Measuring method performance in prospective safety monitoring

Joshua J. Gagne

Division of Pharmacoepidemiology

Brigham and Women's Hospital

Brookings Institution

February 16, 2011

This work is supported by NLM grant 5RC1LM010351-02

Objectives of monitoring methods

- Identify all safety issues of interest (sensitivity)
- Produce as few false positives as possible (specificity)
- But, in prospective monitoring, we also want:
  - True positives identified early and false positives identified late (if they are to occur at all)
  - Performance weighted differently in different scenarios

DoPE
BWH

Imagine the following scenarios…

- Hypothetical monitoring scenario 1:
  - Febrile seizures after administration of one type of vaccine (monitoring vaccine) versus another (comparison vaccine)
  - Monitoring vaccine increases febrile seizure risk by 2-fold
- Hypothetical monitoring scenario 2:
  - AMI in users of an anti-diabetic drug (monitoring drug) versus users of another anti-diabetic drug (comparison drug)
  - Monitoring drug does not increase AMI risk
- Two hypothetical monitoring methods could be used
- New data become available on a quarterly basis, for a total of 10 periods

DoPE BWH

Hypothetical febrile seizure data

	Numbers of new febrile seizure observed in each quarter										
		1	2	3	4	5	6	7	8	9	10	Total
1:1 match	**Monitoring vaccine**	20	20	20	20	20	20	20	20	20	20	200
	Comparison vaccine	10	10	10	10	10	10	10	10	10	10	100

Hypothetical febrile seizure data

	Numbers of new febrile seizure observed in each quarter										
		1	2	3	4	5	6	7	8	9	10	Total
1:1 match	**Monitoring vaccine**	20	20	20	20	20	20	20	20	20	20	200
	Comparison vaccine	10	10	10	10	10	10	10	10	10	10	100

Method A

Hypothetical febrile seizure data

	Numbers of new febrile seizure observed in each quarter										
		1	2	3	4	5	6	7	8	9	10	Total
1:1 match	**Monitoring vaccine**	20	20	20	20	20	20	20	20	20	20	200
	Comparison vaccine	10	10	10	10	10	10	10	10	10	10	100

Method A

Method B

Hypothetical AMI data

1:1 match

Numbers of new AMI observed in each quarter											
	1	2	3	4	5	6	7	8	9	10	Total
Monitoring drug	12	12	12	8	8	8	13	9	8	10	100
Comparison drug	9	10	8	10	10	11	9	12	11	10	100

Hypothetical AMI data

Numbers of new AMI observed in each quarter											
	1	2	3	4	5	6	7	8	9	10	Total
Monitoring drug	12	12	12	8	8	8	13	9	8	10	100
Comparison drug	9	10	8	10	10	11	9	12	11	10	100

1:1 match

Method A

Hypothetical AMI data

Numbers of new AMI observed in each quarter											
	1	2	3	4	5	6	7	8	9	10	Total
Monitoring drug	12	12	12	8	8	8	13	9	8	10	100
Comparison drug	9	10	8	10	10	11	9	12	11	10	100

1:1 match

Method A

Method B

Hypothetical AMI data

Numbers of new AMI observed in each quarter											
	1	2	3	4	5	6	7	8	9	10	Total
Monitoring drug	12	12	12	8	8	8	13	9	8	10	100
Comparison drug	9	10	8	10	10	11	9	12	11	10	100

1:1 match

Method A

Method B

		Truth +	Truth -
Method result	+	1	1
	-	0	0

Sensitivity = 1.00

Specificity = 0.00

An event-based approach

Numbers of exposed events in each scenario in each quarter											
	1	2	3	4	5	6	7	8	9	10	Total
Scenario 1 (vaccine)	20	20	20	20	20	20	20	20	20	20	200
Scenario 2 (drug)	12	12	12	8	8	8	13	9	8	10	100

DoPE
BWH

An event-based approach

Numbers of exposed events in each scenario in each quarter											
	1	2	3	4	5	6	7	8	9	10	Total
Scenario 1 (vaccine)	20	20	20	20	20	20	20	20	20	20	200
Scenario 2 (drug)	12	12	12	8	8	8	13	9	8	10	100

Method A

		Truth +	Truth -
Method result	+	180	27
	-	20	73

Event-based sensitivity = 0.90
Event-based specificity = 0.73

An event-based approach

Numbers of exposed events in each scenario in each quarter											
	1	2	3	4	5	6	7	8	9	10	Total
Scenario 1 (vaccine)	20	20	20	20	20	20	20	20	20	20	200
Scenario 2 (drug)	12	12	12	8	8	8	13	9	8	10	100

Method A

		Truth +	Truth -
Method result	+	180	27
	-	20	73

Event-based sensitivity = 0.90
Event-based specificity = 0.73

Method B

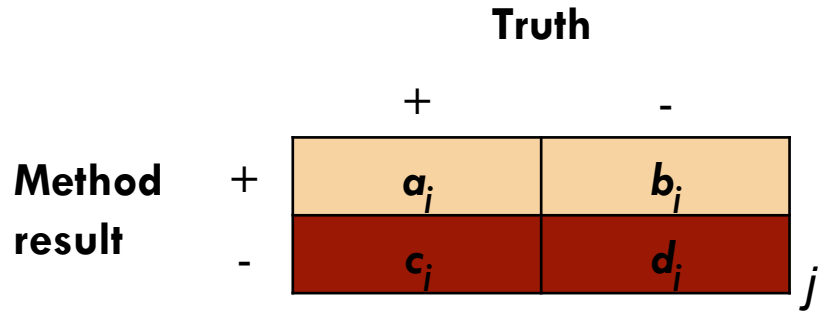
		Truth +	Truth -
	+	120	64
	-	80	36

Event-based sensitivity = 0.60
Event-based specificity = 0.36

Event-based performance metric

$$EBP = \frac{\sum_{j=1}^{n} a_j \cdot w_j}{\sum_{j=1}^{n} a_j + c_j} + \frac{\sum_{j=1}^{n} d_j \cdot (1 - w_j)}{\sum_{j=1}^{n} d_j + b_j}$$

- Where:
  - w_j = weight for sensitivity:specificity preference in scenario *j*
  - *n* = total number of *j* scenarios



A vs. B comparison with *EBP*

Method A

Method result	Truth +	Truth -
+	180	27
-	20	73

Event-based sensitivity = 0.90
Event-based specificity = 0.73

Method B

Method result	Truth +	Truth -
+	120	64
-	80	36

Event-based sensitivity = 0.60
Event-based specificity = 0.36

- $w_i = 0.1$ for febrile seizures; $w_i = 0.2$ for AMI:
  - $EBP_{method\_A} = 0.674$ vs. $EBP_{method\_B} = 0.348$
- $w_i = 0.01$ for febrile seizures; $w_i = 0.01$ for AMI:
  - $EBP_{method\_A} = 0.731$ vs. $EBP_{method\_B} = 0.362$

DoPE BWH

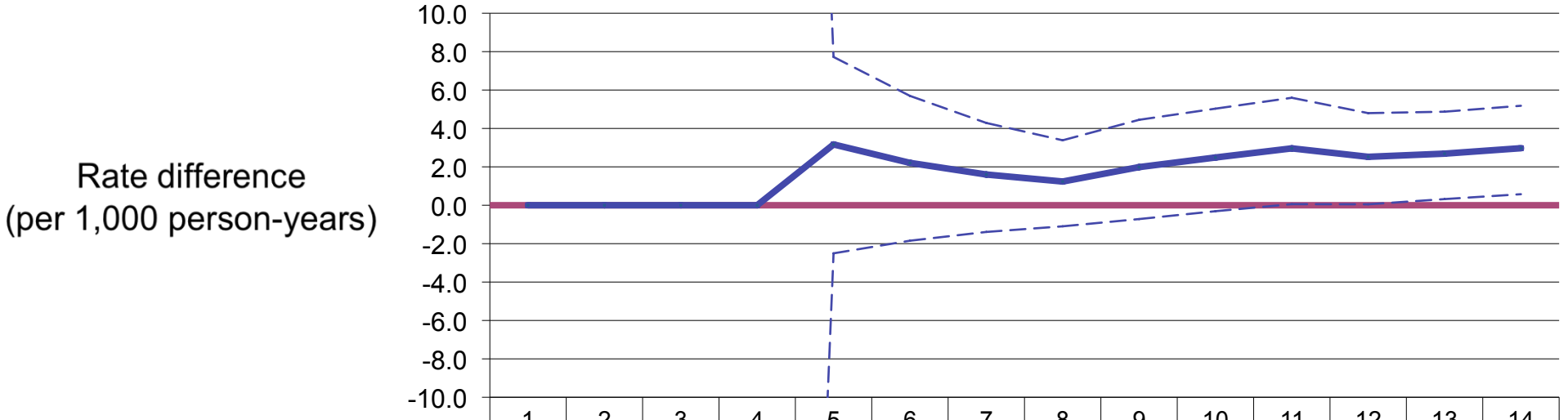
Application in simulated data

- Simulated data resembling signal refinement for statin-induced rhabdomyolysis
- Assumed matched cohort design with 20 periods defined by calendar time (e.g. quarters)
- Compared 93 algorithms
- 100,000 iterations

w_i	Algorithm	Overall sensitivity	Overall specificity	*EBP*
0.01	maxSPRT ($\alpha = 0.01$)	0.18	1.00	0.99
0.10	Pocock-like spending function ($\alpha = 0.10$)	0.32	0.99	0.93
0.20	Pocock-like spending function ($\alpha = 0.40$)	0.49	0.95	0.87

DoPE BWH

Translation to empirical data



	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Lower 95% confidence interval	-99.00	-99.00	-99.00	-99.00	-2.50	-1.85	-1.39	-1.10	-0.72	-0.31	0.06	0.05	0.32	0.58
Cumulative rate difference	0.00	0.00	0.00	0.00	3.17	2.21	1.60	1.24	1.99	2.49	2.96	2.52	2.69	2.97
Upper 95% confidence interval	99.00	99.00	99.00	99.00	7.72	5.70	4.29	3.38	4.45	5.03	5.60	4.80	4.87	5.18
Cumulative exposed events	0	0	0	0	1	1	1	1	2	3	4	4	5	6
Cumulative unexposed events	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cumulative exposed person-years	19	60	124	197	316	453	625	809	1007	1206	1350	1586	1861	2019
Cumulative unexposed person-years	26	65	127	218	348	497	681	877	1084	1299	1462	1715	2010	2275

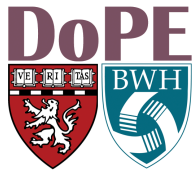
w_i	Algorithm	Alert period
0.01	maxSPRT (α = 0.01)	--
0.10	Pocock-like spending function (α = 0.10)	14
0.20	Pocock-like spending function (α = 0.40)	11

Summary

- *EBP* is useful to compare performance of methods in settings in which truth is known
- Uses exposed events to incorporate time to signaling
  - Most relevant unit for public health decision making (i.e. modifiable)
  - The most meaningful measure of "time" in prospective monitoring
  - Main driver of statistical power
- Allows for transparent weighting of sensitivity-specificity tradeoffs that can vary by monitoring outcome
- Does not depend on "prevalence" of safety issues

DoPE BWH

jgagne1@partner.org



Summarizing event-based performance

- Many ways to use events to summarize performance:
  - Mean average precision, diagnostic odds ratio, F_1 score, Matthews correlation coefficient, Youden's J statistic, accuracy, etc.
  - All are functions of sensitivity, specificity, and (often) "prevalence"
  - Built-in tradeoffs between FP and FN costs are arbitrary and cannot be easily changed
- Sensitivity and specificity are characteristics of the methods whereas prevalence is arbitrary

DoPE
BWH