Using Freshest Feasible Data for Medical Product Safety Surveillance in Mini-Sentinel: Potential and Challenges

W. Katherine Yih, PhD, MPH
Harvard Pilgrim Health Care Institute and Harvard Medical School
January 31, 2013
Inpatient claims data lag, 3 data partners

Data ≥ 90% complete by 6 mo. after care date
Mini-Sentinel data are relatively complete

- Data updated on quarterly basis
- Typical example of timing:

In latest batch of data for M-S:  
First care date  Last care date  Data available

_↓_________↓____________________↓_____

- The most recent data typically 6-9 months old
Advantage of mature (less fresh) data

PRO: data more complete and settled

In latest batch of data for M-S:
First care date     Last care date

Data available
July
Pros and cons of mature (less fresh) data

- **PRO**: data more complete and settled

  In latest batch of data for M-S:
  - First care date: Oct.
  - Last care date: Dec.
  - Data available: July

- **CON**: signal detection delayed
Pros and cons of mature (less fresh) data

- **PRO:** data more complete and settled

  In latest batch of data for M-S:
  - First care date
  - Last care date

  ![Diagram showing vaccination timing]

- **CON:** signal detection delayed
  Especially problematic for influenza vaccine safety monitoring

  ![Diagram showing vaccination timing]
Challenges of influenza vaccine safety monitoring

Influenza vaccination period relatively short, so data must be available soon after exposure to find safety problems in time to make a difference


Typical influenza vaccination timing
Challenges of influenza vaccine safety monitoring

Influenza vaccination period relatively short, so data must be available soon after exposure to find safety problems in time to make a difference

1. Need fresher and frequently updated data
2. Need to adjust for incomplete data
1. Getting fresher and frequent data

Freshest feasible data source is refreshed monthly

- Available toward end of following calendar month (data through Sept. available late Oct., etc.)
- More timely than M-S Distributed Dataset

1. Getting fresher and frequent data

Freshest feasible data source is refreshed monthly

- Available toward end of following calendar month (data through Sept. available late Oct., etc.)
- More timely than M-S Distributed Dataset

1. Getting fresher and frequent data

Freshest feasible data source is refreshed monthly

- Available toward end of following calendar month (data through Sept. available late Oct., etc.)
- More timely than M-S Distributed Dataset

1. Getting fresher and frequent data

Freshest feasible data source is refreshed monthly

- Available toward end of following calendar month (data through Sept. available late Oct., etc.)
- More timely than M-S Distributed Dataset

Files to be created for influenza vaccine safety monitoring

**SDFs (Sequential Data Files)**
- Patient-level data, kept by data partners
- Population = persons with medical claim on or after 9/1/2012

**SCFs (Sequential Case Files)**
- Patient-level data, kept by data partners
- Population = persons per current SDFs with health outcome of interest following influenza vaccination

**SAFs (Sequential Analysis Files)**
- Aggregate data, sent to M-S Operations Center for analysis
- Vaccination population: vaccination per current SDFs
- Cases population: cases per all SCF versions
Expected timing of data refreshes and analyses

- Monthly but unsynchronized data refreshes by data partners
- Biweekly analyses by Operations Center (in weeks in red)

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP1</td>
<td>SDF</td>
<td>SAF</td>
<td></td>
<td>SDF</td>
<td>SAF</td>
<td></td>
<td>SDF</td>
<td>SAF</td>
<td></td>
</tr>
<tr>
<td>DP2</td>
<td>SDF</td>
<td>SAF...</td>
<td></td>
<td>SDF</td>
<td>SAF...</td>
<td></td>
<td>SDF</td>
<td>SAF...</td>
<td></td>
</tr>
<tr>
<td>DP3</td>
<td>SDF</td>
<td>SAF</td>
<td></td>
<td>SDF</td>
<td>SAF</td>
<td></td>
<td>SDF</td>
<td>SAF</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
2. Adjusting for incomplete data

Two kinds of “incompleteness”

A. Lag in data availability →

B. Post-vaccination follow-up interval not fully elapsed

To avoid bias, both must be taken into account.
Near real-time vaccine safety surveillance with partially accrued data†

Sharon K. Greene¹*, Martin Kulldorff¹, Ruihua Yin¹, W. Katherine Yih¹, Tracy A. Lieu¹, Eric S. Weintraub² and Grace M. Lee¹,³

¹Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA
²Immunization Safety Office, Centers for Disease Control and Prevention, Atlanta, GA, USA
³Division of Infectious Diseases and Department of Laboratory Medicine, Children’s Hospital Boston, Boston, MA, USA
Cumulative inactivated H1N1 vaccine doses

Week of Analysis

Cumulative vaccine doses

Nov. 18
Nov. 25
Dec. 2
Dec. 9
Dec. 16
Dec. 23
Dec. 30
Jan. 6
Jan. 13
Jan. 20
Jan. 27
Feb. 3
Feb. 10
Feb. 17
Feb. 24
Mar. 3
Mar. 10
Mar. 17
Mar. 24
Mar. 31
Apr. 14
Cumulative inactivated H1N1 vaccine doses vs. Log-Likelihood Ratio

Critical Value of Log-Likelihood Ratio

Cumulative vaccine doses

Data lag adjustment only

Week of Analysis


0 200,000 400,000 600,000 800,000 1,000,000 1,200,000 1,400,000

0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0

Log-likelihood ratio
Critical Value of Log-Likelihood Ratio

Cumulative vaccine doses

Partial interval adjustment only

Cumulative inactivated H1N1 vaccine doses
Cumulative inactivated H1N1 vaccine doses

Critical Value of Log-Likelihood Ratio

Cumulative vaccine doses

Partial interval and data lag adjustments

Week of Analysis

Log-likelihood ratio

Cumulative inactivated H1N1 vaccine doses

Critical Value of Log-Likelihood Ratio

Cumulative vaccine doses

Partial interval and data lag adjustments

Week of Analysis

Log-likelihood ratio
Conclusion

- **PROS of using fresher data**
  - Gain in timeliness ~5-8 mo.
  - Necessary for influenza vaccine safety monitoring

- **CONS of using fresher data**
  - Some loss of accuracy despite adjustments for data incompleteness and flux
  - Takes extra effort to produce these data—more frequent refreshes, different source files, special file structures
  - Each product needs a separate extract

- We *can* use fresher data, but probably not worthwhile to do so on routine basis
What constitutes a comprehensive safety surveillance system?

• Semi-automated routine surveillance, applying general tools with minor adaptations to address the specific product

But also...

• Ability to bring specialized expertise to bear on specific issue(s) that may arise in product lifecycle