Sentinel

Statistical Power for Use of Tree-Based Scan Statistics for Surveillance of Infant Outcomes Following Maternal Perinatal Medication Use

Presented at ICPE 2021 All Access

Elizabeth A Suarez[1], Michael Nguyen[2], Di Zhang[3], Yueqin Zhao[3], Danijela Stojanovic[2], Monica Munoz[4], Jane Liedtka[5], Abby Anderson[6], Wei Liu[7], Inna Dashevsky[1], Sandra DeLuccia[1], Talia Menzin[1], Jennifer Noble[1], Judith C Maro[1]

1. Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School; 2. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, US Food and Drug Administration; 3. Office of Biostatistics, Center for Drug Evaluation and Research, FDA; 4. Division of Pharmacovigilance, Center for Drug Evaluation and Research, US Food and Drug Administration; 5. Division of Pediatric and Maternal Health, Center for Drug and Evaluation Research, US Food and Drug Administration; 6. Division of Urology, Obstetrics and Gynecology, Center for Drug and Evaluation Research, US Food and Drug Administration; 7. Division of Epidemiology, Center for Drug and Evaluation Research, US Food and Drug Administration

August 23, 2021

Disclosures

- The views expressed in this presentation represent those of the presenters and do not necessarily represent the official views of the U.S. FDA.
- This project was supported by Task Order HHSF22301012T under Master Agreement HHSF223201400030I from the US Food and Drug Administration (FDA).

TreeScan™ and the Role of Signal Identification

- TreeScan is a statistical data mining tool that can be used for signal identification in pharmacovigilance/ pharmacoepidemiologic analyses
- Signal identification: systematic evaluation of potential adverse events related to the use of medical products without pre-specifying an outcome of interest
- TreeScan can supplement current practices (pregnancy exposure registries and administrative database studies) by using large administrative databases to simultaneously scan for new and unsuspected potential safety concerns that can be investigated in targeted studies
- TreeScan has been used with multiple study designs (self controlled, propensity score matched) for studies of vaccine and other medication safety in adults – but is it appropriate for use in pregnant populations?

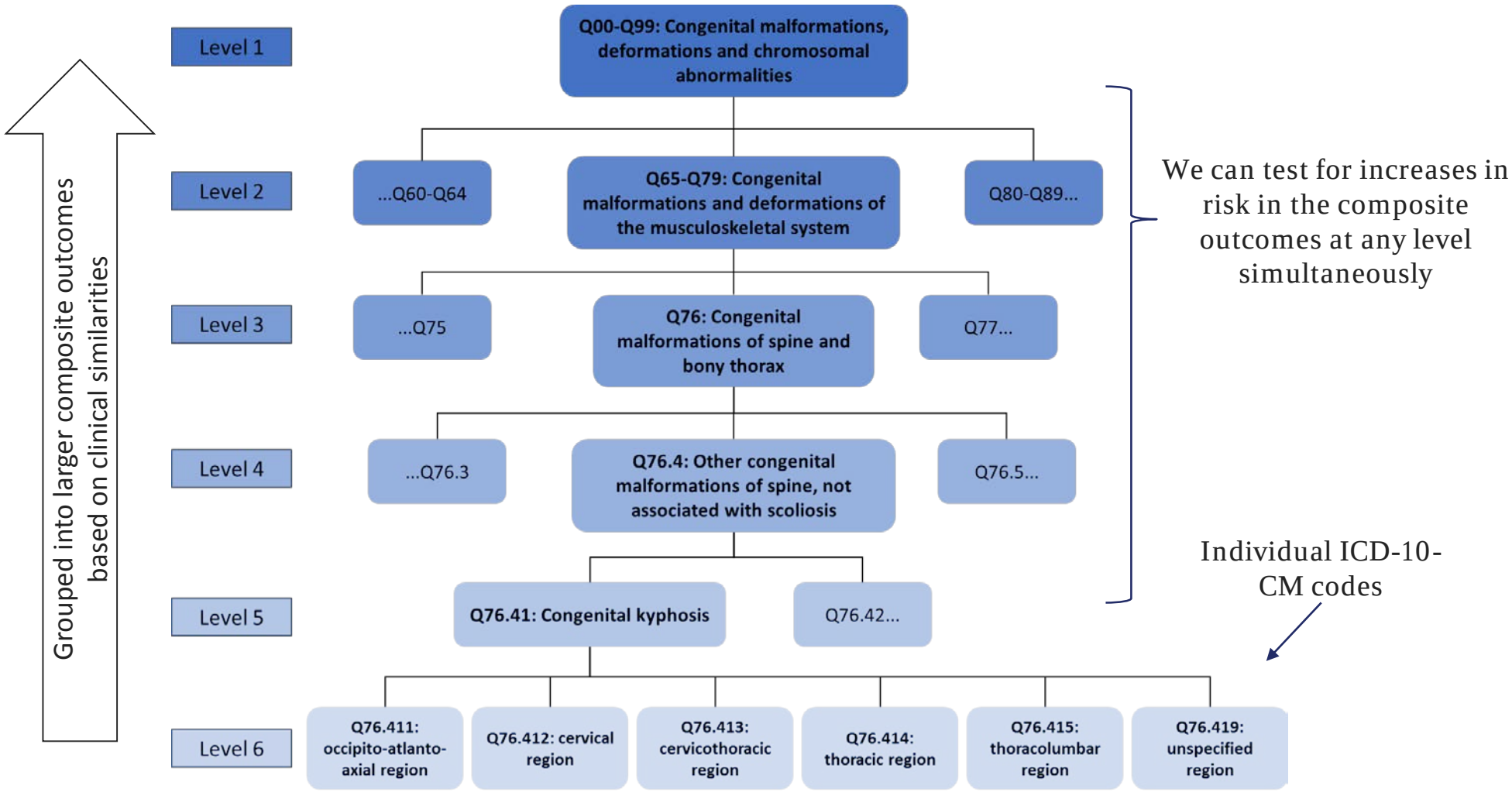
http://www.treescan.org/
Kulldorff, et al. Drug safety data mining with a tree-based scan statistic. 2013. Pharmacoepidemiol Drug Saf.

Study Aims

- To assess the performance of TreeScan under known conditions and with various study design decisions to inform future implementations of TreeScan for pregnancy exposure monitoring in the FDA Sentinel system
- Specifically, to estimate power to detect signals
- Why are we focused on power?
  - To understand the usefulness of TreeScan for rare exposures
  - Power = more timely signal detection

Methods

The Outcome Tree



TreeScan Statistics and p-values for Alerting

- Hypothesis testing:
  - Null: there is no increase in risk across any outcome in the tree in the exposed group
  - Alternative: there is an increase in risk for any outcome in the exposed group
- Formal adjustment for multiple testing to reduce false positives
- A statistical alert occurs when an outcome meets a pre-specified p-value threshold, e.g., <0.05
- Two probability models: Bernoulli and Poisson
  - These models use the referent population in different ways to calculate the expected outcome count in the exposed group
  - Both are compatible with different study designs with propensity scores that we will use to control for confounding:
    - Bernoulli: fixed ratio propensity score matching
    - Poisson: propensity score stratification or weighting
  - We are interested in comparing Bernoulli and Poisson methods for maximizing power, not assessing different confounding control methods

General Simulation Methods

1. Used empirical data to estimate the background incidence of outcomes in our tree
   - IBM MarketScan® Research Database
   - Estimated outcome incidence for each outcome in the tree in an unexposed referent population of pregnant women linked to infants
2. Simulated cohorts with known increases in risk of pre-specified outcomes
   - Selected malformation outcomes with incidence varying from approximately 1 per 10,000 to 1 per 100
   - Increased the risk for that pre-specified outcome by a risk ratio of 1.5, 2, or 4
   - Varied the size of the exposed sample
3. Calculated power to detect the known increase in risk in the simulated cohort using the TreeScan software

Analyses were designed on Sentinel Query Request Package (QRP) version 9.6.0, with Propensity Score Analysis module, Signal Identification module, and ad hoc programming

Methods

Question 1: What is the power to identify signals in scenarios expected in a pregnancy study?

Question 2: Can we improve power with changes to our study design? (propensity score methods)

Question 3: How does the sensitivity or PPV of the outcome definition impact power?

Directly compare the Poisson and Bernoulli models, starting with the same inputs:

- Number of exposed pregnancies
- Relative risk (RR)
- Outcome prevalence

This will inform which model, and therefore study design, may be optimal for routine analyses in the future

Methods

Question 1: What is the power to identify signals in scenarios expected in a pregnancy study?

Question 2: Can we improve power with changes to our study design? (propensity score methods)

Question 3: How does the sensitivity or PPV of the outcome definition impact power?

Bernoulli model used with 1:1 propensity score matching will discard referent pregnancies when the referent group is large

If we match >1 referent pregnancy to each exposed pregnancy (i.e., fixed ratio N:1 matching), do we increase power over a 1:1 matched design?

We simulated propensity score distributions with varying levels of overlap and calculated power after increasing the fixed matching ratio

Methods

Question 1: What is the power to identify signals in scenarios expected in a pregnancy study?

Question 2: Can we improve power with changes to our study design? (propensity score methods)

Question 3: How does the sensitivity or PPV of the outcome definition impact power?

The outcome definition is the same across the tree – a single instance of a diagnosis code in the mother or infant record that meets an incidence criteria

Highly sensitivity but possible low specificity → bias RR estimates towards the null

Does high sensitivity or high positive predictive value (PPV) maximize power?

Results

Question 1: Power estimates varying TreeScan model, outcome incidence, sample size, and RR

	Bernoulli	RR			Poisson	RR		
	# exposed	1.5	2.0	4.0	# exposed	1.5	2.0	4.0
Incidence = 8 per 1000 Q21.0: ventricular septal defect	2000	0.08	0.24	1	2000	0.09	0.49	1
	4000	0.11	0.56	1	4000	0.21	0.89	1
	8000	0.24	0.91	1	8000	0.56	1	1
	15000	0.54	1.00	1	15000	0.92	1	1
	20000	0.77	1.00	1	20000	0.98	1	1
	30000	0.93	1.00	1	30000	1	1	1
Incidence = 1.8 per 1000 Q40.0: pyloric stenosis	2000	0.06	0.08	0.36	2000	0.05	0.09	0.72
	4000	0.05	0.09	0.75	4000	0.06	0.16	0.97
	8000	0.06	0.15	0.98	8000	0.10	0.44	1
	15000	0.07	0.40	1	15000	0.18	0.82	1
	20000	0.11	0.58	1	20000	0.25	0.93	1
	30000	0.15	0.78	1	30000	0.44	0.99	1
Incidence = 0.6 per 1000 Q35.9: cleft palate, unspecified	2000	0.06	0.06	0.06	2000	0.05	0.05	0.16
	4000	0.05	0.05	0.15	4000	0.06	0.08	0.50
	8000	0.06	0.09	0.35	8000	0.07	0.12	0.86
	15000	0.05	0.10	0.77	15000	0.08	0.24	1
	20000	0.06	0.11	0.95	20000	0.08	0.31	1
	30000	0.05	0.16	1	30000	0.11	0.52	1

Question 1: Power estimates varying TreeScan model, outcome incidence, sample size, and RR

	Bernoulli	RR			Poisson	RR		
	# exposed	1.5	2.0	4.0	# exposed	1.5	2.0	4.0
Incidence = 8 per 1000 Q21.0: ventricular septal defect	2000	0.08	0.24	1	2000	0.09	0.49	1
	4000	0.11	0.56	1	4000	0.21	0.89	1
	8000	0.24	0.91	1	8000	0.56	1	1
	15000	0.54	1.00	1	15000	0.92	1	1
	20000	0.77	1.00	1	20000	0.98	1	1
	30000	0.93					1	1
Incidence = 1.8 per 1000 Q40.0: pyloric stenosis	2000	0.06					9	0.72
	4000	0.05					6	0.97
	8000	0.06					4	1
	15000	0.07					2	1
	20000	0.11					3	1
	30000	0.15					9	1
Incidence = 0.6 per 1000 Q35.9: cleft palate, unspecified	2000	0.06					.05	0.16
	4000	0.05	0.05	0.15	4000	0.06	0.08	0.50
	8000	0.06	0.09	0.35	8000	0.07	0.12	0.86
	15000	0.05	0.10	0.77	15000	0.08	0.24	1
	20000	0.06	0.11	0.95	20000	0.08	0.31	1
	30000	0.05	0.16	1	30000	0.11	0.52	1

Poisson has greater power than Bernoulli

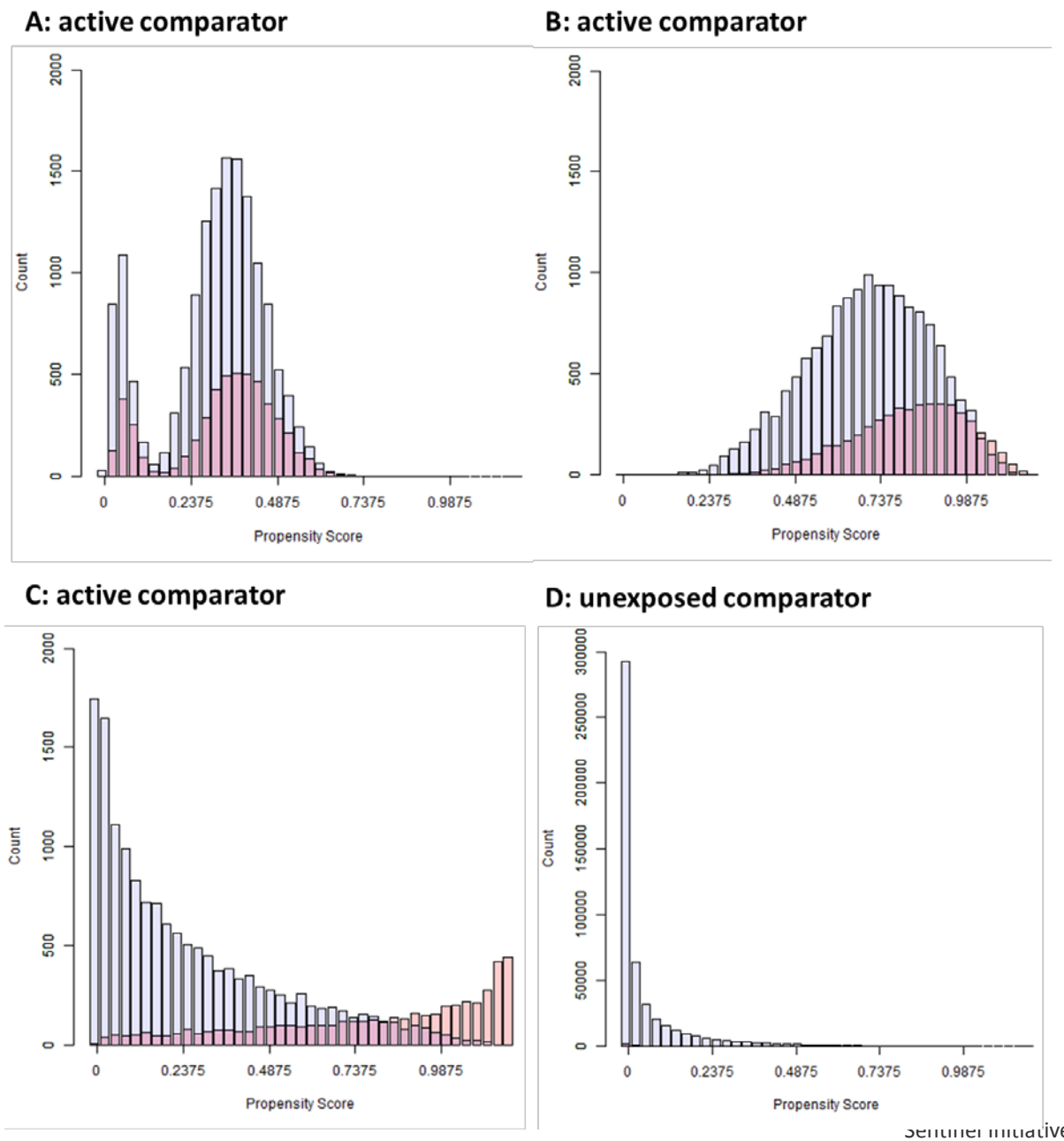
A minimum of 4000 exposed pregnancies is necessary to observe a doubling in risk of common outcomes with approximately 90% power

Question 2: Increasing the fixed matching ratio with the Bernoulli model

Simulated propensity score distributions

Base population:

- 5,000 exposed pregnancies and 20,000 comparator exposed pregnancies for scenarios A-C
- 5,000 exposed and 495,000 unexposed pregnancies for scenario D



Question 2: Increasing the fixed matching ratio with the Bernoulli model

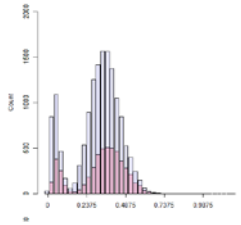
Incidence = 8 per 1000
Q21.0: ventricular septal defect

A

B

C

Active comparator with decreasing overlap in propensity score distributions



D:
Unexposed
comparator

Matching ratio	Exposed N	Change from 1:1	Referent N	Change from 1:1	Full N	Power for Q21.0 for RR=2
1:1	4999		4,999		9,998	0.59
2:1	4997	0%	9,994	100%	14,991	0.79
3:1	4711	-6%	14,133	183%	18,844	0.84

Question 2: Increasing the fixed matching ratio with the Bernoulli model

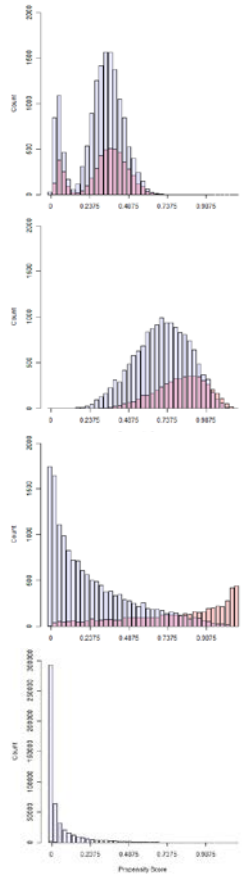
Incidence = 8 per 1000
Q21.0: ventricular septal defect

A

B

C

Active comparator with decreasing overlap in propensity score distributions



D: Unexposed comparator

Matching ratio	Exposed N	Change from 1:1	Referent N	Change from 1:1	Full N	Power for Q21.0 for RR=2
1:1	4999		4,999		9,998	0.59
2:1	4997	0%	9,994	100%	14,991	0.79
3:1					4	0.84
1						0.70
2						0.74
3						0.69
1						0.38
2						0.35
3						0.29
1:1	4892		4,892		9,784	0.65
2:1	4833	-1.2%	9,666	98%	14,499	0.74
3:1	4766	-2.6%	14,298	192%	19,064	0.81

Power improved only when:

a) there was high overlap of propensity scores among groups and enough referent patients

b) there was an unexposed comparator

Not a reliable method for increasing power

Question 3: Sensitivity vs PPV for defining outcomes

Bernoulli model

In each square:

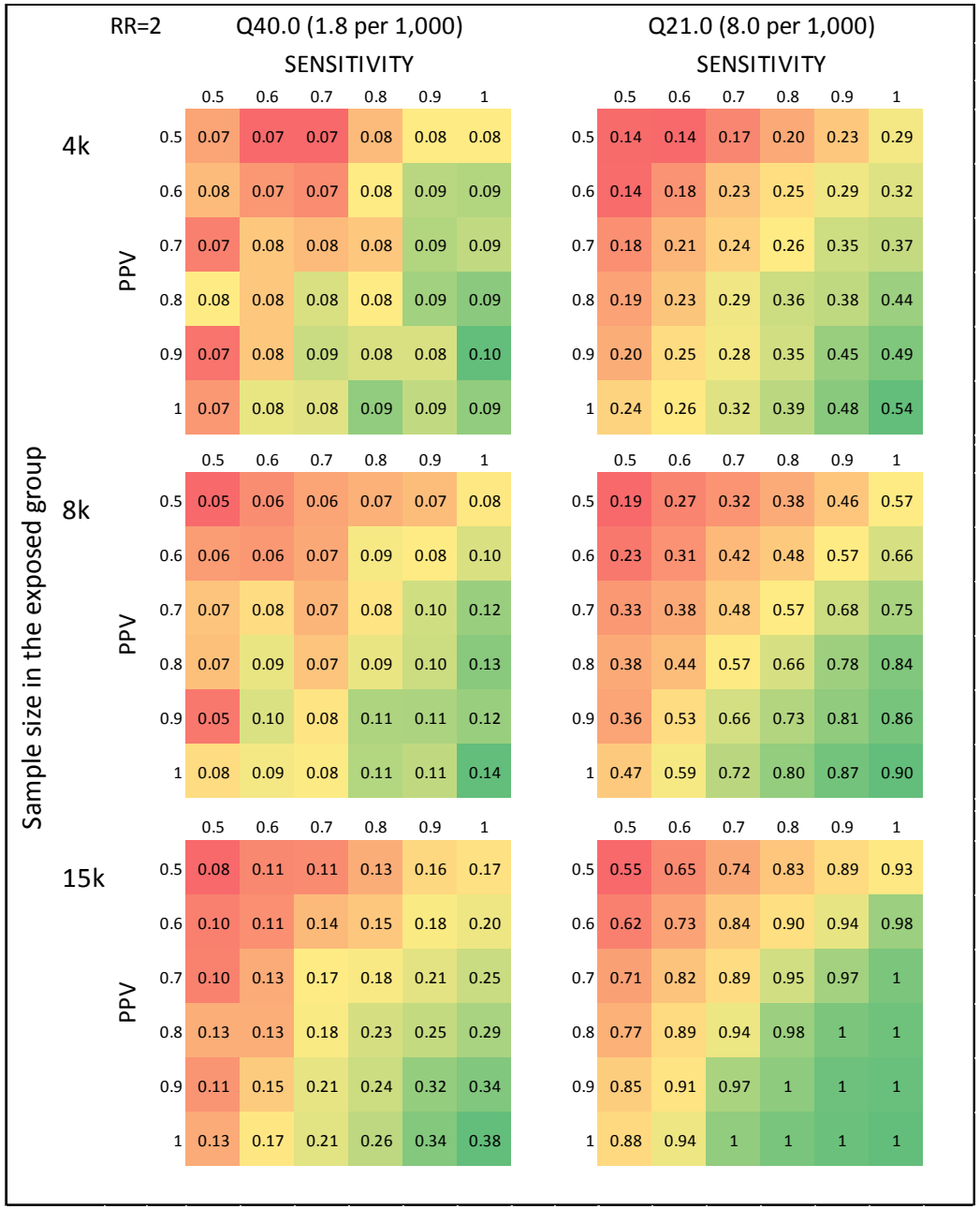
Increasing sensitivity

Increasing PPV

Darker green = greater power

Concentrated on the lower right side, where sensitivity is greater than PPV

A: Bernoulli model



Question 3: Sensitivity vs PPV for defining outcomes

Poisson model

In each square:

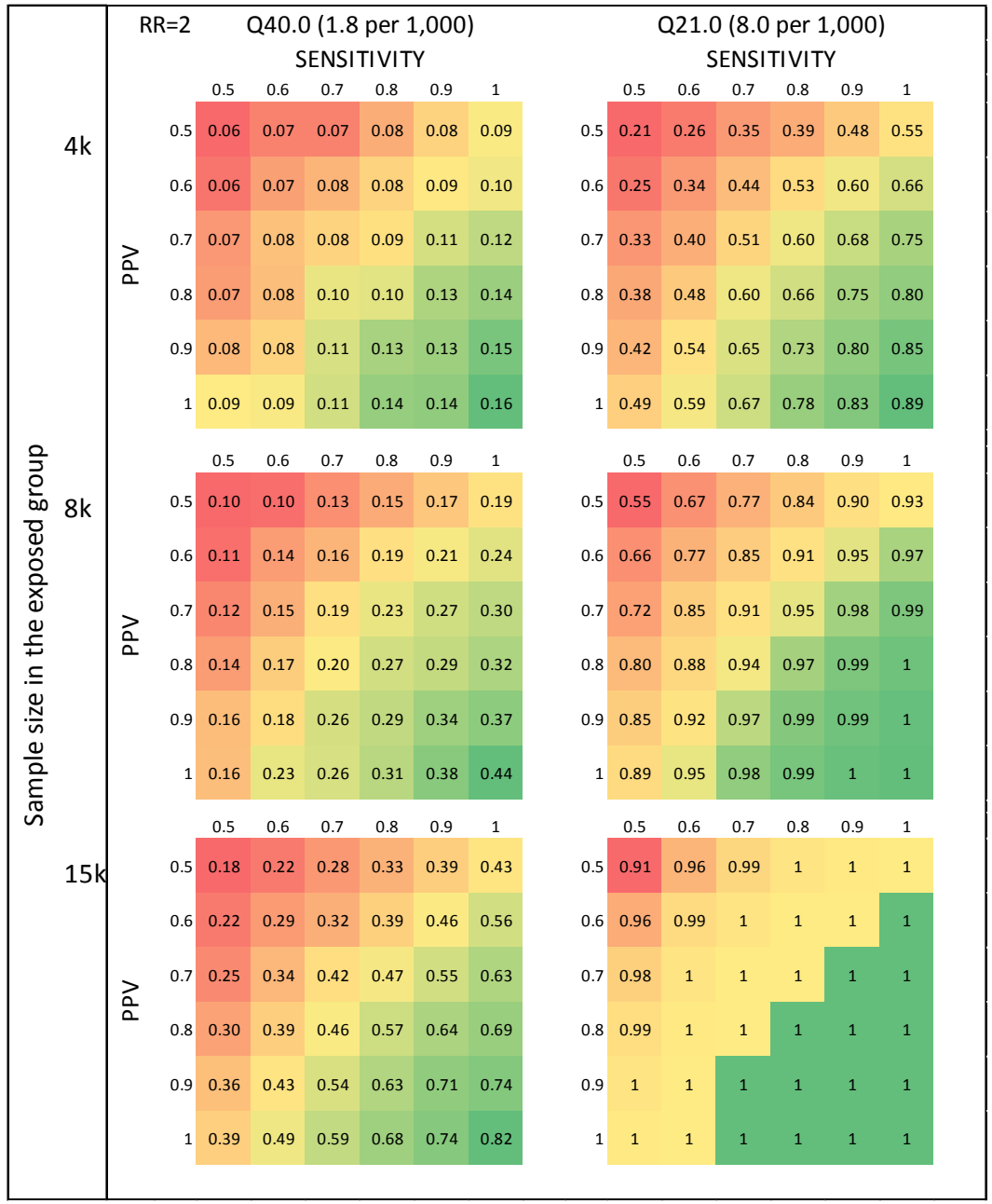
Increasing sensitivity

Increasing PPV

Darker green = greater power

Concentrated on the lower right side, where sensitivity is greater than PPV

B: Poisson model



Question 3: Sensitivity vs PPV for defining outcomes

Poisson model

In each square:

Increasi
sensitiv

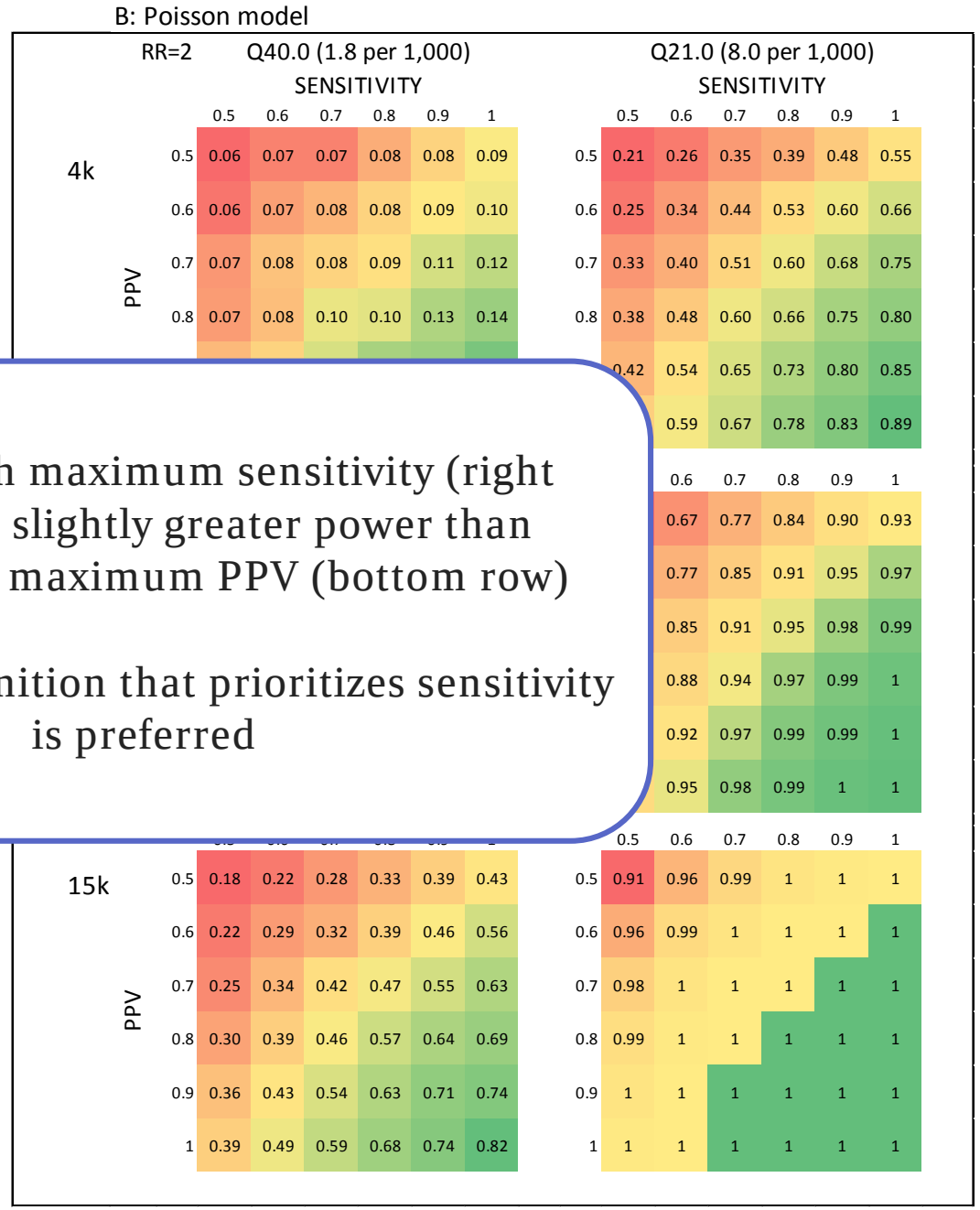
Increasing
PPV

Darker green = greater po

Concentrated on the lower right side, where sensitivity is greater than PPV

Scenarios with maximum sensitivity (right column) had slightly greater power than scenarios with maximum PPV (bottom row)

An outcome definition that prioritizes sensitivity is preferred



Conclusions

- TreeScan is a promising method for use in surveillance of potential adverse infant events following maternal medication exposure during pregnancy
- We recommend using the Poisson model to increase power to observe alerts
  - If less than 4000 exposed pregnancies are available for study, the analysis may be underpowered to detect most alerts
- We attempted to improve power using the Bernoulli method by using N:1 fixed ratio matching, but this proved unreliable as a general strategy
- Our outcome misclassification bias analysis suggests a highly sensitive outcome definition is useful for maintaining power, regardless of TreeScan model used

Thank You

Elizabeth Suarez

Elizabeth_Suarez@harvardpilgrim.org