Data Mining in Pharmacovigilence

Aimin Feng, David Madigan, and Ivan Zorych

dmadigan@rutgers.edu

http://stat.rutgers.edu/~madigan
Safety in Lifecycle of a Drug/Biologic product

- **Pre-clinical**
- **Phase 1**
- **Phase 2**
- **Phase 3**

Safety → Safety → Safety → Safety → Safety

Dose-Ranging → Efficacy → Approval → Safety Concern → Post-Marketing Safety Monitoring
Databases of Spontaneous ADRs

- FDA Adverse Event Reporting System (AERS)
  - Online 1997 – replace the SRS
  - Over 250,000 ADRs reports annually
  - 15,000 drugs - 16,000 ADRs

- CDC/FDA Vaccine Adverse Events (VAERS)
  - Initiated in 1990
  - 12,000 reports per year
  - 50 vaccines and 700 adverse events

- Other SRS
  - WHO - international pharmacovigilance program
The FDA Safety Information and Adverse Event Reporting Program

**A. PATIENT INFORMATION**

<table>
<thead>
<tr>
<th>Patient Identifier</th>
<th>Age at Time of Event</th>
<th>Sex</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>(if applicable)</td>
<td>(mm/dd/yyyy)</td>
<td></td>
<td>(mm/dd/yyyy)</td>
</tr>
</tbody>
</table>

**B. ADVERSE EVENT, PRODUCT PROBLEM OR ERROR**

Check all that apply:
- Adverse Event
- Product Problem: e.g., defective/bad batch
- Wrong Product
- Wrong Patient/Other

3. Outcomes attributed to adverse event:
- Death
- Disability or Permanent Damage
- Life Threatening
- Other

4. Diagnosis or Reason for Use (Indication)

5. Description of Event, Problem or Product Use Error

**D. SUSPECT PRODUCT(S)**

1. Name, Strength, Manufacturer (from product label):

<table>
<thead>
<tr>
<th>Name</th>
<th>Strength</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Drug or Amount

<table>
<thead>
<tr>
<th>Drug</th>
<th>Amount</th>
<th>Frequency</th>
<th>Route</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Dates of Use (Start and Stop Dates) (if known, plus duration)

<table>
<thead>
<tr>
<th>Start Date</th>
<th>Stop Date</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>(mm/dd/yyyy)</td>
<td>(mm/dd/yyyy)</td>
<td></td>
</tr>
</tbody>
</table>

4. Event Related to Use (Same as above)

5. Occurrence After Use (Stop or Dose Reduced?)

<table>
<thead>
<tr>
<th>Start Date</th>
<th>Stop Date</th>
<th>Did it Appear?</th>
</tr>
</thead>
<tbody>
<tr>
<td>(mm/dd/yyyy)</td>
<td>(mm/dd/yyyy)</td>
<td>Yes/No</td>
</tr>
</tbody>
</table>

**E. SUSPECT MEDICAL DEVICE**

1. Brand Name

2. Common Device Name

3. Manufacturer Name, City, and State

4. Model #

5. Operator of Device

6. Catalog #

7. Expiration Date (mm/dd/yyyy)

8. Serial #

9. Other #

10. If Implantable, Give Date (mm/dd/yyyy)

11. If Expired, Give Date (mm/dd/yyyy)

12. Is this a Single-use Device that was Reprocessed and Reused on a Patient?

13. Yes/No

14. If Yes to Item No. 13, Enter Name and Address of Reprocessor

**F. OTHER (CONCOMITANT) MEDICAL PRODUCTS**

<table>
<thead>
<tr>
<th>Product Name and Therapy Details (Excludes treatment of event)</th>
</tr>
</thead>
</table>

**G. REPORTER (See Confidentiality Section on Back)**

1. Name and Address

2. Phone #

3. E-mail

4. Also Reported To:

**C. PRODUCT AVAILABILITY**

<table>
<thead>
<tr>
<th>Product Available for Evaluation? (Do not send product to FDA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
</tr>
</tbody>
</table>

If you do not want your identity disclosed to the manufacturer, place an "X" in this box:

**Additional Information**

[Additional text not visible]
Weakness of SRS Data

- Passive surveillance
  - Underreporting
- Lack of accurate “denominator”, only “numerator”
  - “Numerator”: No. of reports of suspected reaction
  - “Denominator”: No. of doses of administered drug
- No certainty that a reported reaction was causal
- Missing, inaccurate or duplicated data
Existing Methods

- Multi-item Gamma Poisson Shrinker (MGPS)
  - US Food and Drug Administration (FDA)

- Bayesian Confidence Propagation Neural Network
  - WHO Uppsala Monitoring Centre (UMC)

- Proportional Reporting Ratio (PRR and aPRR)
  - UK Medicines Control Agency (MCA)

- Reporting Odds Ratios and Incidence Rate Ratios
  - Other national spontaneous reporting centers and drug safety research units
Existing Methods (Cont’d)

- Focus on 2X2 contingency table projections

<table>
<thead>
<tr>
<th></th>
<th>AE $j = \text{Yes}$</th>
<th>AE $j = \text{No}$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug $i = \text{Yes}$</td>
<td>a=20</td>
<td>b=100</td>
<td>120</td>
</tr>
<tr>
<td>Drug $i = \text{No}$</td>
<td>c=100</td>
<td>d=980</td>
<td>1080</td>
</tr>
<tr>
<td>Total</td>
<td>120</td>
<td>1080</td>
<td>1200</td>
</tr>
</tbody>
</table>

- 15,000 drugs * 16,000 AEs = 240 million tables
- Most $N_{ij} = 0$, even though $N_{..}$ very large
The Different Measures

<table>
<thead>
<tr>
<th>Measure of Association</th>
<th>Formula</th>
<th>Probabilistic Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR Relative Risk*</td>
<td>[\frac{a * (a + b + c + d)}{(a + c) * (a + b)}]</td>
<td>[\frac{\Pr(ae \mid drug)}{\Pr(ae)}]</td>
</tr>
<tr>
<td>PRR Proportional Reporting Ratio</td>
<td>[\frac{a}{(a + b)} \quad \frac{c}{(c + d)}]</td>
<td>[\frac{\Pr(ae \mid drug)}{\Pr(ae \mid \neg drug)}]</td>
</tr>
<tr>
<td>ROR Reporting Odds Ratio</td>
<td>[\frac{a}{c} \quad \frac{b}{d}]</td>
<td>[\frac{\Pr(ae \mid drug)/\Pr(\neg ae \mid drug)}{\Pr(ae \mid \neg drug)/\Pr(\neg ae \mid drug)}]</td>
</tr>
<tr>
<td>Information Component</td>
<td>[\text{Log}_2 \frac{a * (a + b + c + d)}{(a + c) * (a + b)}]</td>
<td>[\text{log}_2 \frac{\Pr(ae \mid drug)}{\Pr(ae)}]</td>
</tr>
</tbody>
</table>
Relative Reporting Ratio

\( (RR_{ij} = N_{ij}/E_{ij}) \)

- Advantages
  - Simple
  - Easy to interpret

- Disadvantages
  - Extreme sampling variability when baseline and observed frequencies are small
    \((N=1, E=0.01 \text{ v.s. } N=100, E=1)\)
  - GPS provides a shrinkage estimate of RR that addresses this concern.
Same Relative Reporting Ratio!

<table>
<thead>
<tr>
<th>Drug</th>
<th>Not Drug</th>
<th>AE</th>
<th>Not AE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>a=1</td>
<td>b=5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c=5</td>
<td>d=49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a=20</td>
<td>b=100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c=100</td>
<td>d=980</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a=200</td>
<td>b=1000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c=1000</td>
<td>d=9800</td>
</tr>
</tbody>
</table>

Chi-square = 0.33
Chi-square = 6.58
Chi-square = 65.8
GPS/MGPS

- Denote by $\rho_{ij}$ the true RR for Drug $i$ and AE $j$
- Assumes the $\rho_{ij}$’s arise from a particular 5-parameter distribution
- Use empirical Bayes to use the data to estimate these five parameters.
GPS-EBGM

- Define $\lambda_{ij} = \mu_{ij} / E_{ij}$, where
  - $N_{ij} \sim \text{Poisson}(\mu_{ij})$
  - $\lambda_{ij} | \lambda \sim p \cdot g(\lambda; \alpha_1, \beta_1) + (1-p) \cdot g(\lambda; \alpha_2, \beta_2)$
    a mixture of two Gamma Distributions

- EBGM = Geometric mean of Post-Dist. of $\lambda_{ij}$
  - Estimates of $\mu_{ij} / E_{ij}$
  - “Shrinks” $N_{ij} / E_{ij} \to 1$
  - Smaller variances than $N_{ij} / E_{ij}$