Likelihood Ratio Test on Safety Data

Division of Biometrics VII
Office of Biostatistics
Office of Translational Sciences
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

For the ASA Safety Monitoring Working Group

August 22, 2017

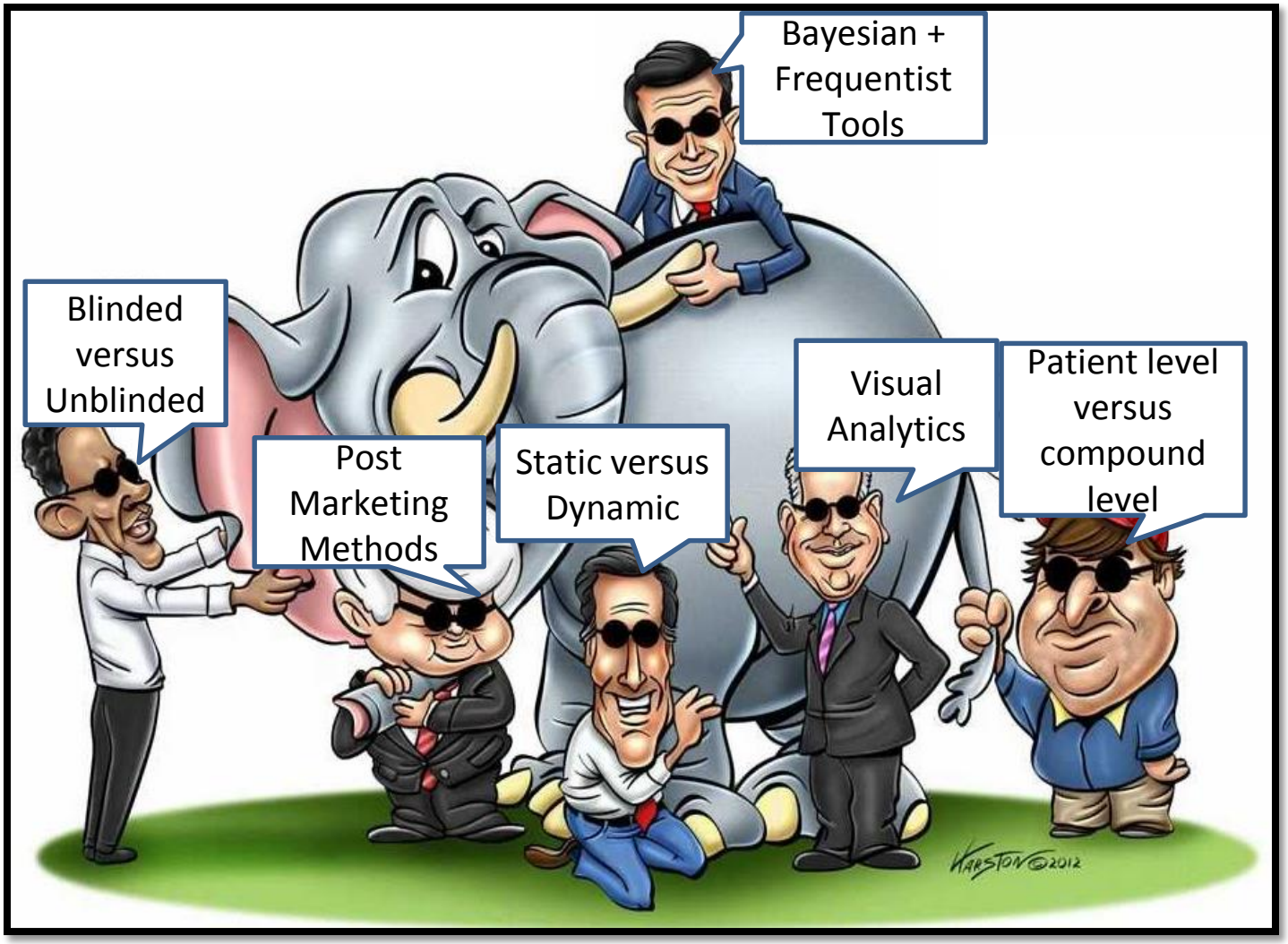
2017 DIA China Quantitative Science Forum

U.S. FOOD & DRUG ADMINISTRATION

Disclaimer

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies.

An Approach to Look at Safety Monitoring Elephant Metaphor



Adapted from: https://www.youtube.com/watch?v=ygT-UwgEaSo

3

ASA Safety Monitoring Working Group

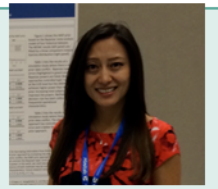
William Wang, Chair

WS1: Industry Practice & Regulation

Greg Ball Susan Duke

- Faiz Ahmad (Galderma)
- Greg Ball (Co-lead, Merck)
- Amit Bhattacharya (ACI Clinical)
- Brenda Crowe (Lilly)
- Susan Duke (Co-lead, Drug Safety Counts)
- Michael Fries (CSL Behring)
- Robert (Mac) Gordon (Janssen)
- Barbara Hendrickson* (AbbVie)
- Esteban Herrero-Martinez¥ (AbbVie)
- Juergen Kuebler† (Qscicon)
- Qi Jiang (Amgen)
- Dennis O'Brien* (BI)
- Lothar Tremmel (AstraZeneca)
- Wenquan Wang (Morphotek)
- William Wang (Chair, Merck)

WS2: Methodology



Judy Li Melvin Munsaka

- Michael Colopy (UCB)
- Michael Fries (CSL Behring)
- Karolyn Kracht (AbbVie)
- Judy Li (Co-lead, Regeneron)
- Li An Lin (Merck)
- Yong Ma (FDA)
- Melvin Munsaka (Co-lead, Takeda)
- Matilde Sanchez (Arena)
- Sourav Santra (Cytel)
- Krishan Singh (GSK)
- Ed Whalen (Pfizer)
- William Wang (Chair, Merck)
- Brian Waterhouse (AbbVie)
- Kefei Zhou (Theravance)
- Yueqin Zhao (FDA)

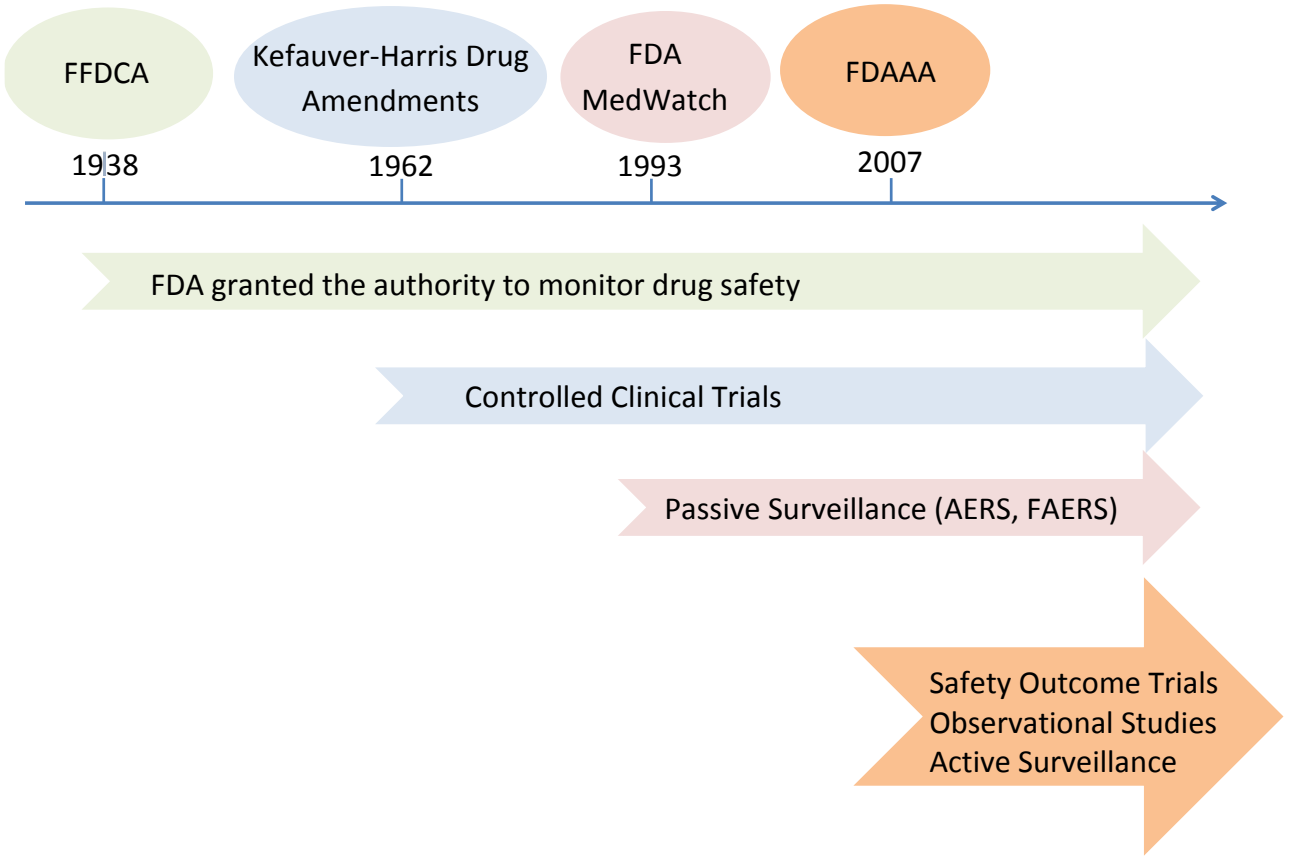
Special guest members
* Safety physician
¥ Regulatory affairs PV specialist
† European statistician

FDA U.S. FOOD & DRUG ADMINISTRATION

Outline

- Drug safety monitoring- key events
- Statistical methods in safety monitoring
- LRT method overview
- Details of the LRT methods

Key Events Leading to Changes in Drug Safety Monitoring



Pre-Marketing Safety Monitoring

- Largely descriptive
  - AE tables and SAE tables with proportions or rates
  - Recently, graphical display
- Three-tiered approach to AEs (Crowe et al, 2009)
  - Tier 1: Pre-specified hypothesis being defined (p values for formal comparison)
  - Tier 2: No pre-specified hypothesis and "common" AEs (proportions and confidence intervals)
  - Tier 3: Rare events (proportions or rates only)
- Multiplicity Issue:
  - While family wise error rate (FWER) control is commonly used for efficacy; False discovery rate (FDR) control is more appropriate in the context of safety, especially for Tier 2 AEs
  - Mehrotra & Heyse (2004) and Mehrotra & Adewale (2012) proposed a double FDR (DFDR) procedure to flag body systems/specific AEs
  - Berry & Berry (2004), Xia, Carlin & Ma (2011) proposed alternative Bayesian hierarchical mixed models to account for multiplicities in AE assessment

Post-Marketing Safety Monitoring (I) - Passive Surveillance

- Systems such as FAERS and VAERS
  - Conducted by the regulatory agencies
  - Data reported by consumers, physicians, manufacturers…
  - Both mandatory and voluntary reporting
  - Signal detection

- Statistical methods
  - Descriptive statistics with graphical display https://openfda.shinyapps.io/dash/
  - Reporting Odds Ratio (ROR)
  - Proportional Reporting Ratio (PRR)
    - https://openfda.shinyapps.io/RR_D/
    - https://openfda.shinyapps.io/RR_E/
    - https://openfda.shinyapps.io/dynprr/
  - Likelihood ratio – based method
    - https://openfda.shinyapps.io/LRTest/
    - https://openfda.shinyapps.io/LRTest_E/

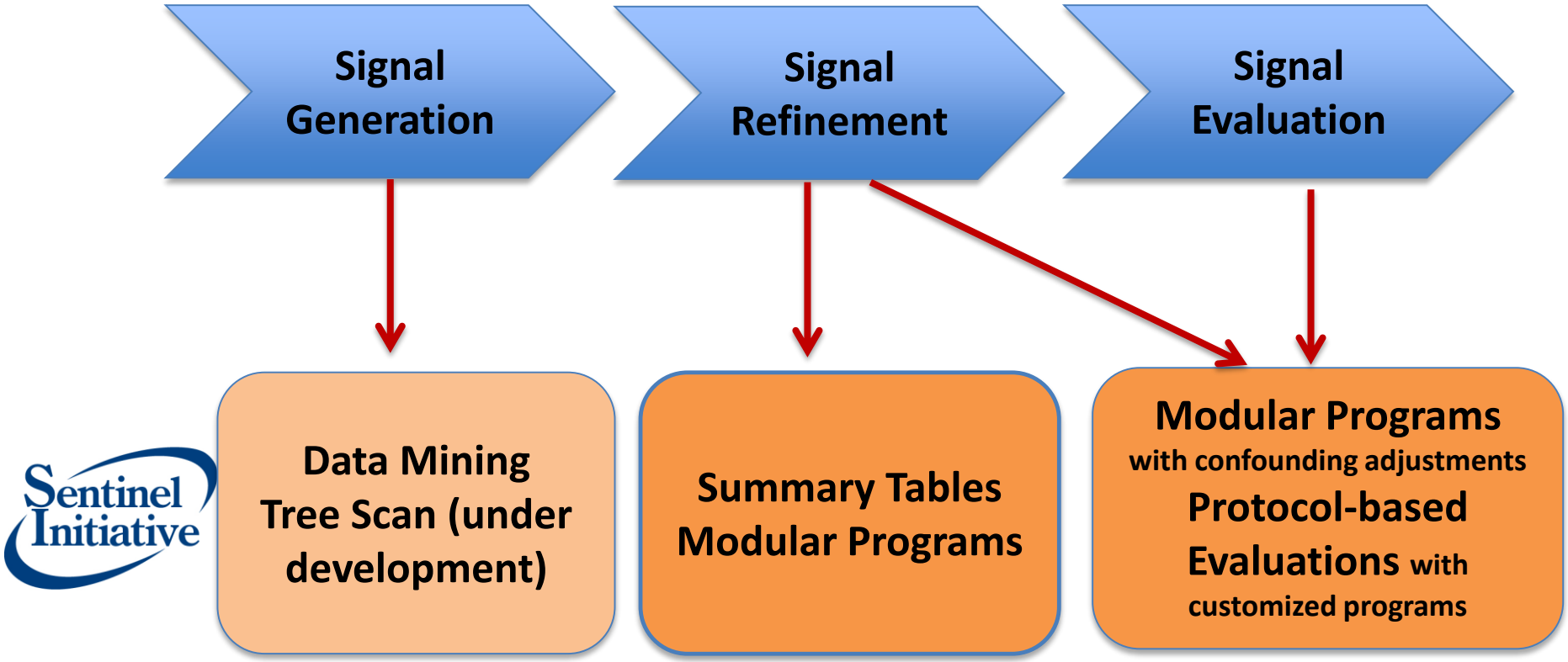
- Issues to consider: multiplicity, FDR

Post-Marketing Safety Monitoring (II) - Safety Outcome Trials

- Prospective, randomized, controlled trial
- Issued as post-marketing requirements (PMR) and conducted by the sponsors
- Similar to method used in non-inferiority efficacy trials – to rule out a pre-specified amount of risk called risk margin; noted as Δ
  - H_0: $\rho \geq \Delta$
  - H_1: $\rho < \Delta$
- Pre-specified safety outcome
- Study needs to be adequately powered
- Two-stage approach, group sequential designs, multiplicity adjustments….

Post-Marketing Safety Monitoring (III) - Active Surveillance

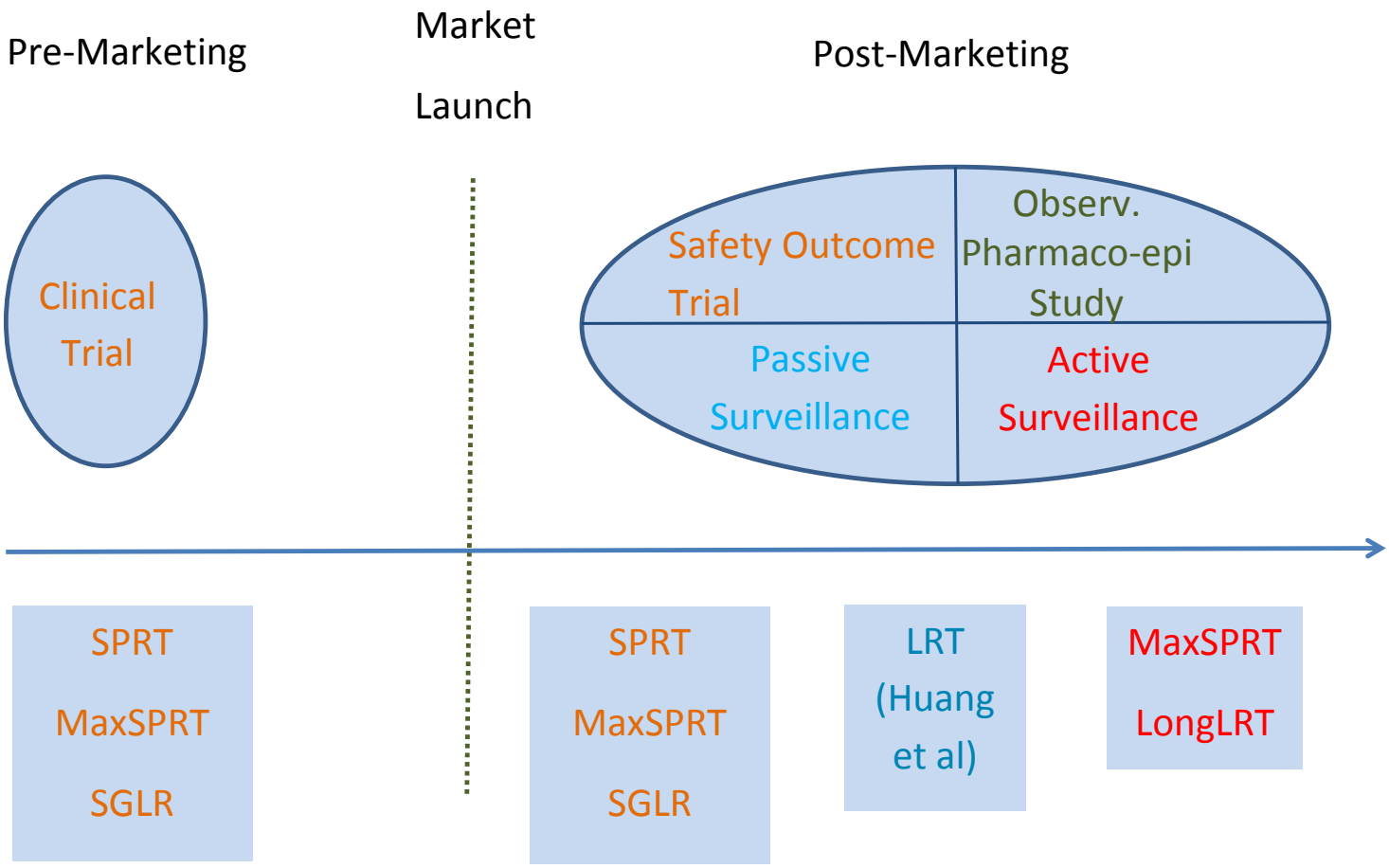
- The Sentinel System- https://www.sentinelinitiative.org/
- Signal detection/generation, refinement and validation



Post-Marketing Safety Monitoring (IV) - Observational Pharmacoepidemiologic Studies

- Designed to assess a serious risk associated with a drug exposure
- Conducted by the sponsor or the regulatory agencies
- Protocol-based with pre-specified hypothesis, mostly with a control group
- Data sources include administrative health care claims data, electronic medical records, registries, prospectively collected observational data, others
- Statistical methods:
  - Meta-analysis
  - Propensity score matching to 'mimic' clinical trials
  - Inverse probability weighted method
  - Marginal Structural Models (MSM) to control for time-dependent confounding

Likelihood Ratio-Based Methods in Safety Studies



SPRT: Sequential Probability Ratio Test
MaxSPRT: Maximized SPRT
SGLR: Sequential Generalized Likelihood Ratio Test

LRT: Likelihood Ratio Test
LongLRT: Longitudinal Likelihood Ratio Test

Sequential Probability Ratio Test (SPRT)

- Developed by Abraham Wald in 1945 for use in quality control
- A continuous sequential monitoring test with pre-determined upper and lower bound
- For testing a simple null hypothesis *($H_{0:}\ p=p_0$)* against a simple alternative hypothesis *($H_{a:}\ p=p_a$)*
- Let C_t be the random variable representing the number of adverse events during the time period [0, t].
- Log-likelihood ratio test statistic *(LLR_t)* is computed at every time point $t > 0$ as additional data are collected
- $LLR_t = ln \frac{P(C_t=ct \mid H_a)}{P(C_t=ct \mid H_0)}$
- Continuously in time after each new observation enters the study

SPRT….contd.

- LLR_t is sequentially monitored for all values of t > 0 until

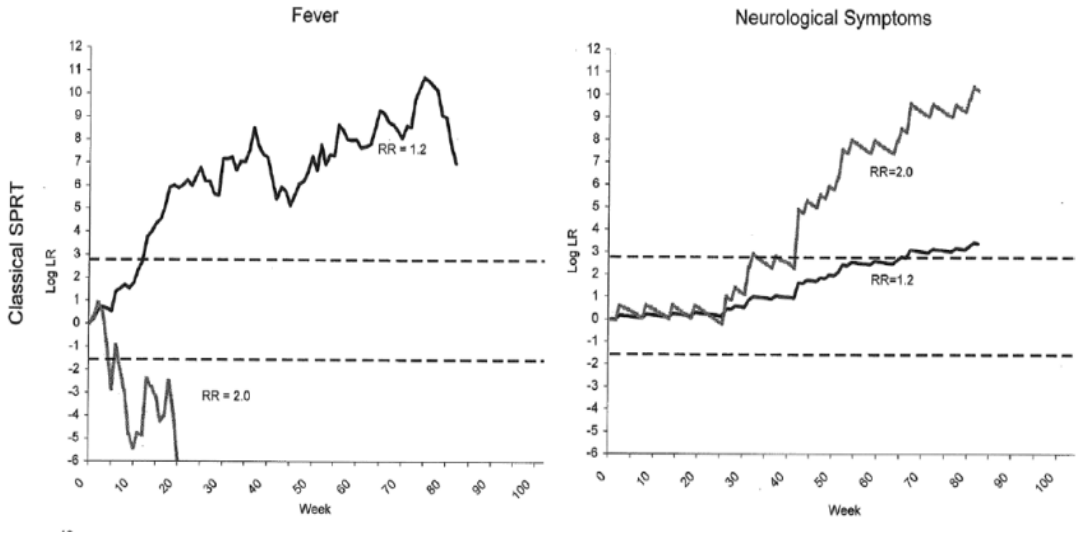
  $LLR_t \geq ln\,[1 - \beta)/\alpha]$ reject H_0, or until

  $LLR_t \leq ln\,[\beta/(1 - \alpha)]$ accept H_0

- Pros: looking at the data in a continuous fashion or as often as needed while controlling type 1 error
- Cons:
  - Designed to monitor a <u>known</u> safety problem, not for signal detection/generation
  - Highly sensitive to the choice of H_a. Inappropriate choice of H_a can delay the detection of a safety signal or completely miss it
  - Practical implications can be problematic, because often it is difficult to know beforehand excess relative risk we should look for.

Example: SPRT method

- Pediarix vaccine: historical data from CDC-VSD project
  - Combination vaccine protecting children from 5 different diseases
  - prospective weekly monitoring to evaluate risk of fever or neurological symptoms within 28 days after vaccination
- SPRT applied with Poisson and binomial models
  - Results highly dependent on relative risk *RR* used (*RR* = 1.2 and *RR* = 2.0) for the alternate hypothesis H_a



- When $H_a = 2.0$ is used, the detection of fever signal is missed, whereas when $H_a = 1.2$ is used, the detection of neurological symptoms signal is delayed.

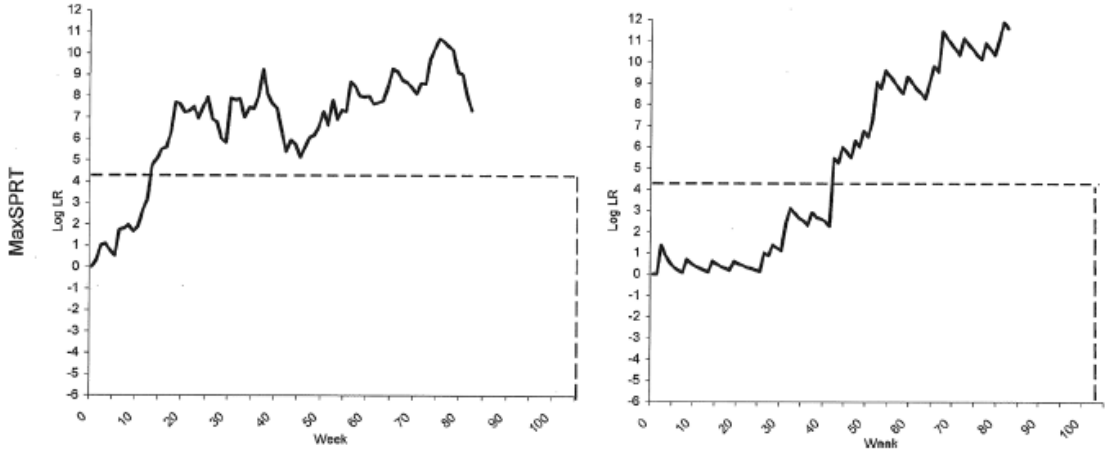
Kulldorff (2011)

Maximized SPRT (MaxSPRT)

- Developed by Kulldorff et al, 2011 to overcome the drawbacks with SPRT
- In MaxSPRT, H_a is a composite, e.g., H_a: $RR > 1$
- The test statistic is the maximum likelihood under H_a divided by the likelihood under H_0

$$LR_t = \max_{H_a} \frac{P(Ct=ct \mid H_a)}{P(Ct=ct \mid H_0)}$$

- No closed form formula for the upper/lower bound, Monte Carlo simulations or recursive numerical integration are used to compute them
- Applying MaxSPRT to the Pediarix vaccine data, safety signal was detected for both fever and neurological symptoms (Lower bound is not defined in this example as there is no reason to stop the study if the drug/vaccine is safe)



Kulldorff (2011)

Sequential Generalized Likelihood Ratio (SGLR) Test

- Shih et al (2010) proposed a new class of sequential generalized likelihood ratio (SGLR) tests for evaluating AEs in vaccine studies
- The SGLR test can be used for 2-arm pre-licensure clinical trials and 1-arm post-licensure surveillance studies
- A key feature of the SGLR tests is the exponential family models of the rare events sequence under the Poisson arrival of adverse events
- Shih et al illustrated the test using data from a blinded placebo-controlled randomized rotavirus efficacy and safety trial (REST).
- Simulation studies show the superior performance of the SGLR test compared to the SPRT and MaxSPRT in both 2-arm randomized clinical trials and 1-arm post-licensure studies in terms of expected total number of events and probability of rejecting H_O
- A truncated version of MaxSPRT when applied to 1-arm surveillance studies is in fact a SGLR test without a lower boundary; introducing a suitable lower boundary can lead to substantial savings in sample size with little loss of power
- R package using available at http://med.Stanford.edu/ClinicalTrialMethodology.html

Example: SGLR method

Table I. Power and expected number of events for various sequential tests of $H_0: \lambda_V/\lambda_C = 1$ versus $H_1: \lambda_V/\lambda_C \geqslant 3$ in a single-armed post-licensure study.

		SPRT†			
λ_V/λ_C	GLR*	$\gamma = 2.0$	3.0	5.0	MaxSPRT‡
(a) Expected total number of events					
1.0	3.6	9.7	4.4	2.5	957.9
2.0	7.4	14.7	9.8	4.9	20.4
3.0	6.7	8.7	7.1	5.7	9.6
4.0	5.2	7.1	5.5	4.9	6.7
5.0	4.2	6.4	4.8	4.2	5.4
(b) Probability of rejecting H_0					
1.0	0.049	0.041	0.038	0.030	0.049
2.0	0.523	0.949	0.648	0.305	1.000
3.0	0.900	1.000	0.966	0.734	1.000
4.0	0.956	1.000	0.997	0.925	1.000
5.0	0.957	1.000	1.000	0.980	1.000

*The thresholds $b_0 = 3.435$, $b_1 = 1.822$ are chosen such that $p_{\lambda_V/\lambda_C=1}(\text{reject } H_0) \leqslant 0.05$, $p_{\lambda_V/\lambda_C=3}(\text{accept } H_0) \leqslant 0.10$.

†Truncated at $n^* = 1000$; γ is the assumed alternative value of λ_V/λ_C in the likelihood ratio statistic.

‡Truncated at $n^* = 1000$; the threshold $b = 4.306$ is chosen such that $p_{\lambda_V/\lambda_C=1}(\text{reject } H_0) \leqslant 0.05$.

Shih (2010)

LRT in Safety Signal Detection

- Signal detection with multiple AEs and multiple drugs in the FAERS system
- Huang et al (2011) proposed LRT method based on 2x2 tables

	$Drug_j$	Other drugs	
AE_i	n_{ij}	Subtracted	$n_{i.}$
Other AEs	Subtracted	Subtracted	Subtracted
	$n_{j.}$	Subtracted	$n_{..}$

- Hypothesis:

$$n_{ij} \sim Pois(n_{i.} p_i)$$

$$n_{.j} - n_{ij} \sim Pois((n_{..} - n_{i.}) \times q_i)$$

$H_0 : p_i = q_i = p^*$ for all AEs, i

$H_a : p_i > q_i$ for all at least one AE, i

- Test statistic

$$LR_{ij} = \frac{L_a(\hat{p}, \hat{q})}{L_0(\hat{p}^*)} = \left[\frac{\left(\frac{n_{ij}}{n_{i.}}\right)^{n_{ij}} \left(\frac{n_{.j} - n_{ij}}{n_{..} - n_{i.}}\right)^{n_{.j} - n_{ij}}}{\left(\frac{n_{.j}}{n_{..}}\right)^{n_{.j}}} \right],$$

Longitudinal Likelihood Ratio Test (LongLRT)

- There was a need to factor in the exposure information for active surveillance in large observational or clinical trial databases
- Huang et al (2014) proposed the longitudinal likelihood ratio test (LongLRT)
- Authors present several definitions of drug exposure (e.g., event-time, person-time, and exposure-time) that can be used as a denominator
- All AE cases that occur during the exposure period are called countable cases
- If recurrent AEs and all exposure-time are considered, Poisson distribution is used
- Asymptotic distribution of the LongLRT statistic is analytically not tractable, Monte Carlo simulations are used to obtain the empirical distribution and to compute the critical boundaries
- LongLRT controls (family-wise) type I error. One does not have to stop the analysis after a signal is detected at a look, the number of looks for the LongLRT method is usually not pre-specified.

20

Conclusion Remarks

- LRT methods applied to safety data have gained popularity in the recent decade
- LRT can be very useful in drug safety monitoring
  - In passive or active surveillance
  - In sequential safety monitoring
  - Others in the future?

Thanks to

- Lan Huang
- Mat Soukup
- Rongmei Zhang
- Clara Kim
- William Wang
- Mark Levenson

For slides and reviews

Reference

1. Crowe BJ, Xia HA, Berlin JA, Watson DJ, Shi H, Lin SL, Kuebler J, Schriver RC, Santanello NC, Rochester G, Porter JB, Oster M, Mehrotra DV, Li Z, King EC, Harpur ES, Hall DB. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clinical Trials*. 2009 Oct;6(5):430-40.
2. Chuang-Stein, Christy, and H. Amy Xia. The practice of pre-marketing safety assessment in drug development. *Journal of Biopharmaceutical Statistics* 23.1 2013: 3-25.
3. Wald, Abraham. Sequential Tests of Statistical Hypotheses. *Annals of Mathematical Statistics*. 1945, **16** (2): 117–186.
4. Martin Kulldorff , Robert L. Davis , Margarette Kolczak , Edwin Lewis , Tracy Lieu & Richard Platt A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance, *Sequential Analysis*, 2011 30:1, 58-78
5. Shih, M.-C., Lai, T. L., Heyse, J. F. and Chen, J. (2010), Sequential generalized likelihood ratio tests for vaccine safety evaluation. *Statist. Med*., 2010, 29: 2698–2708.
6. Lan Huang, Jyoti Zalkikar, and Ram Tiwari. A likelihood based method for signal detection with application to FDA's drug safety data. *Journal of the American Statistical Association (JASA), 2011,* 106 (496), 1230-1241
7. Lan Huang, Jyoti Zalkikar, Ram Tiwari. Likelihood ratio tests for longitudinal drug safety data. *Statistics in Medicine*, 2014, 33(14), 2408-2424
8. Lan Huang, Ted Gou, Jyoti Zalkikar, Ram Tiwari. A review of statistical methods for safety surveillance. *Therapeutic Innovation & Regulatory Science*, 2014, 48 (1), 98-108


FDA
U.S. FOOD & DRUG
ADMINISTRATION