# Active pharmacovigilance using BCPNN-based machine-learning

Douglas McNair MD PhD

*Engineering Fellow & President, Cerner Math Inc*

# Motivation

- **Detect safety signals (negative events, sequences, processes)**
  - With greater sensitivity and specificity than prior art methods
  - With faster time-to-detection than is possible with “spontaneous adverse drug report” regulatory agency systems
- **Affordably, continuously, automatically, scalably, sustainably**
  - Using large observational de-identified, confidentiality-protected, privacy-compliant EHR-derived datawarehouse
- **With appropriate False Discovery Rate (FDR) control**

Cerner

# Cerner at a Glance

- **Founded in 1979, based in Kansas City**
- **Publicly traded since 1986**
- **Largest standalone health care IT company in world**
  - ~10,000 facilities around the world
  - Comprehensive suite of health care solutions & services
  - Contemporary, scalable architecture and cloud platform
- **2012 Revenues of $2.7B**
- **13,500 Associates Worldwide**
  - ~6,000 in Professional / Managed Services
  - Over 2,500 in Engineering / Intellectual Property

Cerner

# Cerner Global Presence


millennium+

*Cerner Millennium* live and localised in 17 countries and four languages: English, Spanish, French and German

**24**

Total countries

More than 10,000 facilities on 6 continents use Cerner Solutions

**400+**

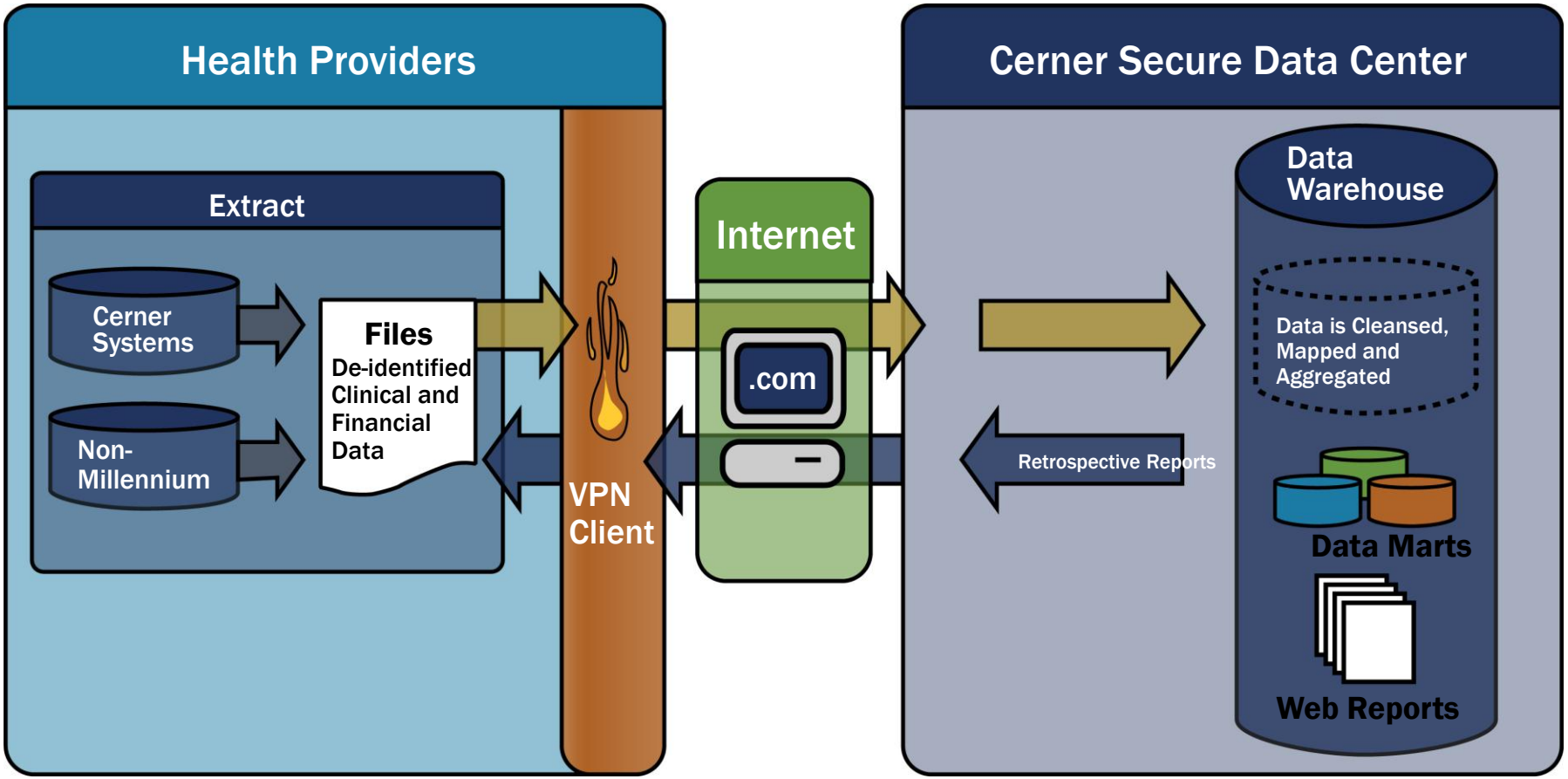
*Millennium* facilities outside the USA

**2,600+**

associates work outside USA, ~20% of company

Cerner

# Cerner Health Facts® De-identified Data Warehouse



# Health Facts® – 516 U.S. institutions are participants

## Total Patient and Episodes

| Patient Type | Patients | Encounters |
|---|---|---|
| Inpatient | 9,453,000 | 12,033,000 |
| Emergency | 20,150,000 | 27,713,000 |
| Outpatient | 70,980,000 | 190,323,000 |

## Total Clinical Data Elements

| Type | Items |
|---|---|
| Medication Orders | 303,856,000 |
| Lab Results | 5,022,000,000 |
| Flowsheet items | 1,109,000,000 |

# Pharmacovigilance prior art has limitations…

- Human beings' finite attention spans and fragmented scope of responsibility and authority with regard to noticing and comprehending the exposures and outcomes to submit ADRs
- Multiple variables and multiple events and outcomes too complex for humans to notice and understand, especially in context of older, sicker patient pop (w/ many comorbid factors)
- Some [undetected] patterns arise longitudinally, outside the range of any one clinician's oversight
- Submitters receive no compensation for reporting, which leads to significant under-reporting and sparsity of ADR submissions [hence, no signal to detect]
- Humans perceive substantial medical malpractice or other risk that would be incurred were they to undertake to report an event, which makes them less likely to report

Cerner

# More limitations of prior art…

- **Some safety event types arise with such low frequency that, although serious or life-threatening, it is improbable that any one practitioner would encounter even one of them during their practice lifetime**
  **➔ disposition to regard them as 'incidental' or idiosyncratic/random events and not report them**
- **Neglect longitudinal patient-level information are neglected, often omitted from regulatory agencies' spontaneous reporting forms**
  **➔ only event-level information is captured**
- **Seriousness of AEs and ADRs is recorded on reporting forms and electronic file formats at the "case" level and not at the "event" level**
  **➔ inability to discriminate 'serious' from 'non-serious' adverse events… pooling these together dilutes all signals [under-detection]**
- **Excessive false-negative (Type II) involving serious adverse events and false-positive (Type I) errors involving non-serious and/or incidental adverse events**
  - Simple PRR > 2.0 gives false-discovery rate (FDR) > 2%

# Yet more limitations of prior art…

- **If there is a physiologic mechanism whereby an agent causes an outcome, one expects to find a quantitative dose-response relationship in the data**
  - ↑ exposures are associated ➔ ↑ frequent, severe, long-lasting, irreversible outcomes or death
- **Unable to discern differences among different strengths or mg/kg dosages or particular dosage-forms or routes of administration**
  - coarse, binomial "exposed"/"not-exposed" status
- **Lack detailed date-timestamped minute-wise timing of when medication-administration exposures occurred and when any events materialized**
  - only able to ascertain is simple statistical 'association', not causality
- **Do not record excursions when physicians modify patients' orders over time**
  - dose-range adjustment of the dose or concomitant medications over a subsequent period of time
  - discontinuation of therapy ("de-challenge")
  - re-prescribing of the drug again over a period of time ("re-challenge")
  - fail to take advantage of this naturally-occurring causality evidence
- **Poor sensitivity and low PPV due to low prevalence of AEs and Rx_AE pairs**

Cerner

# Cerner Datamining EHR PV Process

- **Select important safety 'contexts' in data warehouse**
  - Usually these involve one or more specific conditions, diagnoses, or therapies, or sequences of interventions (protocols, care plans)
  - May involve specific gender, age, race, venue, role, genomic biomarker or other attributes
- **Extract records for the target group, for all exposures**
  - Identify prevalent comorbid diagnoses and other attributes in this pop
  - Identify frequent association-sets of concomitant meds and procedures
  - Extract records for attribute-matched comparator group(s)
- **Assemble counts of outcome events, for all event-types**
- **Enrich with first-pass BCPNN to overall AE prevalence > 5%**
- **Compute BCPNN, MGPS p-values**
- **Compute FDR rates for each statistically significant "exposure(s)-event" pairs/triples, to control Type I error rate**

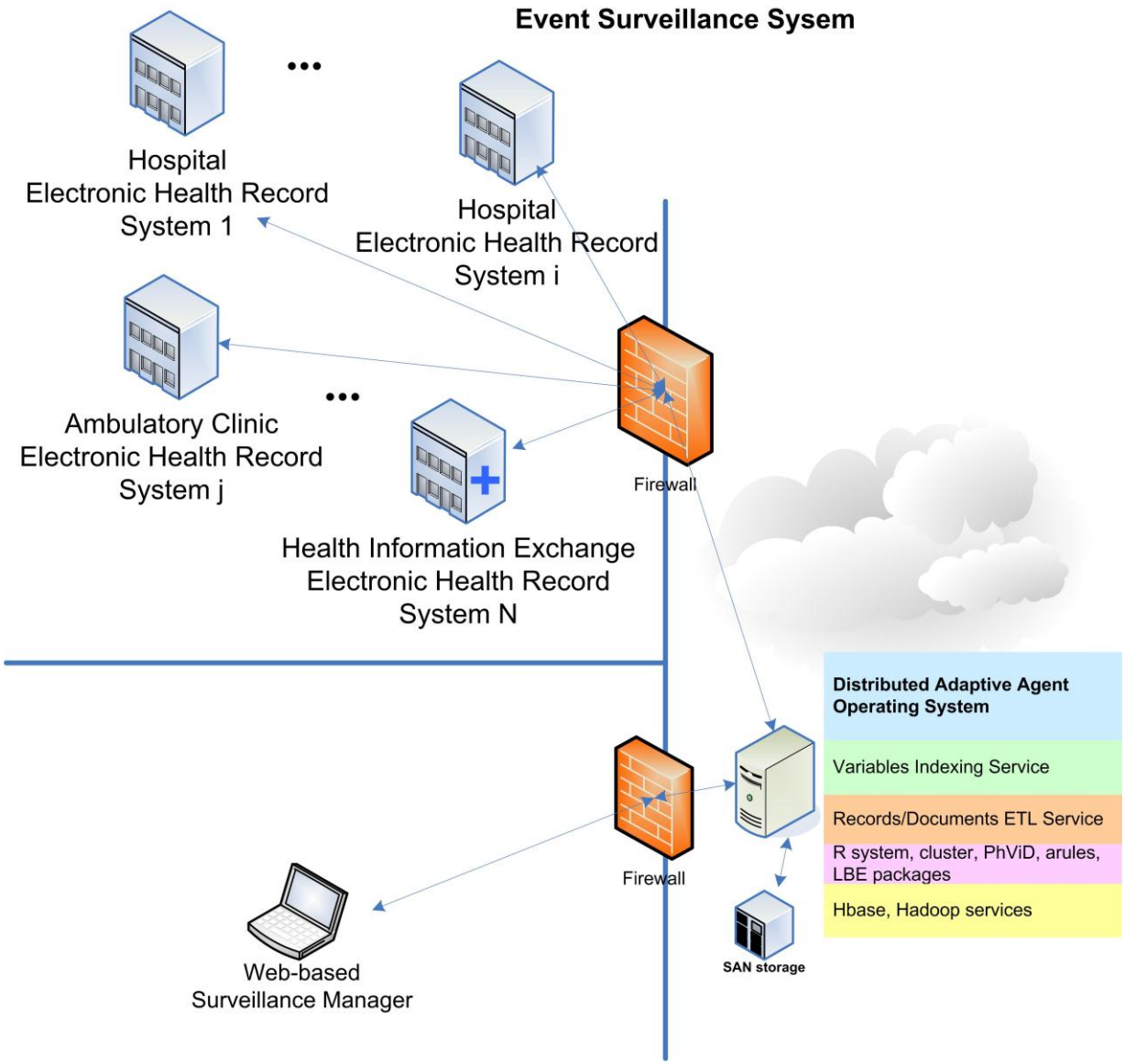
Cerner

# Cerner Datamining EHR PV Features

- Context-sensitive – able to identify safety signals that only arise in subset of overall populations exposed to the therapeutic product or procedure
- Sensitive – able to detect second- or higher-order patterns of multiple variables that would not be noticeable to human observers
- Accurate – FDR and FNR control, to minimize false-positive (Type I) and false-negative (Type II) errors
- Independently validatable – signals detected can be confirmed by re-running the analysis on separate cohorts of EMR-derived records
- Longitudinal – can be used to track the effectiveness of label-change or training or other measures that are undertaken to mitigate risks
- Causality – can use "challenge-dechallenge-rechallenge" and longitudinal sequence of information to establish causal relationship between exposure and event
- Performance / Scalability – sustainable with millions of "exposure-event" pairs

Cerner

# Cerner Parallelized Active PV System



# Example: Drug safety data mining in AMI in-patients

- AMI patients typically have multiple pre-existing comorbid conditions that must receive treatment during the AMI episode, plus any medications associated with PCI/CABG, hospital-acquired pneumonia or other infectious complication subsequent to coronary revascularization procedure
- More than 300 instantiated single-med and multi-med regimens with N > 20
- Lots of diabetes, hypertension, depression, lots of other prevalent conditions
- Misc. conditions that are not all that uncommon (epilepsy, gout, arthritis, etc.)
- Many patients receive more than 10 ADME-concomitant medications during hospital admission for AMI (context-specific “polypharmacy”)
- Extracted 6,699 patients with ‘complete’ data from Health Facts® with EKG-proven AMI coded as ICD-9 410.xx … restricting to relatively healthy population between 35 and 60 years old who survived at least 72 hours (long enough for exposures to meds; long enough for liver function, kidney function, and other lab tests to show significant acute abnormalities, if they arise)
- 131 client institutions, admission dates 01-JAN-2008 thru 31-DEC-2010
- Excluded patients lacking prior encounters or who did not have previous encounters where liver function or kidney function test values were not measured or were not in normal range

# BCPNN screen – In-hospital Mortality

- **Mostly agents that prolong the EKG QT interval → fatal arrhythmias -OR-**
- **Agents that augment risk of other organ-system impairments (e.g. NSAIDs)**

| Medication or Combo or Concomitant Meds | Received Med or Combo (% of total) | Actual Nbr Died (% Rcvd) | Expected Nbr Died | Relative Risk | Act/Exp | p-value** | NNH |
|---|---|---|---|---|---|---|---|
| Ciprofloxacin | 386 (5.8%) | 129 (33.4%) | 28.3 | **5.9** | 1.8 | < 0.0001 | 3.1 |
| Amiodarone +Phenytoin | 39 (0.6%) | 13 (33.3%) | 3.7 | **5.9** | 3.5 | < 0.0002 | 3.6 |
| Ibuprofen +Azithromycin | 25 (0.4%) | 7 (28.0%) | 2.2 | **4.9** | 3.1 | < 0.01 | 4.5 |
| Ibuprofen +Ciprofloxacin | 24 (0.4%) | 6 (25.0%) | 2.5 | **4.4** | 2.4 | < 0.05 | 5.2 |
| Valproic acid | 53 (0.8%) | 12 (22.6%) | 4.8 | **4.0** | 2.5 | < 0.004 | 5.8 |
| Amiodarone +Levofloxacin | 277 (4.1%) | 49 (17.7%) | 26.5 | **3.1** | 1.8 | < 0.0001 | 8.0 |
| Acetaminophen +Phenytoin | 135 (2.0%) | 24 (17.8%) | 8.5 | **3.1** | 2.8 | < 0.0001 | 8.1 |
| Phenytoin | 158 (2.4%) | 27 (17.1%) | 9.7 | **3.0** | 2.8 | < 0.0001 | 8.6 |
| Amiodarone +Azithromycin | 65 (1.0%) | 9 (13.8%) | 4.6 | **2.4** | 1.9 | < 0.05 | 12.1 |
| Acetaminophen +Azithromycin | 277 (4.1%) | 31 (11.2%) | 18.0 | **2.0** | 1.7 | < 0.003 | 17.4 |
| Levofloxacin | 736 (11.0%) | 79 (10.7%) | 51.4 | **1.9** | 1.5 | < 0.0001 | 17.6 |
| Azithromycin | 327 (4.9%) | 33 (10.1%) | 20.8 | **1.8** | 1.6 | < 0.006 | 21.5 |
| Ondansetron +Ciprofloxacin | 245 (3.7%) | 24 (9.8%) | 16.6 | **1.7** | 1.4 | < 0.0001 | 23.4 |
| Fluoxetine | 131 (2.0%) | 11 (8.4%) | 4.9 | **1.5** | 2.2 | < 0.02 | 36.0 |
| Entire cohort | 6,699 (100%) | 380* (5.7%) | N/A | 1.0 | N/A | N/A | N/A |

# BCPNN screen – In-hospital Grade 4 Liver Injury

- Drugs and combos that have known liver toxicity ➔ exacerbated risk in AMI pop
- Drug that is ordinarily low-risk for liver tox ➔ significant risk in AMI pop

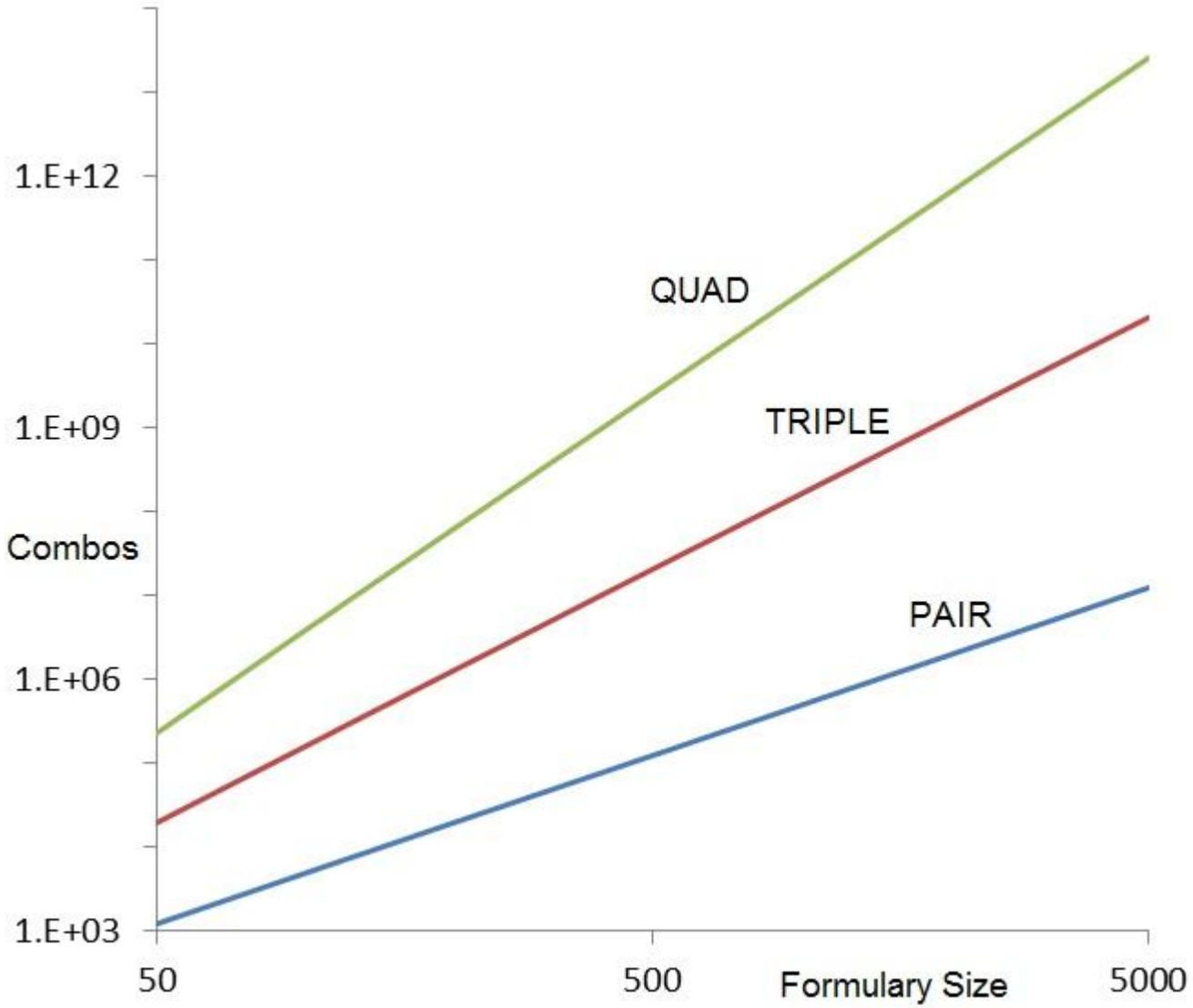
| Medication or Combo or Concomitant Meds | Received Med or Combo (% of total) | Actual Nbr Gr.4 liver inj (% Rcvd) | Expected Nbr Gr. 4 liver inj | Relative Risk | Act/Exp | p-value** | NNH |
|---|---|---|---|---|---|---|---|
| Ibuprofen +Levofloxacin | 58 (0.9%) | 9 (15.5%) | 3.5 | **5.3** | 2.5 | < 0.02 | 7.8 |
| Amiodarone +Levofloxacin | 277 (4.1%) | 35 (12.6%) | 18.1 | **4.3** | 1.9 | < 0.0003 | 9.8 |
| Amiodarone +Ibuprofen | 110 (1.6%) | 12 (10.9%) | 5.8 | **3.8** | 2.1 | < 0.02 | 12.2 |
| Levofloxacin | 736 (11.0%) | 54 (7.3%) | 35.7 | **2.5** | 1.5 | < 0.001 | 19.9 |
| Ondansetron +Levofloxacin | 439 (6.6%) | 32 (7.3%) | 20.4 | **2.5** | 1.6 | < 0.01 | 21.1 |
| Venlafaxine | 111 (1.7%) | 8 (7.2%) | 3.6 | **2.5** | 2.2 | < 0.04 | 22.7 |
| Entire cohort | 6,699 (100%) | 192* (2.9%) | N/A | 1.0 | N/A | N/A | N/A |

# Remarks – PV on Massively-Parallel Processors

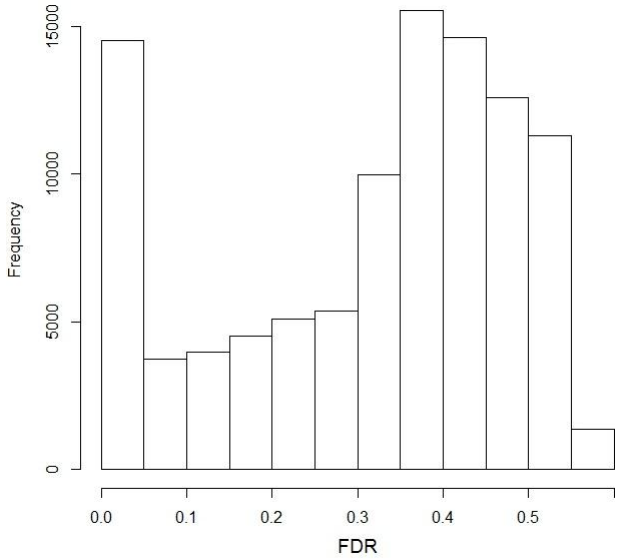
- **Reduction-to-practice successful**
  - Discovered 20 important safety signals not previously recognized
    - *Discovered 14 exposures that were associated with statistically significant ($p < 0.05$) increased risk of in-hospital mortality, elevated up to 5.9-fold above the mortality risk experienced by the cohort as a whole*
    - *Discovered 6 exposures that were associated with up to 5.3-fold increased risk of Grade 4 liver injury while the patients were in-hospital*
  - Performed on 32-node HP Vertica® MPP cluster (8*32=256GB RAM)
  - Utilized Benjamini-Hochberg FDR control
  - Signals confirmed/validated in independent sample
  - Could become the basis for personalized, refined order-sets and plans of care (for AMI treatment, in this example)
- **Parallelization via Hadoop Map-Reduce on private cloud**
  - Implemented for large number of conditions and populations
  - Implemented for high-cardinality formularies (pairs, triples, quads)

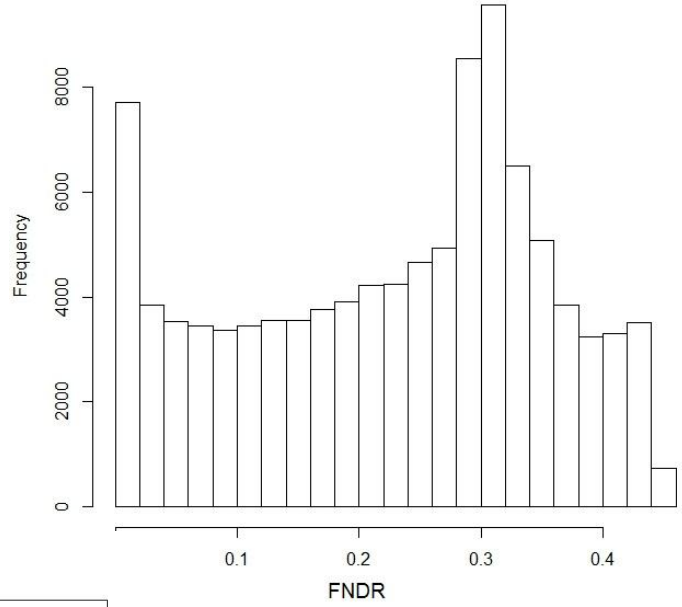
Cerner

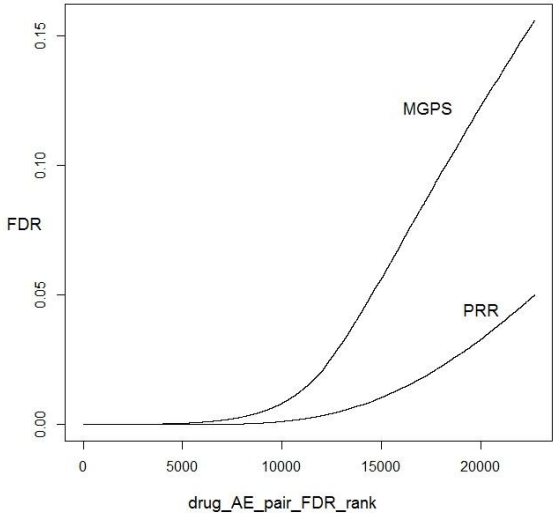
# Combinatorial explosion of concomitant meds combos



# R fdrtool() FDR and FNR distributions - MGPS



# Serial vs. Parallel Signal-Detection

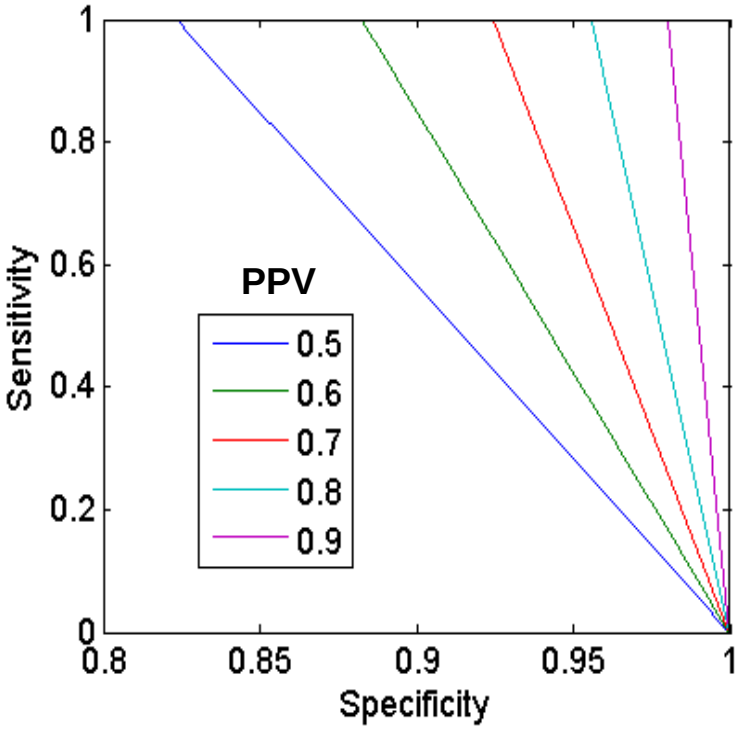
- **Serial / sequential … decimates *sensitivity***
  - net_sensitivity = sens_t1 * sens_t2
  - net_specificity = spec_t1 + [spec_t2 * (1 – spec_t1)]
- **Parallel / simultaneous … decimates *specificity***
  - net_sensitivity = sens_t1 + sens_t2 – (sens_t1 * sens_t2)
  - net_specificity = spec_t1 * spec_t2

| t | PRR | BCPNN | rBCPNN | MGPS | JSS | p(rBCPNN +MGPS) |
|---|---|---|---|---|---|---|
| sens | 49% | 46% | 55% | 27% | 51% | 67% |
| spec | 94% | 99.5% | 89% | 99.99% | 99.7% | 89% |

# 2-Stage Method with FDR Control

$$PPV = \frac{Sensitivity \times Prevalence}{Sensitivity \times Prevalence + (1 - Specificity) \times (1 - Prevalence)}$$

**Enrich putative AEs in mined dataset via Stage-1 screen to a target Prevalence ~ 15%**




**To achieve PPV > 0.50, we need this relationship between sensitivity and specificity:**

$$\frac{Sensitivity}{1 - Specificity} > \frac{1 - Prevalence}{Prevalence}$$

$$\frac{0.60}{1 - 0.90} > \frac{1 - 0.15}{0.15}$$

$$\frac{0.50}{1 - 0.95} > \frac{1 - 0.10}{0.10}$$

**Because of modest prevalence of AEs in as-treated population, to get higher PPV the specificity of the Stage-2 measure (MGPS) dominates.**

**Objective:**

- **Specificity > 0.90**
- **Sensitivity > 0.60**

# Example – BCPNN-enriched to overall AE prevalence > 5%

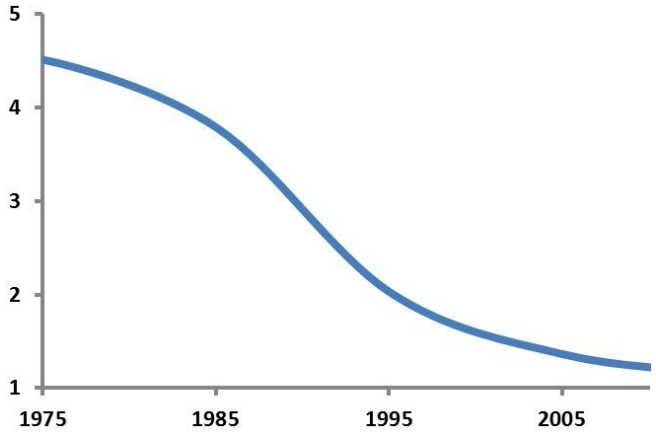
- Exploring 35-60 YO in-patients with non-STEMI acute myocardial infarction + other conditions
- These patterns have never before been detected… not in any regulatory agency or other dataset

| Medication or Combo or Concomitant Meds | Received Med or Combo (% of total) | Actual Nbr Died (% Rcvd) | Expected Nbr Died | Relative Risk | Act/Exp | p-value** | NNH |
|---|---|---|---|---|---|---|---|
| Ciprofloxacin | 386 (5.8%) | 129 (33.4%) | 28.3 | **5.9** | 1.8 | < 0.0001 | 3.1 |
| Amiodarone +Phenytoin | 39 (0.6%) | 13 (33.3%) | 3.7 | **5.9** | 3.5 | < 0.0002 | 3.6 |
| Ibuprofen +Azithromycin | 25 (0.4%) | 7 (28.0%) | 2.2 | **4.9** | 3.1 | < 0.01 | 4.5 |
| Ibuprofen +Ciprofloxacin | 24 (0.4%) | 6 (25.0%) | 2.5 | **4.4** | 2.4 | < 0.05 | 5.2 |
| Valproic acid | 53 (0.8%) | 12 (22.6%) | 4.8 | **4.0** | 2.5 | < 0.004 | 5.8 |
| Amiodarone +Levofloxacin | 277 (4.1%) | 49 (17.7%) | 26.5 | **3.1** | 1.8 | < 0.0001 | 8.0 |
| Acetaminophen +Phenytoin | 135 (2.0%) | 24 (17.8%) | 8.5 | **3.1** | 2.8 | < 0.0001 | 8.1 |
| Phenytoin | 158 (2.4%) | 27 (17.1%) | 9.7 | **3.0** | 2.8 | < 0.0001 | 8.6 |
| Amiodarone +Azithromycin | 65 (1.0%) | 9 (13.8%) | 4.6 | **2.4** | 1.9 | < 0.05 | 12.1 |
| Acetaminophen +Azithromycin | 277 (4.1%) | 31 (11.2%) | 18.0 | **2.0** | 1.7 | < 0.003 | 17.4 |
| Levofloxacin | 736 (11.0%) | 79 (10.7%) | 51.4 | **1.9** | 1.5 | < 0.0001 | 17.6 |
| Azithromycin | 327 (4.9%) | 33 (10.1%) | 20.8 | **1.8** | 1.6 | < 0.006 | 21.5 |
| Ondansetron +Ciprofloxacin | 245 (3.7%) | 24 (9.8%) | 16.6 | **1.7** | 1.4 | < 0.0001 | 23.4 |
| Fluoxetine | 131 (2.0%) | 11 (8.4%) | 4.9 | **1.5** | 2.2 | < 0.02 | 36.0 |
| Entire cohort | 6,699 (100%) | 380* (5.7%) | N/A | 1.0 | N/A | N/A | N/A |

# Remarks – PV for Subgroups and Personalized Medicine

"Drugs tested during the 1970s had a median effectiveness odds-ratio of 4.51 . This dropped to 3.78 in the 1980s, 2.02 in the 1990s, and 1.36 in the 2000s. It is possible that the easiest-to-discover and most-powerful treatments have already been discovered… In terms of drug development, it may make sense to start focusing on more homogeneous phenotypic patient subgroups, such as those that have well-defined biomarkers. It may make sense to give greater priority to improving the tolerability of existing drugs…"

*—Mark Olfson & Steven Marcus, 2013*

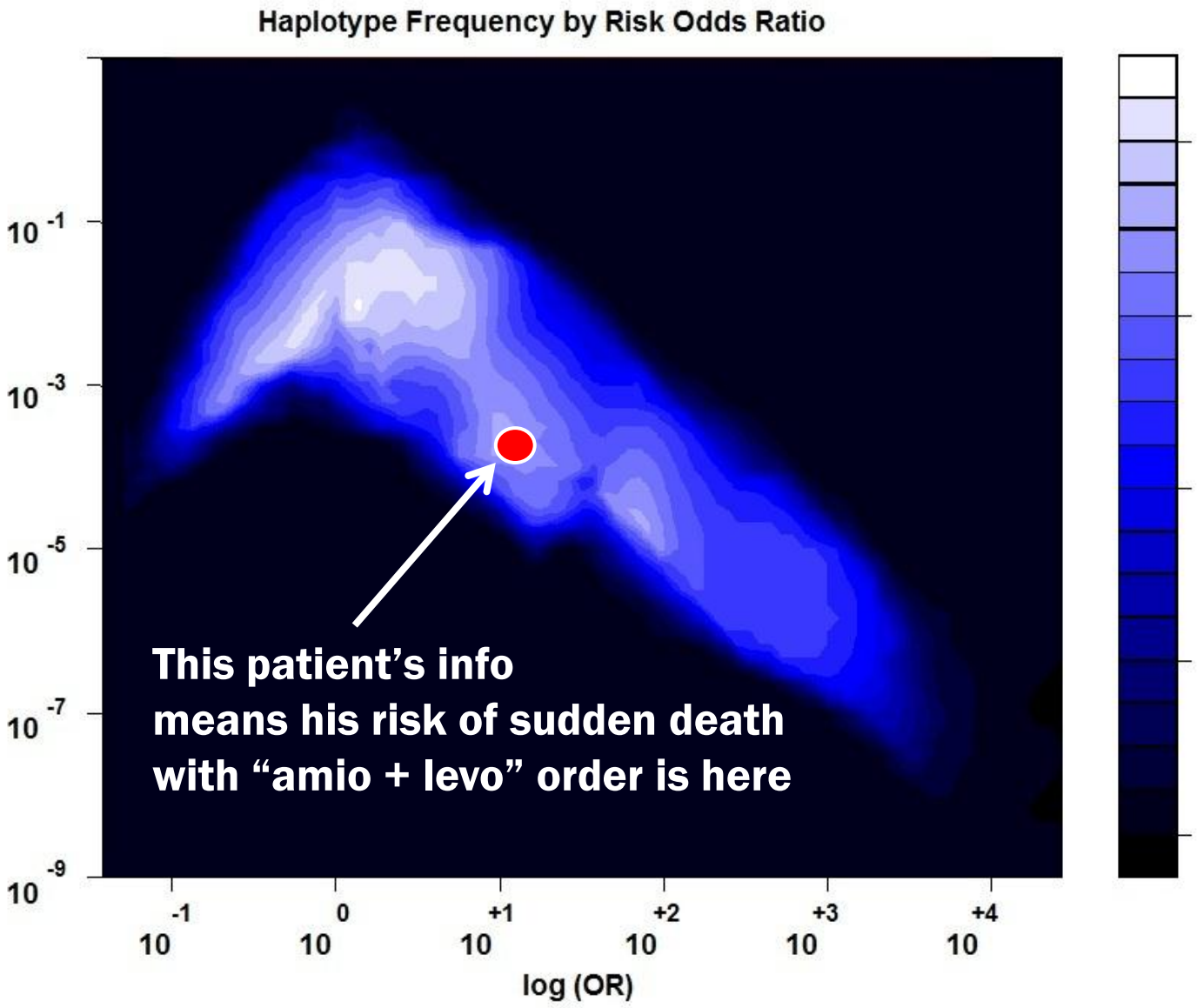



Olfson M, Marcus S. Decline in placebo-controlled trial results suggests new directions for comparative effectiveness research. Health Aff. 2013;32(6):1116-25.
Agres T. Drug Discov & Devel 2013; 16(4):6-7.

# Datamining in Health Facts and Healthe Intent

- **Large cohorts in observational datawarehouses**
  - STEMI and non-STEMI  AMI
  - Mild hypokalemia often
  - QT interval frequently in high-normal range
  - Some prevalent SNP variations in NOD1AP  (not rare ones in KCNH2 and KCNQ1 and SCN5A genes)
  - CYP1A2*1C (slow-metabolizer) vs. CPY1A2*1F (rapid-metabolizer) genotypes
- **Case Scenario**
  - 52 year-old Caucasian male presents with AMI and community-acquired pneumonia, receives PCI with stenting of the LAD coronary artery, and is placed on conventional AMI protocol in CCU.  On Day-2 of hospitalization he develops a pattern of intermittent, hemodynamically unstable ventricular tachycardia and is placed on amiodarone.
  - No other organ system abnormalities, and liver and kidney function tests normal. Mild hypokalemia (potassium 3.1 mEq/L). The patient's EKG shows a depressed ST segment and variable T-wave inversion consistent with acute MI,
  - QTc = 475 msec (normal range between 350-470 msec in adult male).
  - No known history of long-QT syndrome or sudden cardiac death in the family. No QT-prolonging meds on-board.

# Prevalence of Genotype-Phenotype Vectors



# Summary - 1

- **In FDA AERS data, <5% of ever-reported drug-event pairs generate signals (95%LCL EB-adjusted O/E ratio ($EB_{05}$) >= 2) [Szarfman 2002].**
  - But in obs EHR data that are not affected by human reporting sociology, more than 7% of ever-incident drug-event pairs generate signals
  - And in obs EHR data of ever-concomitant triples and quads have a rate over 10%
- **Machine-learning in EHR (VLDB) requires 'vertical' and 'parallelized' computing platforms to deal with terabyte and petabyte scale**
  - Sample size adequate for statistical power, in general, for individual care/prevention use-cases & population use-cases
  - Requires appropriate data rights, de-identification, and consenting for 'secondary-use' observational research (IRB reviewed)

Cerner

# Summary - 2

- **Toolbox:  R packages, MATLAB, Weka**
- **Bayesian confidence-propagation neural network, empirical Bayes, MGPS, James-Stein Shrinker, SVM, K-NN, other**
- **'Ensemble' models combine evidence from multiple models**
  - AdaBoost, Gradient Boost, RandomForest, Alternating Decision Trees
- **Superior specificity and sensitivity, compared to traditional PV**
- **But requires careful FDR control to avoid excessive Type I error**

Cerner

# Summary - 3

- **'Translational medicine' is not just "bench-to-bedside"**
- **It is also translating the other direction: "bedside-to-bench"**
- **Single-method PV suffers either from poor sensitivity or poor specificity or both**
- **Sequential 2-stage 'screen-confirm' has inadequate sensitivity**
- **Parallel BCPNN, MGPS, JSS (and perhaps other) signal detection can jointly optimize sensitivity and specificity**
- **Parallelization enables scale-up to large cohorts accruing in de-identified EHR-derived repositories on modern cloud computing systems**
- **Find true signals quicker!**
  - Manufacturer, Regulator, PBM, Health Plan

Cerner

# References


- Ahmed I, et al. Early detection of pharmacovigilance signals with automated methods based on false discovery rates. Drug Saf. 2012;35:495-506.
- Ahmed I, et al. Pharmacovigilance data mining with methods based on false discovery rates: a comparative simulation study. Clin Pharmacol Ther. 2010;88:492-8.
- Ahmed I, et al. False discovery rate estimation for frequentist pharmacovigilance signal detection methods. Biometrics. 2010;66:301-9.
- Ahmed I, et al. Bayesian pharmacovigilance signal detection methods revisited in a multiple comparison setting. Stat Med. 2009;28:1774-92.
- Almenoff J, et al. Perspectives on the use of data mining in pharmaco-vigilance. Drug Saf 2005;28:981-1007.
- An L, et al. Mining pharmacovigilance data using Bayesian logistic regression with James-Stein type shrinkage estimation. J Biopharm Stat. 2010;20:998-1012.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Roy Statist Soc Ser B 1995;57:289-300.
- Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Statist 2001;29:1165-88.
- Brown J, et al. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. Pharmacoepi & Drug Saf 2007;16:1275-84.
- Constantine N, et al. Improved classification of recent HIV-1 infection by employing a two-stage sensitive/less-sensitive test strategy. J AIDS 2003;32:94–103.
- Held L, Ranyimbo A. A Bayesian approach to estimate and validate the false negative fraction in a two-stage multiple screening test. Methods Inf Med. 2004;43:461-4.
- Hochberg A, Hauben M. Time-to-signal comparison for drug safety data-mining algorithms vs. traditional signaling criteria. Clin Pharmacol & Ther 2009;85:600-6.
- Levine J, et al. Reply: Evaluation of datamining methods for simultaneous and systematic detection of safety signals in large databases. Br. J. Clin. Pharm 2005;61:105-13.
- Matsushita Y, et al. Criteria revision and performance comparison of three methods of signal detection applied to the spontaneous reporting database of a pharmaceutical manufacturer. Drug Saf. 2007;30:715-26.
- Olfson M, Marcus S. Decline in placebo-controlled trial results suggests new directions for comparative effectiveness research. Health Aff. 2013;32:1116-25.
- Park M, et al. A novel algorithm for detection of adverse drug reaction signals using a hospital electronic medical record database. Pharmacoepi & Drug Saf 2011;20:598-607.
- Shu Y, et al. Sequential evaluation of a medical diagnostic test with binary outcomes. Stat Med. 2007;26:4416-27.
- Storey J. The positive false discovery rate: a Bayesian interpretation and the q-value. Ann Statist 2003;31:2013-35.
- Szarfman A, Machado S, O'Neill R. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. Drug Saf. 2002;25:381-92.
- Tubert-Bitter P, et al. Comparison of two drug safety signals in a pharmacovigilance data mining framework. Stat Methods Med Res. 2012 Oct 14. [Epub ahead of print]

# *Thank you!*

Cerner