OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Highlights from research in the Observational Medical Outcomes Partnership

Paul Stang, PhD
Patrick Ryan, PhD
Global Epidemiology
Janssen Research and Development

1

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Observational Medical Outcomes Partnership

- ***Public-Private Research Partnership established to inform the appropriate use of observational healthcare databases for studying the effects of medical products:***
  - Conducting methodological research to empirically evaluate the performance of various analytical methods on their ability to identify true associations and avoid false findings
  - Developing tools and capabilities for transforming, characterizing, and analyzing disparate data sources across the health care delivery spectrum
  - Establishing a shared resource so that the broader research community can collaboratively advance the science

2

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Why Most Databases Were Created

- Payors track expenditures to manage costs
  - Managed care
  - Insurance company
  - Employer
  - Government - some track more than health care
- Clinicians track patients to manage care
  - Mayo Clinic, GPRD, Electronic Medical Records
- Surveys are carried out in the population
  - National Comorbidity Survey, National Household Survey (UK), Whitehall, MONICA, Framingham
- ***The data reflect the underlying healthcare delivery system***
- ***Most were not designed for research***

3

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Types of Health Care Databases

- **Claims data**: MD, goods, prescriptions
- **Electronic medical record**
- **Laboratory/diagnostic testing**
- **Symptom data**: eg, NHANES
- **Facility**: eg, Hospital data (Premier)
- **Cost database**: Prescription AWP
- **Surveillance data**: SEER, CDC, AERS, MAUDE
- **Prospective cohorts**: Framingham
- **Prescription data**: Medco, IMS
- **Cross-sectional studies**: NCS
- **Registries**: Cancer, pregnancy, disease, exposure, devices
- **Linked National Databases**: Sweden, Denmark

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

To understand the data, you must understand….

- Underlying healthcare delivery system
- How 'long' and 'deep' the capture is for a given patient
- Limitations of coding that may affect interpretation
  - Financial incentives in coding in some countries
  - 'Rule-out' diagnoses
- How different methods perform in different databases and populations

5

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

What can we do with longitudinal observational databases?

Descriptive

- Disease natural history
  - What is the background rate of a disease?
  - What are risk factors for developing the disease?
  - Does having the disease impact risk of subsequent outcomes?
- Drug utilization patterns
  - How many patients have exposure to medical product?
  - Amongst patients with exposure, what are common comorbidities and concomitant medications?

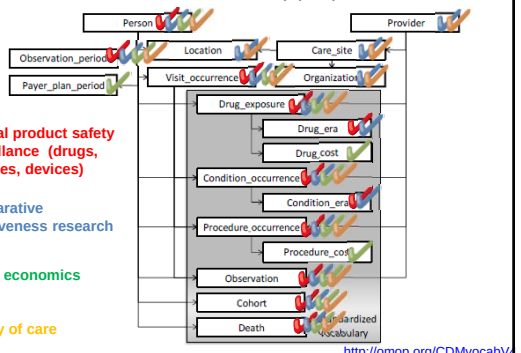
Analytical

- Exploring the effects of medical products
  - Safety signal detection and evaluation
  - Comparative effectiveness research

6

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Prepare data once for research, use it multiple times for many purposes



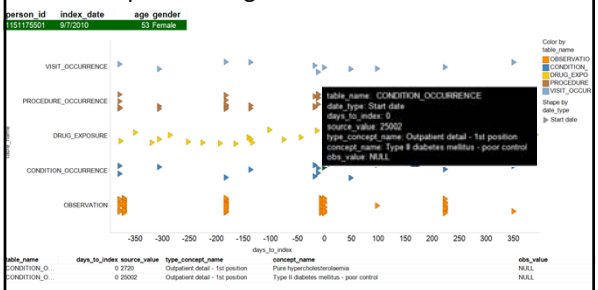
- Medical product safety surveillance (drugs, vaccines, devices)
- Comparative effectiveness research
- Health economics
- Quality of care

http://omop.org/CDMvocabV4

7

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

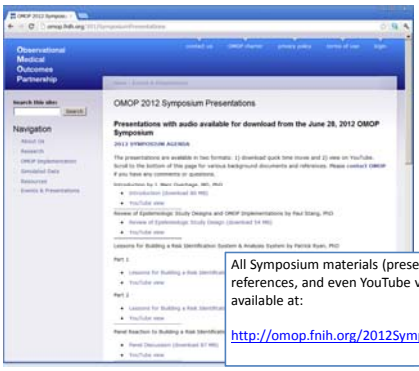
Patient profile: longitudinal record with diabetes



8

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

A brief summary from a long journey



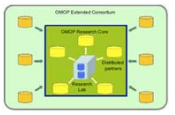
All Symposium materials (presentations, handouts, references, and even YouTube videos) are all publicly available at:

http://omop.fnih.org/2012SymposiumPresentations

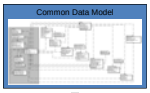
9

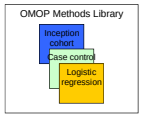
OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

OMOP 2010/2011 Research Experiment



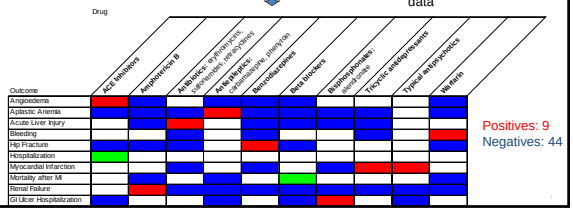
- Open-source
- Standards-based





- 10 data sources
- Claims and EHRs
- 200M+ lives

- 14 methods
- Epidemiology designs
- Statistical approaches adapted for longitudinal data



Positives: 9
Negatives: 44

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

OMOP 2011/2012 Research Agenda

Drug-outcome pairs

	Positives	Negatives
Total	165	234
Myocardial Infarction	36	66
Upper GI Bleed	24	67
Acute Liver Injury	81	37
Acute Renal Failure	24	64

+ EU-ADR replication

- Improve HOI definitions
- Explore false positives

- Evaluate study design decisions (EDDIE)

Methods development

Methods enhancements
- *Multivariate self-controlled case series*

Increased parameterization
- *Case-control, new user cohort designs*

Application of existing tools
- *ICTPD, OS, LGPS, DP*

- Expand CDM for additional use cases

Observational data

Real-world performance: Thomson MarketScan, GE + OMOP Distributed Partners + EU-ADR network

Simulated data: signal → OSIM2

- Strength (RR)
- Type (timing)

11

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Ground truth for OMOP 2011/2012 experiments

isoniazid, fluticasone, indomethacin, clindamycin, ibuprofen, loratadine, sertraline, pioglitazone

	Positive controls	Negative controls	Total
Acute Liver Injury	81	37	118
Acute Myocardial Infarction	36	66	102
Acute Renal Failure	24	64	88
Upper Gastrointestinal Bleeding	24	67	91
Total	165	234	399

Criteria for positive controls:
- Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label
- Drug listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"
- Literature review identified no powered studies with refuting evidence of effect

Criteria for negative controls:
- Event not listed anywhere in any section of active FDA structured product label
- Drug not listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"
- Literature review identified no powered studies with evidence of potential positive association

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Exploring isoniazid and acute liver injury

CMAJ RESEARCH

Adverse events associated with treatment of latent tuberculosis in the general population

Benjamin M. Smith MD, Kevin Schwartzman MD MPH, Gillian Bartlett PhD, Dick Menzies MD MSc

ABSTRACT

Background: Guidelines recommend treatment of latent tuberculosis in patients at increased risk for active tuberculosis. Studies investigating the association of therapy with serious adverse events have not included the entire treated population nor accounted for comorbidities or occurrence of similar events in the untreated general population. Our objective was to estimate the risk of adverse events requiring hospital admission that were associated with therapy for latent tuberculosis infection in the general population.

Methods: Using administrative health data from the province of Quebec, we created a historical cohort of all residents dispensed therapy for latent tuberculosis between 1998 and 2003. Each patient was matched on age, sex and postal region with two untreated residents. The observation period was 18 months (from 6 months before to 12 months after initiation of therapy). The primary outcome was hospital admission for therapy-associated adverse events.

Results: During the period of observation, therapy for latent tuberculosis was dispensed to 9145 residents, of whom 95% started isoniazid and 5% started rifampin. Pretreatment comorbid illness was significantly more common among patients receiving such therapy compared with the matched untreated cohort. Of all patients dispensed therapy, 45 (0.5%) were admitted to hospital for a hepatic event compared with 15 (0.1%) of the untreated patients. For people over age 65 years, the odds of hospital admission for a hepatic event among patients treated for latent tuberculosis infection was significantly greater than among matched untreated people after adjustment for comorbidities (odds ratio [OR] 6.4, 95% CI 2.2–18.3). Excluding patients with comorbid illness, there were two excess admissions to hospital for hepatic events per 100 patients initiating therapy compared with the rate among untreated people over 65 years (95% CI 0.1–3.87).

Interpretation: The risk of adverse events requiring hospital admission increased significantly among patients over 65 years receiving treatment for latent tuberculosis infection. The decision to treat latent tuberculosis infection in elderly patients should be made after careful consideration of risks and benefits.

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Dr. Dick Menzies, dick.menzies@mcgill.ca

CMAJ 2011. DOI:10.1503/cmaj.091824

CMAJ, February 22, 2011, 183(3)

13

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Smith et al. 2011 study design and results

- Data source: Administrative claims from health insurance board of Quebec
- Study design: Cohort
- Exposure: all patients dispensed >=30d of therapy, 180d washout
- Unexposed cohort: 2 patients per exposed, matched by age, gender, and region, with no tuberculosis therapy
- Time-at-risk: Length of exposure + 60 days
- Events: Incident hospital admission for noninfectious or toxic hepatitis
- "Event ratio" estimated with conditional logistic regression
- Covariates: prior hospitalization, Charlson score, comorbidities

Table 2: Event rates and odds ratios for outcomes of interest, by cohort

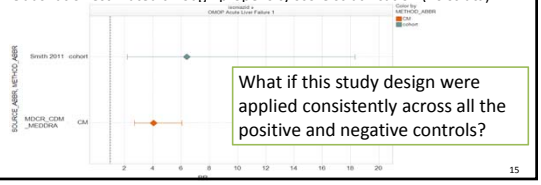
Outcome; age, yr	Crude event rate, events/total (rate per 100 patients)		Event ratio, cohort treated for LTBI v. untreated cohort (95% CI)		
	LTBI therapy cohort	Untreated cohort*	Crude OR†	Adjusted OR‡	Adjusted OR§
Hospital admission for hepatic event of interest¶					
Total	45/9145 (0.5)	15/18 290 (0.1)	6.5 (3.8–11.1)	3.7 (2.0–6.9)	2.7 (1.3–5.6)
≤ 35	5/4523 (0.1)	1/9046 (0.0)	10.0 (1.2–85.6)	NC	NC
36–50	8/2533 (0.3)	7/5066 (0.1)	2.6 (1.0–6.9)	2.0 (0.6–6.9)	1.5 (0.4–5.6)
51–65	10/1232 (0.8)	4/2464 (0.2)	7.0 (2.3–21.3)	2.9 (0.7–12.0)	2.6 (0.4–16.0)
> 65	22/857 (2.6)	3/1714 (0.2)	10.8 (4.2–28.0)	6.4 (2.2–18.3)	4.2 (0.9–11.7)

14

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Revisiting the isoniazid – acute liver injury example

- Data source: MarketScan Medicare Beneficiaries (MDCR)
- Study design: Cohort
- Exposure: all patients dispensed new use of isoniazid, 180d washout
- Unexposed cohort: Patient with indicated diagnosis (e.g. pulmonary tuberculosis) but no exposure to isoniazid; negative control drug referents
- Time-at-risk: Length of exposure + 30 days, censored at incident events
- Covariates: age, sex, index year, Charlson score, number of prior visits, all prior medications, all comorbidities, all priority procedures
- "Odds ratio" estimated through propensity score stratification (20 strata)

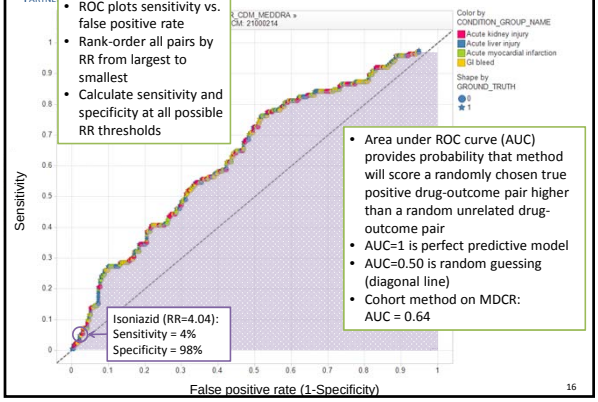


What if this study design were applied consistently across all the positive and negative controls?

15

Receiver Operating Characteristic (ROC) curve

- ROC plots sensitivity vs. false positive rate
- Rank-order all pairs by RR from largest to smallest
- Calculate sensitivity and specificity at all possible RR thresholds

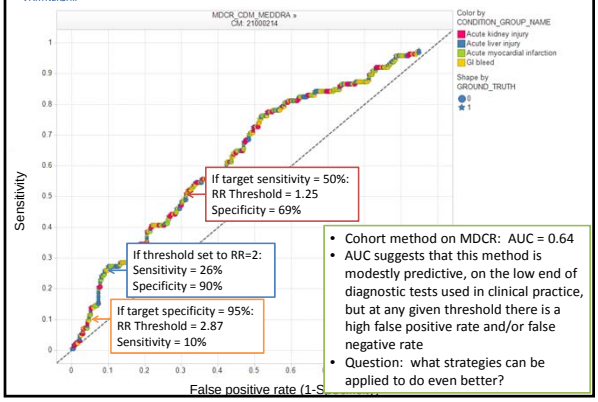


- Area under ROC curve (AUC) provides probability that method will score a randomly chosen true positive drug-outcome pair higher than a random unrelated drug-outcome pair
- AUC=1 is perfect predictive model
- AUC=0.50 is random guessing (diagonal line)
- Cohort method on MDCR: AUC = 0.64

Isoniazid (RR=4.04):
Sensitivity = 4%
Specificity = 98%

16

Setting thresholds from an ROC curve



If target sensitivity = 50%:
RR Threshold = 1.25
Specificity = 69%

If threshold set to RR=2:
Sensitivity = 26%
Specificity = 90%

If target specificity = 95%:
RR Threshold = 2.87
Sensitivity = 10%

- Cohort method on MDCR: AUC = 0.64
- AUC suggests that this method is modestly predictive, on the low end of diagnostic tests used in clinical practice, but at any given threshold there is a high false positive rate and/or false negative rate
- Question: what strategies can be applied to do even better?

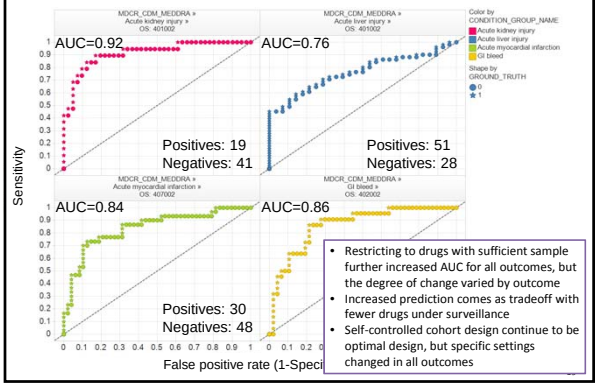
Strategies to improve predictive accuracy

- Stratify results by outcome
- Tailor analysis to outcome
- Restrict to sufficient sample size
- Optimize analysis to the data source

18

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Performance after applying these strategies



- Restricting to drugs with sufficient sample further increased AUC for all outcomes, but the degree of change varied by outcome
- Increased prediction comes as tradeoff with fewer drugs under surveillance
- Self-controlled cohort design continue to be optimal design, but specific settings changed in all outcomes

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

To recap the improvements that could be achieved by following these ideas…

Before: One method applied to all test cases

If sensitivity = 50%:

Outcome	AUC	Threshold	Specificity
All	0.64	1.25	69%

After: Partitioning, tailoring, restriction

If sensitivity = 50%:

Outcome	AUC	Threshold	Specificity
Acute kidney injury	0.92	2.69	95%
Acute liver injury	0.76	1.51	89%
Acute myocardial infarction	0.84	1.59	92%
GI bleed	0.86	1.87	94%

In MDCR

20

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Optimal methods (AUC) by outcome and data source

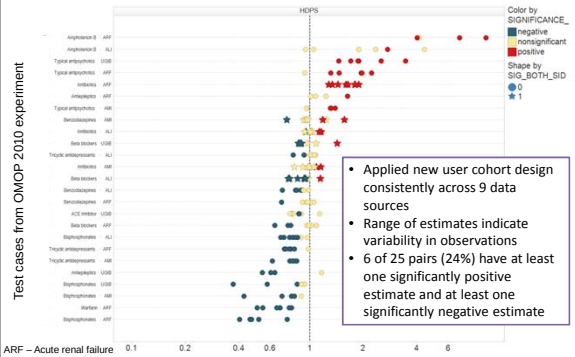
Data source	Acute kidney injury	Acute liver injury	Acute myocardial infarction	GI bleed
MDCR	OS: 401002 (0.92)	OS: 401002 (0.76)	OS: 407002 (0.84)	OS: 402002 (0.86)
CCAE	OS: 404002 (0.89)	OS: 403002 (0.79)	OS: 408013 (0.85)	SCCS: 1931010 (0.82)
MDCD	OS: 408013 (0.82)	OS: 409013 (0.77)	OS: 407004 (0.80)	OS: 401004 (0.87)
MSLR	SCCS: 1939009 (1.00)	OS: 406002 (0.84)	OS: 403002 (0.80)	OS: 403002 (0.83)
GE	SCCS: 1949010 (0.94)	OS: 409002 (0.77)	ICTPD: 3016001 (0.89)	ICTPD: 3034001 (0.89)

- Self-controlled designs are optimal across all outcomes and all sources, but the specific settings are different in each scenario
- AUC > 0.80 in all sources for acute kidney injury, acute MI, and GI bleed
- Acute liver injury has consistently lower predictive accuracy
- No evidence that any data source is consistently better or worse than others

21

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Heterogeneity due to data source



- Applied new user cohort design consistently across 9 data sources
- Range of estimates indicate variability in observations
- 6 of 25 pairs (24%) have at least one significantly positive estimate and at least one significantly negative estimate

ARF – Acute renal failure
ALI – Acute liver injury
AMI – Acute myocardial infarction
UGIB – GI bleed

22

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Takeaways from research toward risk identification

- Performance of different methods
  - Self-contro
- Evaluating a
  - Broader definitions have better coverage and comparable performance to more specific definitions
- Performance across different signal sizes
  - A risk iden
    effects wit
- Data source heterogeneity
  - Substantia
    has value
- Method parameter sensitivity
  - Each meth
    than other
    outcome e

Ryan PB et al, Statistics in Medicine, 2012: "Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership"

Ryan PB et al, Statistics in Biopharmaceutical Research, 2013: "Learning from Epidemiology: Interpreting Observational Database Studies for the Effects of Medical Products"

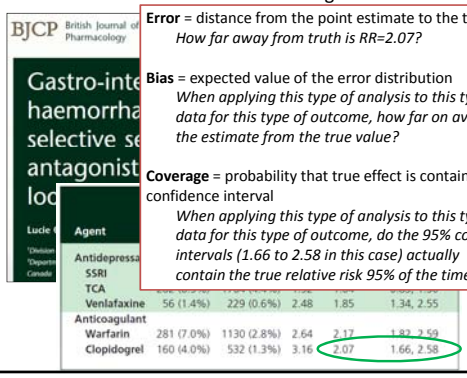
Madigan D et al, American Journal of Epidemiology, 2013: "Evaluating the Impact of Database Heterogeneity on Observational Study Results"

Madigan D et al, Therapeutic Advances in Drug Safety, 2013: "Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies"

23

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Consider a typical observational database study: Exploring clopidogrel and upper gastrointestinal bleeding



p<.001

Error = distance from the point estimate to the true effect
How far away from truth is RR=2.07?

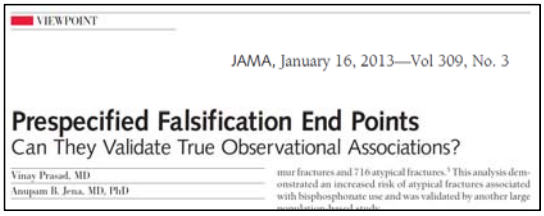
Bias = expected value of the error distribution
When applying this type of analysis to this type of data for this type of outcome, how far on average is the estimate from the true value?

Coverage = probability that true effect is contained within confidence interval
When applying this type of analysis to this type of data for this type of outcome, do the 95% confidence intervals (1.66 to 2.58 in this case) actually contain the true relative risk 95% of the time?

24

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Learning from what's already known



VIEWPOINT

JAMA, January 16, 2013—Vol 309, No. 3

Prespecified Falsification End Points

Can They Validate True Observational Associations?

Vinay Prasad, MD
Anupam B. Jena, MD, PhD

mur fractures and 716 atypical fractures.[3] This analysis demonstrated an increased risk of atypical fractures associated with bisphosphonate use and was validated by another large

- Their recommendation: Use 3-4 negative controls, in addition to target outcome, as a means of assessing the plausibility of an observational analysis result
- Our recommendation: Use a large sample of negative (and positive) controls to empirically measure analysis operating characteristics and use them to calibrate your study finding

25

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

OMOP approach to methodological research

- Develop a standardized implementation of the analysis strategy
  - Study design: Case-control
  - Nesting within indication (unstable angina)
  - Case definition: First episode of upper GI hemorrhage
  - 10 controls per case, matched on age, gender, and index
  - Exposure definition: Length of exposure + 30d
  - Exclusion criteria: <180d of observation before case

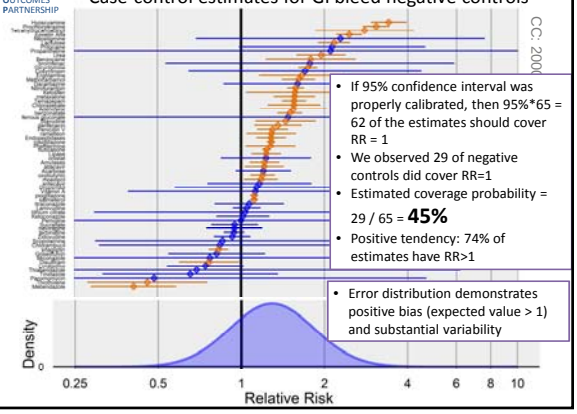
Standard approach yields similar results as initial study:
Opatrny 2008 in CRPD:
2.07 (1.66, 2.58)
OMOP 2012 in CCAE:
1.86 (1.79, 1.93)

- Systematically apply the analysis across a network of databases, consistently for a large sample of positive and negative controls
  - GI Bleeding: **24** positive controls, **67** negative controls
  - Criteria for negative controls:
    - Event not listed anywhere in any section of active FDA structured product label
    - Drug not listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases"
    - Literature review identified no evidence of potential positive association
- Record all effect estimates (RR, CI) from all analysis-database-drug-outcome combinations and summarize analysis*database performance
  - If we assume drugs identified as negative controls truly have no effect on outcome, then we can assume true RR = 1 as a basis for measuring error

26

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Case-control estimates for GI bleed negative controls

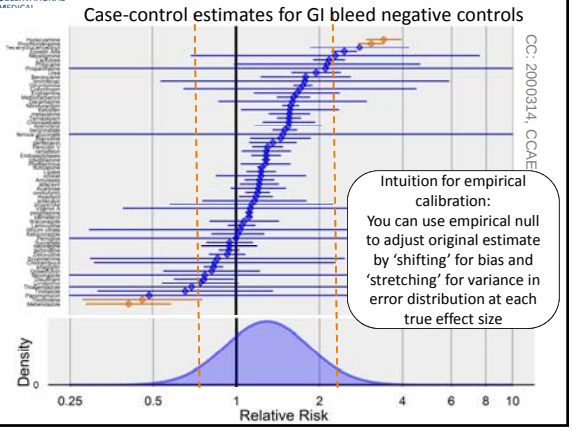


- If 95% confidence interval was properly calibrated, then 95%*65 = 62 of the estimates should cover RR = 1
- We observed 29 of negative controls did cover RR=1
- Estimated coverage probability = 29 / 65 = **45%**
- Positive tendency: 74% of estimates have RR>1

- Error distribution demonstrates positive bias (expected value > 1) and substantial variability

OBSERVATIONAL MEDICAL

Case-control estimates for GI bleed negative controls

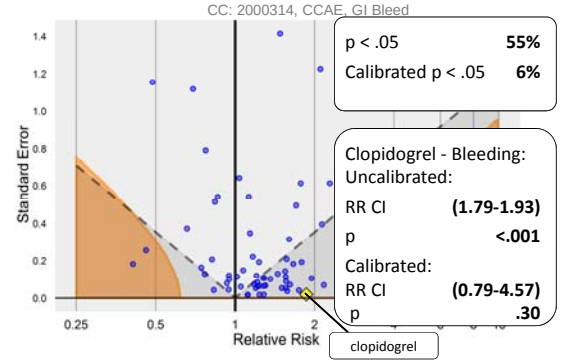


Intuition for empirical calibration:
You can use empirical null to adjust original estimate by 'shifting' for bias and 'stretching' for variance in error distribution at each true effect size

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

p-value calibration plot

CC: 2000314, CCAE, GI Bleed



p < .05	**55%**
Calibrated p < .05	**6%**

Clopidogrel - Bleeding:
Uncalibrated:

RR CI	**(1.79-1.93)**
p	**<.001**

Calibrated:

RR CI	**(0.79-4.57)**
p	**.30**

clopidogrel

29

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Research Article

Statistics in Medicine

Received 12 November 2012, Accepted 3 July 2013, Published online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/sim.5925

Interpreting observational studies: why empirical calibration is needed to correct *p*-values

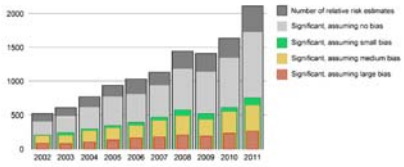


Figure 4. Effect estimates extracted from MEDLINE abstracts of observational studies using healthcare databases, by publication year. The number of estimates reaching statistical significance ($p < 0.05$) is estimated using four assumption on the null distribution: no bias (traditional significance testing): mean = 0, SD = 0; small bias: mean = 0, SD = 0.25; medium bias: mean = 0.25, SD = 0.25; large bias: mean = 0.5, SD = 0.5.

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Performance of Pharmacovigilance Signal-Detection Algorithms for the FDA Adverse Event Reporting System

R Harpaz[1], W DuMouchel[2,3], P LePendu[1], A Bauer-Mehren[1], P Ryan[3,4] and NH Shah[1]

Clinical Pharmacology & Therapeutics (2013); **93** 6, 539–546. doi:10.1038/clpt.2013.24

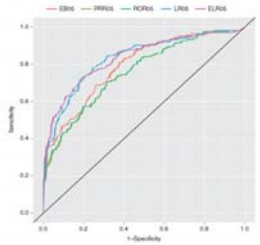


Figure 2 Receiver operating characteristic (ROC) curves for the EBGS, PRRGS, RORGS, LRGS, and ELRGS signal scores. The pattern of containment between the LR-based results and the disproportionality-based results implies that the

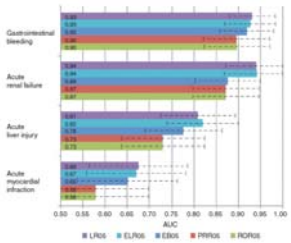


Figure 3 Signal-detection algorithm performance (AUC) classified by event. Error bars reflect AUC 95% confidence intervals. AUC, area under the receiver operating characteristic curve.

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Takeaways from research toward risk identification

- Systematic exploration of negative and positive controls can be used to measure analysis operating characteristics of a signal detection system
  - Errors in observational studies were observed to be differential by analysis design, data source, and outcome
  - Magnitude and direction of bias varied, but all analyses had error distributions far from nominal
  - OMOP results suggest a risk identification system can perform at AUC>0.80
    - Definite potential for useful information, but…
    - Imperfect predictions mean substantial risk of false positive and false negative findings
- Traditional interpretation of p-values and 95% confidence intervals may be misleading in the context of observational database studies
- Empirical calibration is one approach to attempt to account for residual error that should be expected within any observational analysis
- Advancing the science of observational research requires a transparent, reproducible approach to methodology and systematic application

32

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Lessons for building a risk identification system

- Strategies to improve performance:
  - Partition results by outcome
  - Tailor analysis to outcome
  - Restrict to sufficient sample size
  - Optimize analysis to the data source
- OMOP's experimental evidence suggests that following these strategies may yield predictive accuracy at or better than most clinical screening tools used in standard practice

33

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Lessons for building a risk identification system

- Where we are now:
  - Given the diversity in performance and heterogeneity in estimates, we caution against generalizing these results to other outcomes or other data sources
  - If you want to apply risk identification to different outcomes and/or different data sources, we suggest performing an empirical assessment to establish best practice and benchmark performance
- Potential next step:
  - conduct similar experiment for additional 19 outcomes identified by EUADR[1] as high-priority safety issues
  - Once 23 HOIs complete, re-assess whether patterns emerge that would allow generalization to other outcomes

[1]Trifiro et al, PDS 2009

34

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Conclusions

- Using the OMOP approach, a risk identification system can perform at AUC>0.80
- Traditional p-values and confidence intervals require empirical calibration to account for bias in observational studies
- Advancing the science of observational research requires an empirical and reproducible approach to methodology and systematic application

35

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Contact

Paul Stang
pstang@its.jnj.com

Patrick Ryan
pryan4@its.jnj.com

OMOP website: http://omop.org

Look for supplement in Drug Safety, published in Nov2013

OMOP 2013 Symposium, November 5-6, 2013

36