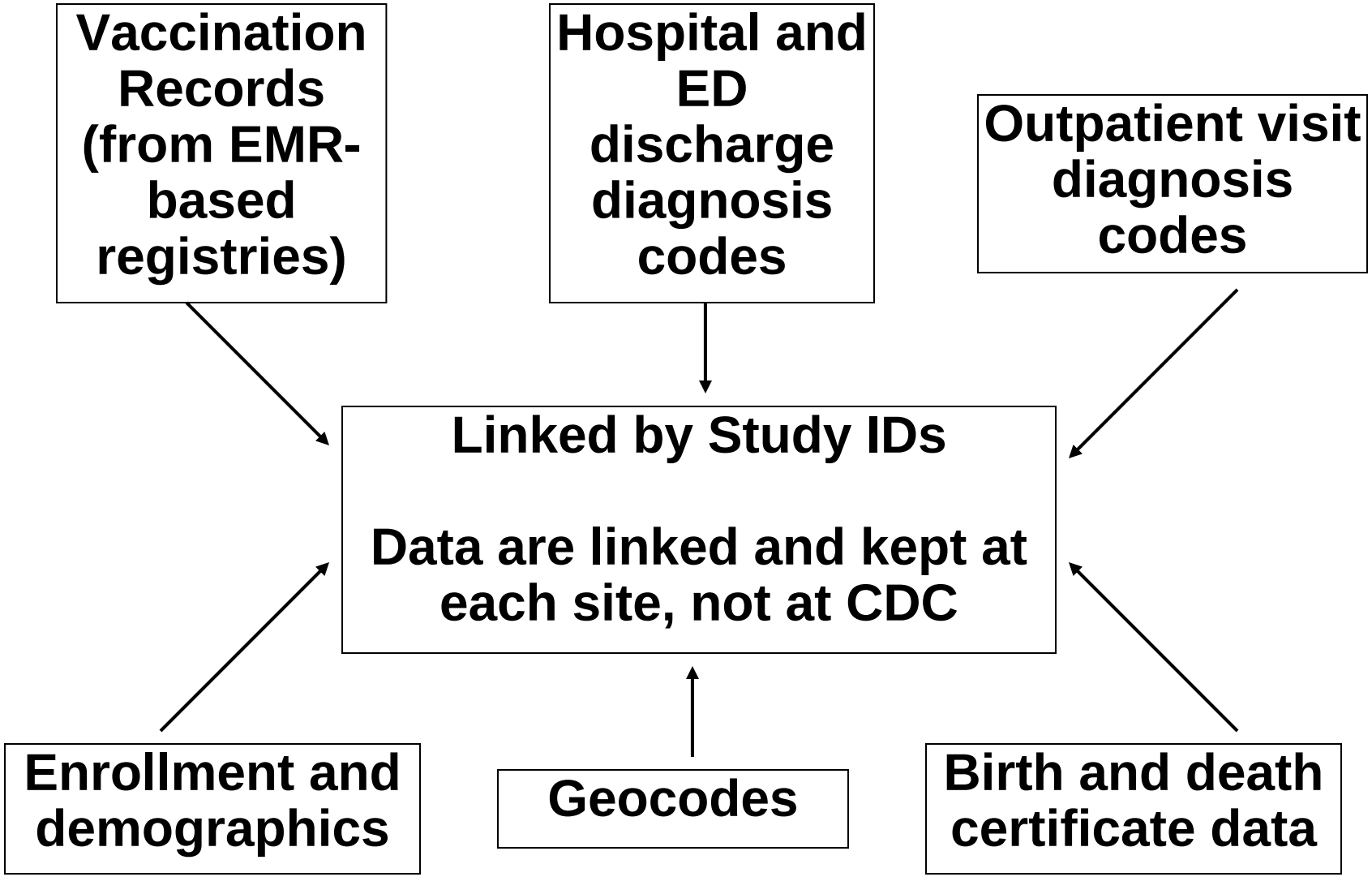
Examples of Data Needs for Signal Refinement: The Vaccine Safety Datalink Project

- 8 HMORN sites
- Sponsored and coordinated by CDC since 1991
- Current cohort is ~ 9 million persons

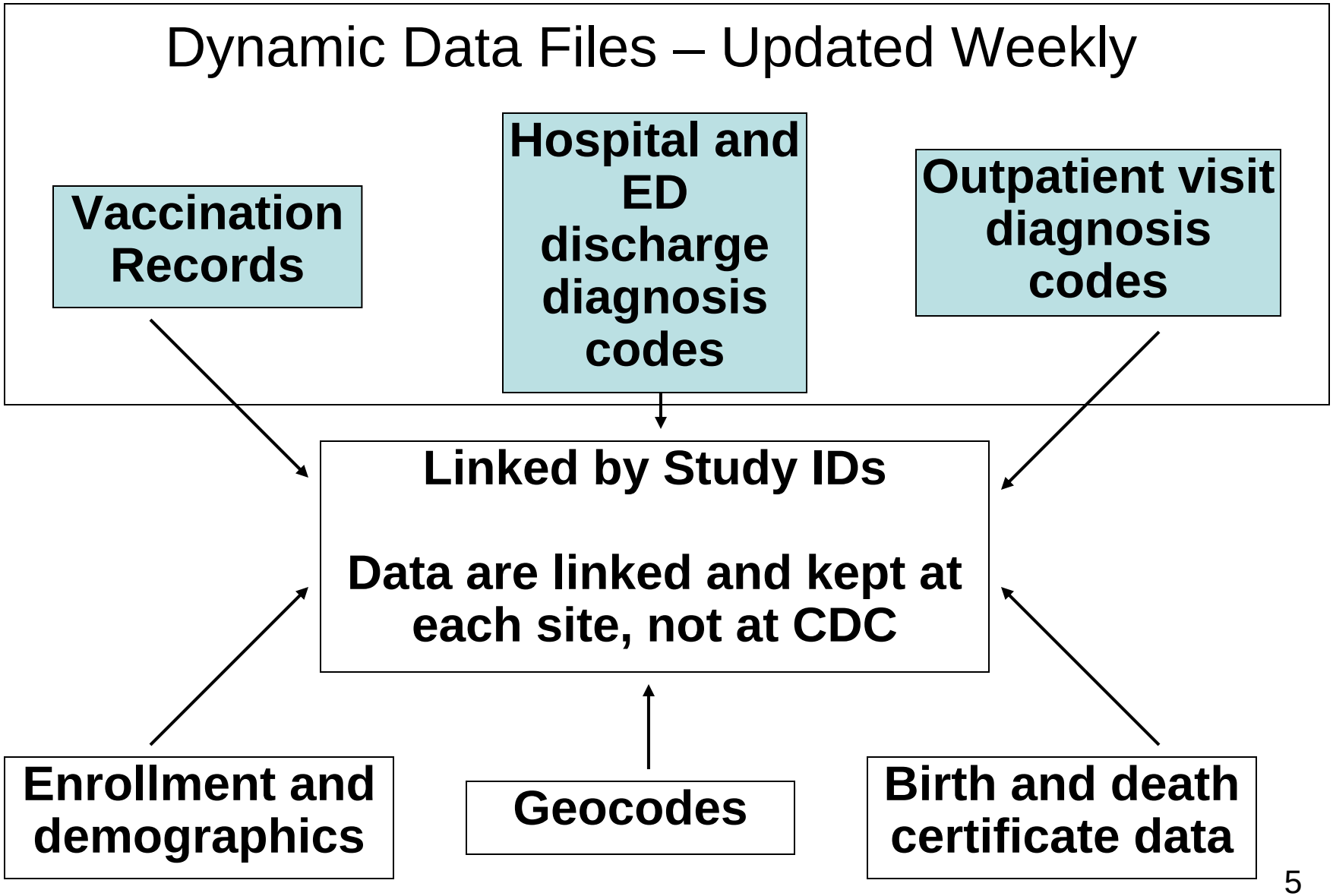
  ~ 3% of U.S. population

  – Birth cohort = 95,000
- All VSD sites participate in Mini-Sentinel

VSD Data Files
-- Distributed Data Model



VSD's Rapid Cycle Analysis



Rapid Cycle Analysis

- Data are updated on all vaccines and all outcomes every week
- Monitor pre-specified outcomes
  - Selected based on literature and reports
- Evaluate the number of outcomes in vaccinated persons
- Compare this to the expected number of outcomes based on a comparison group

Lieu Med Care 2007

Signals Occurring in VSD's RCA

- RCA began in 2005-06
- Vaccines monitored:
  - Menactra, Tdap, MMRV, RotaTeq, HPV, influenza
- In the first 5 vaccines, we monitored 30 vaccine-outcome pairs
- 10 signals occurred; one represented a true association (MMRV and febrile seizures)

Yih Pediatrics in press

Signal Refinement – Recommended Steps

1. Check the Data
2. Check the Analytic Programs
3. Descriptive Statistics
4. Time from Vaccine to Adverse Event
5. Adjust for Additional Confounders
6. Use Other Comparison Groups
7. Chart Review

Modified from Kulldorff, Yih, presentations for VSD

Signal Refinement – Recommended Steps

8. Compare with Other Existing Data
9. Studies with New Data
10. Compare Similar Outcomes

Modified from Kulldorff, Yih, presentations for VSD

Check the Data

- Observed counts
- Expected counts
- Example: MMRV and allergic reactions
  - Initial background rates were from 2000-06
  - After signal, somewhat higher rates noted in later years
  - Rates for 2005-06 used instead
  - Signal did not persist

Check the Data

- Example: HPV and allergic reactions
  - Review of the historical data showed a limited number of cases
  - Analysis was considered biased (Type I error probability >0.05)
  - Newly developed refinement, conditional maxSPRT, was implemented to adjust for this uncertainty
  - Signal did not persist

Li & Kulldorff Stat Med 2010

Check the Data

- Compare incidence rates with the literature
- Example: MMRV and thrombocytopenia
  - Original background rate was from all person-time, due to sparse numbers after MMR
  - France et al. (2008) provided post-MMR incidence for 1-year-olds (twice the rate initially used)
  - New post-MMR rate substituted
  - Signal did not persist

Descriptive Statistics

- By age, gender, site
- Look at secular and seasonal trends
- Example: MMRV and Ataxia
  - One site had 3 times as many cases as expected
  - Chart review there found 20/21 cases were miscoded
  - Miscoding not correctable so site was excluded
  - Signal did not persist

Time from Vaccine to Adverse Event

- Descriptive histograms
- Temporal scan statistic
- Look at different risk windows
- Example: MMRV and seizures
  - Temporal scan showed clustering in days 7-10 after MMRV
  - Logistic regression showed increased risk of seizure in days 7-10 after MMRV, compared with MMR + V
  - Chart review and reanalysis confirmed a >2-fold risk of febrile seizures

Adjust for Additional Counfounders

. . . in non-sequential analyses

- More accurate age adjustments
- Seasonal and secular trends
- Concomitant vaccines
- Example: RotaTeq and GI Bleeding #1
  - Examination of age distribution of cases suggested GI bleeding was age-dependent
  - Rates were adjusted by age
  - Signal did not persist

Use Other Comparison Groups

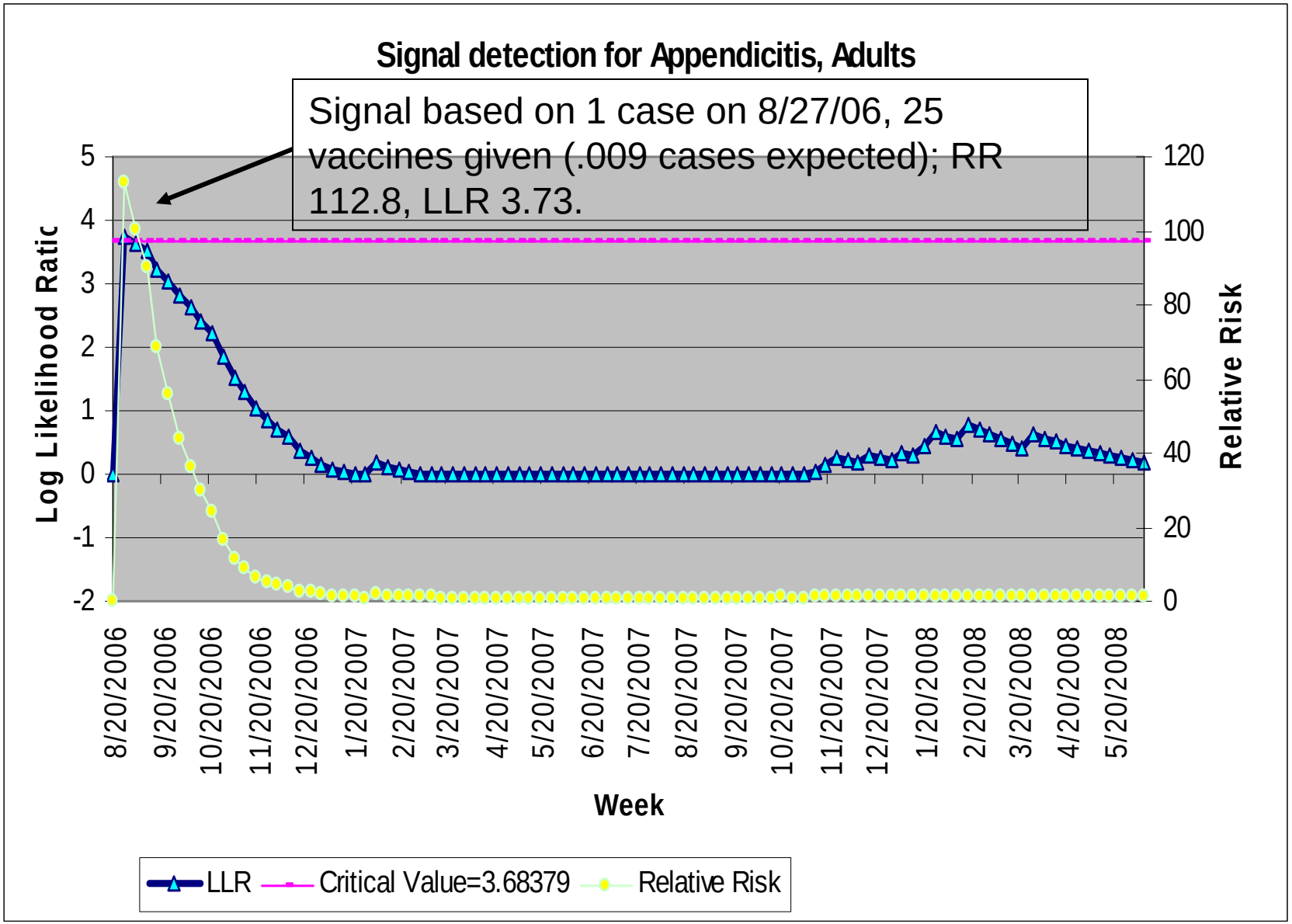
- Different historical time periods
- Other vaccines
- Concurrent well-care visits
- Self-control case series
- Example: RotaTeq and GI Bleeding #2
  - New signal occurred
  - Logistic regression using visits for other vaccines in the concurrent period as the comparison found no difference in risk
  - Secular trend deemed possible source of error in background rates

Chart Review

- Review a random sample of cases
- Or review all exposed and unexposed cases
- Re-do analysis with chart confirmed cases
- Example: Menactra and Guillain-Barre syndrome
  - 5^{th} case produced a signal
  - 0.97 cases expected under the null
  - 0 cases confirmed on chart review

Studies with New Data

- Continue data collection in current system
- Case-control or self-control studies
- Example: HPV and appendicitis
  - 1 case in Week 2 of data produced a retrospective, transient signal
  - At time of the actual look, the relative risk and test statistic (log likelihood ratio) had decreased to null values
  - “Old” signal ascribed to chance



Data from 8/20/06 to 6/8/08: 19 cases, 108,184 vaccines given (16.4 cases expected); RR=1.2, LLR 0.1995.

Summary: Reasons for False Signals (N=9)

- **Error in estimated background rates**
  - **Miscoding of data (2)**
  - **Inappropriate group (1)**
  - **Low #s of cases in historical data (1)**
- **Confounding**
  - **Secular trends (1)**
  - **Age (1)**
- **Chance (3)**

Conclusions

- **In signal evaluation, data quality should be a key focus**
- **Refined analyses of existing data often lead to an answer**
  - **The additional data needed are sometimes in the distributed data files**
  - **May not have been foreseen during protocol development**
- **Chance is an important factor**