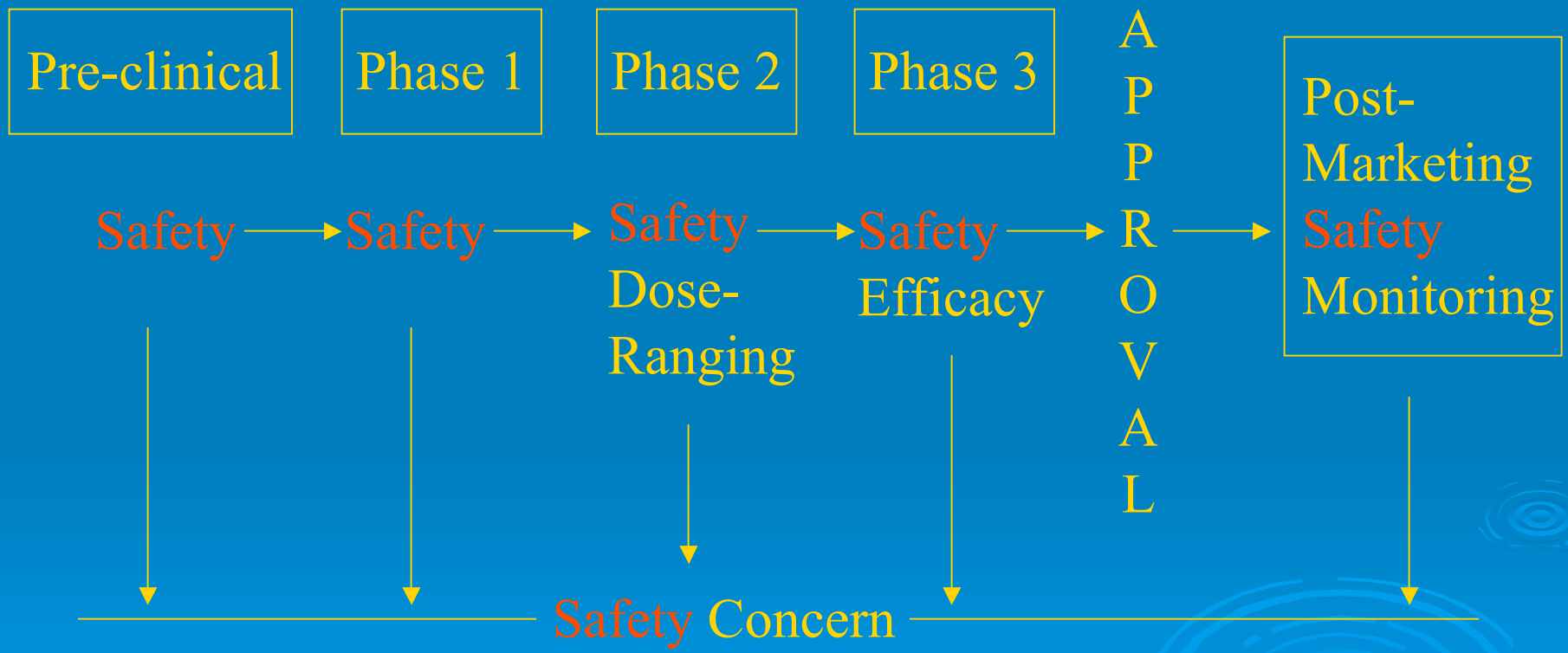
Data Mining in Pharmacovigilence

Aimin Feng, David Madigan, and Ivan Zorych

dmadigan@rutgers.edu

http://stat.rutgers.edu/~madigan

Safety in Lifecycle of a Drug/Biologic product



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

U.S. Department of Health and Human Services

MEDWATCH

The FDA Safety Information and Adverse Event Reporting Program

For VOLUNTARY reporting of adverse events, product problems and product use errors

Page ____ of ____

Form Approved: OMB No. 0910-0291, Expires: 10/31/08
See OMB statement on reverse.

FDA USE ONLY
Triage unit
sequence #

PLEASE TYPE OR USE BLACK INK

A. PATIENT INFORMATION

1. **Patient Identifier** — In confidence
2. **Age at Time of Event, or Date of Birth:**
3. **Sex** — ☐ Female ☐ Male
4. **Weight** — _____ lb or _____ 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)* ☐ Disability or Permanent Damage
   ☐ Life-threatening ☐ 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**
6. **Relevant Tests/Laboratory Data, Including Dates**
7. **Other Relevant History, Including Preexisting Medical Conditions** *(e.g., allergies, race, pregnancy, smoking and alcohol use, liver/kidney problems, etc.)*

C. PRODUCT AVAILABILITY

Product Available for Evaluation? *(Do not send product to FDA)*

☐ Yes ☐ No ☐ Returned to Manufacturer on: _____ *(mm/dd/yyyy)*

D. SUSPECT PRODUCT(S)

1. **Name, Strength, Manufacturer** *(from product label)*
   #1
   #2
2. **Dose or Amount** / **Frequency** / **Route**
   #1
   #2
3. **Dates of Use** *(If unknown, give duration) from/to (or best estimate)*
   #1
   #2
4. **Diagnosis or Reason for Use** *(Indication)*
   #1
   #2
5. **Event Abated After Use Stopped or Dose Reduced?**
   #1 ☐ Yes ☐ No ☐ Doesn't Apply
   #2 ☐ Yes ☐ No ☐ Doesn't Apply
6. **Lot #**
   #1
   #2
7. **Expiration Date**
   #1
   #2
8. **Event Reappeared After Reintroduction?**
   #1 ☐ Yes ☐ No ☐ Doesn't Apply
   #2 ☐ Yes ☐ No ☐ Doesn't Apply
9. **NDC # or Unique ID**

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**

F. OTHER (CONCOMITANT) MEDICAL PRODUCTS

Product names and therapy dates *(exclude treatment of event)*

G. REPORTER *(See confidentiality section on back)*

1. **Name and Address**
   Phone # / **E-mail**
2. **Health Professional?** ☐ Yes ☐ No
3. **Occupation**
4. **Also Reported to:** ☐ Manufacturer ☐ User Facility ☐ Distributor/Importer
5. **If you do NOT want your identity disclosed to the manufacturer, place an "X" in this box:** ☐

FORM FDA 3500 (10/05) Submission of a report does not constitute an admission that medical personnel or the product caused or contributed to the event.

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

	AE j = Yes	*AE j = No*	*Total*
Drug *i* = Yes	*a*=20	*b*=100	120
Drug *i* = No	*c*=100	*d*=980	1080
Total	*120*	*1080*	1200

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

The Different Measures

Measure of Association	*Formula*	*Probabilistic Interpretation*
RR Relative Risk*	$\frac{a * (a + b + c + d)}{(a + c) * (a + b)}$	$\frac{\Pr(ae \mid drug)}{\Pr(ae)}$
PRR Proportional Reporting Ratio	$\frac{a / (a + b)}{c / (c + d)}$	$\frac{\Pr(ae \mid drug)}{\Pr(ae \mid \neg drug)}$
ROR Reporting Odds Ratio	$\frac{a / c}{b / d}$	$\frac{\Pr(ae \mid drug) / \Pr(\neg ae \mid drug)}{\Pr(ae \mid \neg drug) / \Pr(\neg ae \mid drug)}$
Information Component	$Log_2 \frac{a * (a + b + c + d)}{(a + c) * (a + b)}$	$\log_2 \frac{\Pr(ae \mid drug)}{\Pr(ae)}$

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 v.s. N=100, E=1)
  - GPS provides a shrinkage estimate of RR that addresses this concern.

$$RR_{ij}=N_{ij}*N../N_{i.}N_{.j}$$

	AE_j	Not AE_j	
$Drug_i$	N_{ij}		$N_{i.}$
Not $Drug_i$			
	$N_{.j}$		$N..$

Same Relative Reporting Ratio!

	AE_j	Not AE_j
$Drug_i$	*a=1*	*b=5*
Not $Drug_i$	*c=5*	*d=49*

Chi-square = 0.33

	AE_j	Not AE_j
$Drug_i$	*a=20*	*b=100*
Not $Drug_i$	*c=100*	*d=980*

Chi-square = 6.58

	AE_j	Not AE_j
$Drug_i$	*a=200*	*b=1000*
Not $Drug_i$	*c=1000*	*d=9800*

Chi-square = 65.8

GPS/MGPS

- Denote by ρ_{ij} the true RR for Drug *i* and AE *j*
- Assumes the ρ_{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
  - $N_{ij} \sim \text{Poisson}(\mu_{ij})$
  - $\lambda_{ij} \mid \lambda \sim p * g(\lambda; \alpha_1,\beta_1) + (1-p) * g(\lambda; \alpha_2,\beta_2)$

    a mixture of two Gamma Distributions
- EBGM = Geometric mean of Post-Dist. of λ_{ij}
  - Estimates of μ_{ij} / E_{ij}
  - "Shrinks" $N_{ij} / E_{ij} \rightarrow 1$
  - Smaller variances than N_{ij} / E_{ij}