Common Evidence Model (CEM)

Core Development Team:

Erica A. Voss, Lee Evans, Anthony Sena, & Rich Boyce

Support:

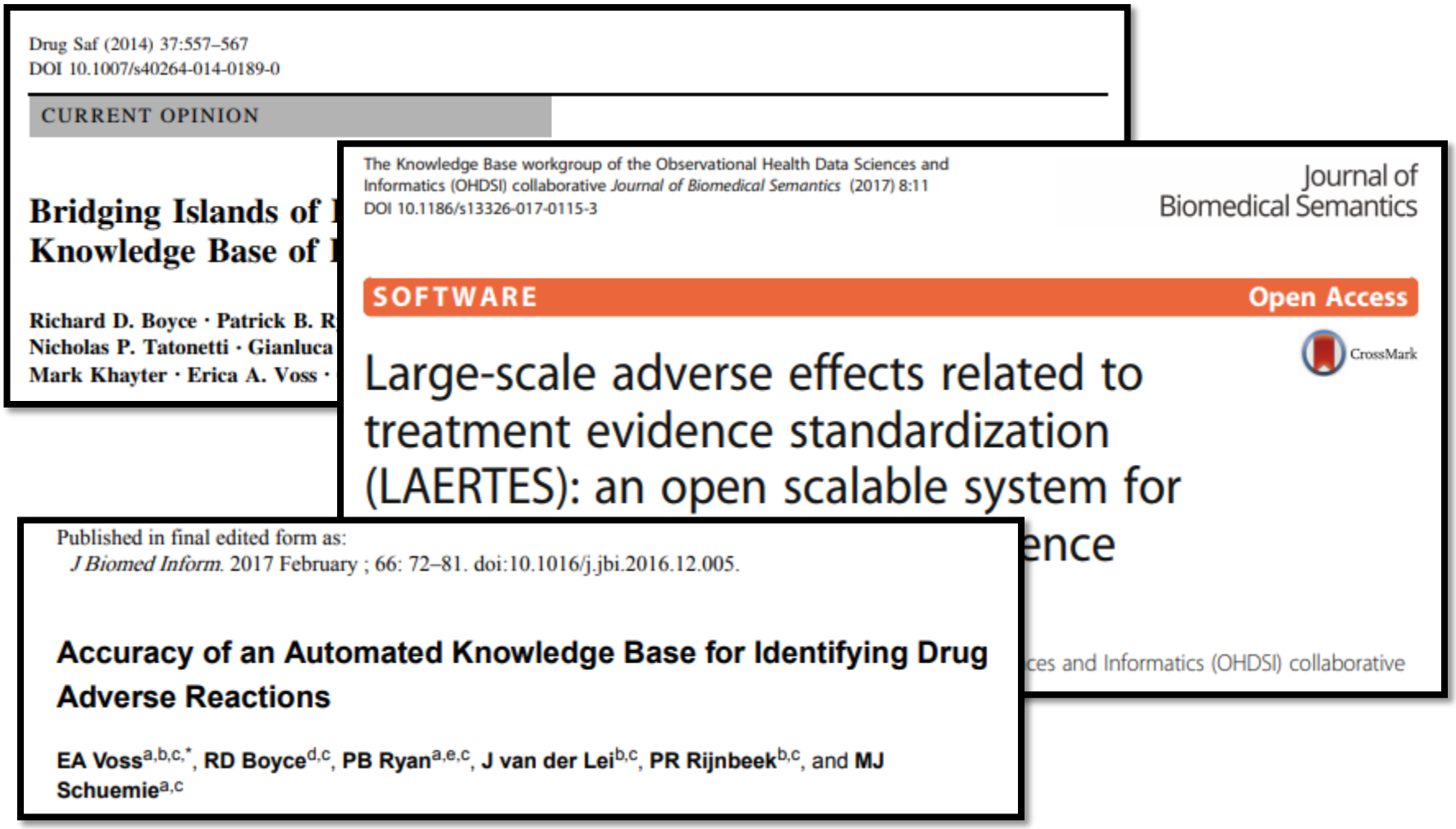
Martijn Schuemie, Peter Rijnbeek, & Patrick Ryan

Topics

- What is the Common Evidence Model?
- Initial Use Case: Finding Negative Controls

What is the Common Evidence Model?

History



Drug Saf (2014) 37:557–567
DOI 10.1007/s40264-014-0189-0

CURRENT OPINION

Bridging Islands of I
Knowledge Base of I

Richard D. Boyce · Patrick B. R
Nicholas P. Tatonetti · Gianluca
Mark Khayter · Erica A. Voss ·

The Knowledge Base workgroup of the Observational Health Data Sciences and Informatics (OHDSI) collaborative *Journal of Biomedical Semantics* (2017) 8:11
DOI 10.1186/s13326-017-0115-3

Journal of Biomedical Semantics

SOFTWARE Open Access

CrossMark

Large-scale adverse effects related to treatment evidence standardization (LAERTES): an open scalable system for
ence

ces and Informatics (OHDSI) collaborative

Published in final edited form as:
J Biomed Inform. 2017 February ; 66: 72–81. doi:10.1016/j.jbi.2016.12.005.

Accuracy of an Automated Knowledge Base for Identifying Drug Adverse Reactions

EA Voss[a,b,c,*], RD Boyce[d,c], PB Ryan[a,e,c], J van der Lei[b,c], PR Rijnbeek[b,c], and MJ Schuemie[a,c]

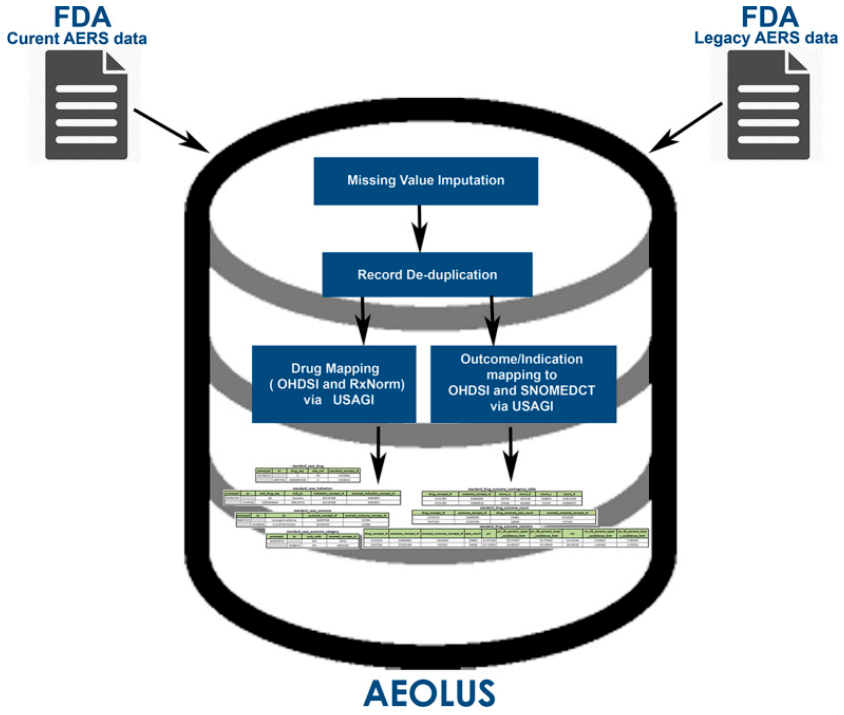
CommonEvidenceModel (CEM)

- Database bridging islands of information together with goal of supporting research of existing evidence about drugs and outcomes of interest.
- CEM replaces LAERTES
- CEM pulls together and standardizes the following types of data:

Spontaneous Reports

- U.S. FDA's Adverse Event Reporting System (FAERS)
- "Curated and standardized version of **FAERS** removing duplicate case records, applying standardized vocabularies with drug names mapped to RxNorm concepts and outcomes mapped to SNOMED-CT, and pre-computed summary statistics about drug-outcome relationships for general consumption."

AEOLUS

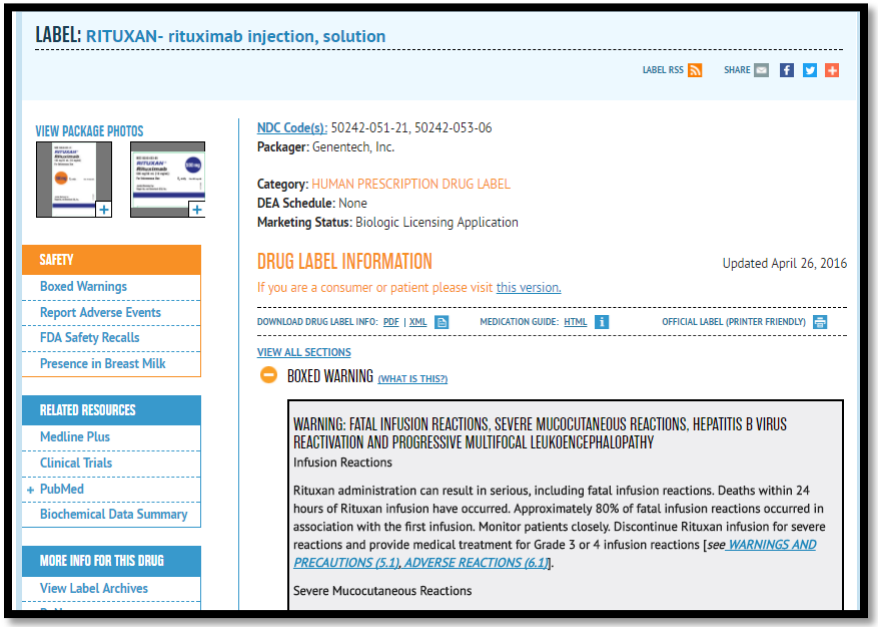


Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH. A curated and standardized adverse drug event resource to accelerate drug safety research. Sci Data. 2016 May 10;3:160026. doi: 10.1038/sdata.2016.26. PubMed PMID: 27193236; PubMed Central PMCID: PMC4872271.

Product Labels

- SPLICER, a tool that reads and parses United States Structured Product Labels (SPLs) for drugs and HOIs in the sections “Adverse Drug Reactions” or “Postmarketing”

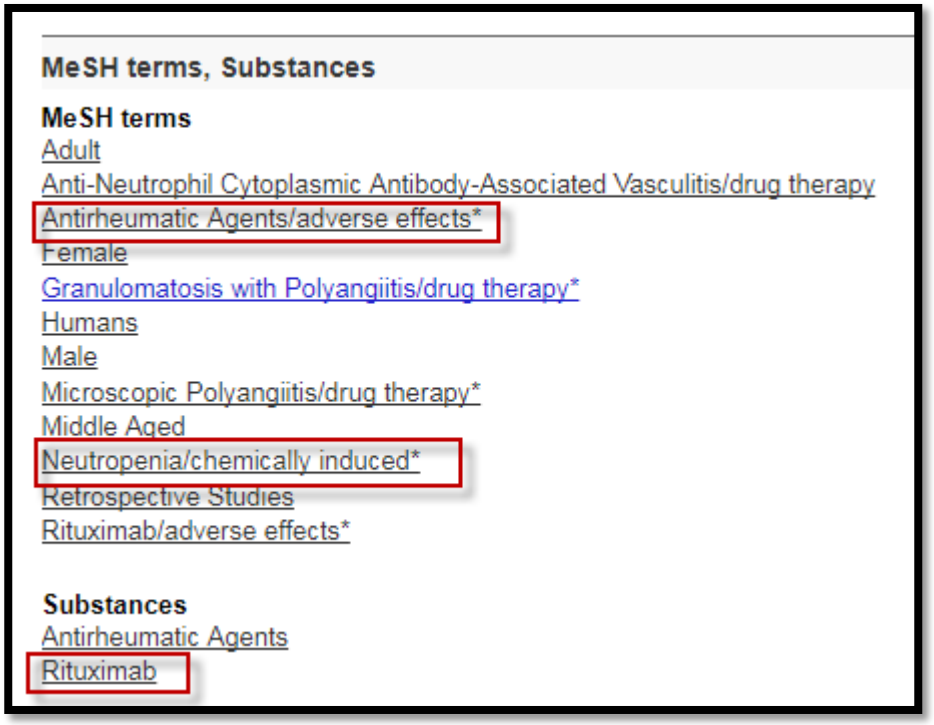
- SPLICER already utilizes the OMOP Vocabulary and maps drugs to RxNorm and HOIs to MedDRA terms.



Duke J, Friedlin J, Li X. Consistency in the safety labeling of bioequivalent medications. Pharmacoepidemiol Drug Saf. 2013 Mar;22(3):294-301. doi: 10.1002/pds.3351. Epub 2012 Oct 8. PubMed PMID: 23042584.

Published Literature

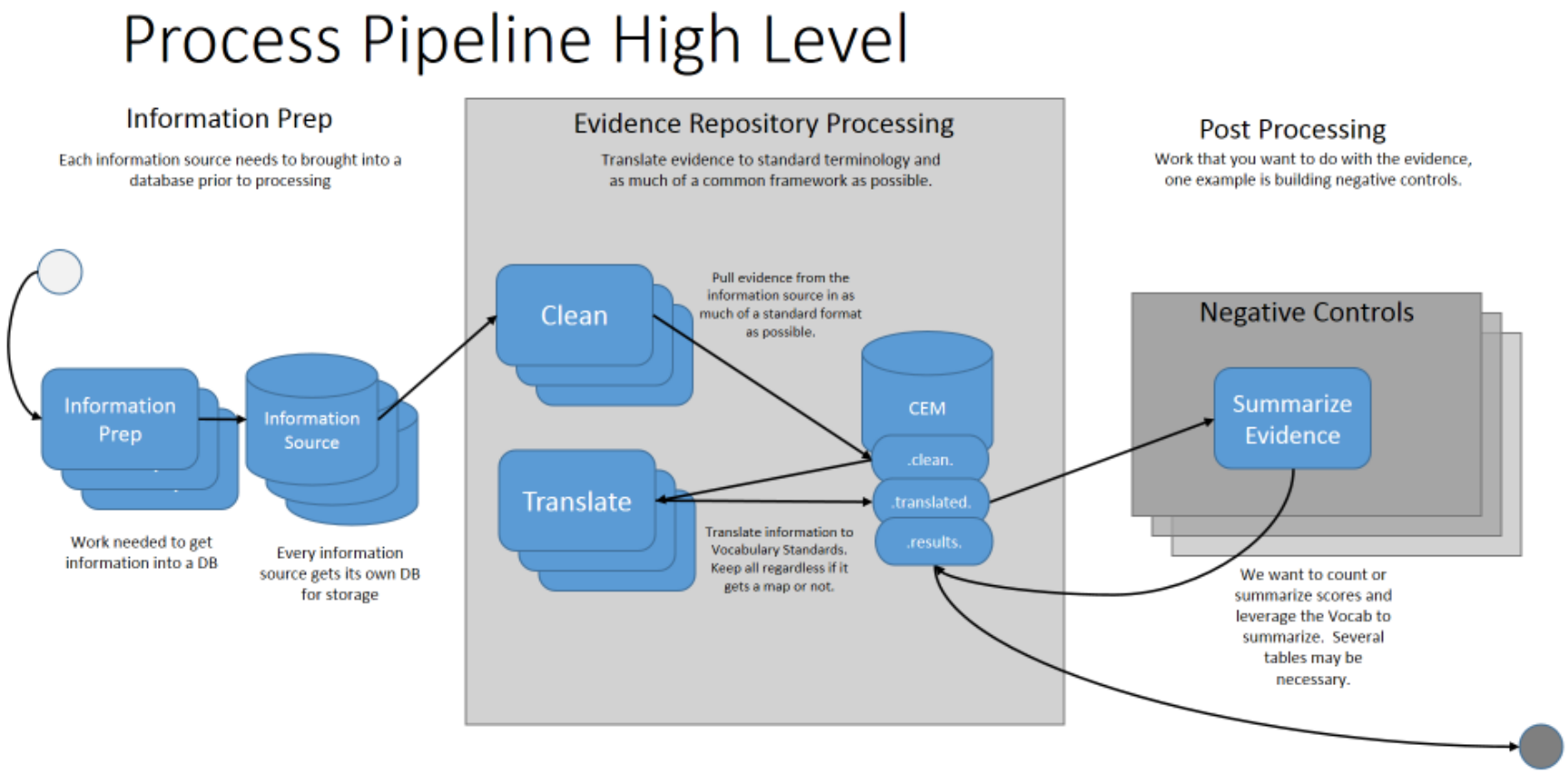
- We have several methods that access data from published literature:
  - Using **MeSH tagged publications from Medline** looked for adverse drug reactions based on the co-occurrence of a drug and an adverse event on the same citation
  - Co-occurrence of a drug and condition MeSH tag or found in the Title of Abstract of a publication. Leverages **PubMed API**
  - **Semantic Medline DB** uses natural language processing to extract semantic predictions from titles and text. CEM pulls those in an "adverse event" relationship



Winnenburg R, Sorbello A, Ripple A, Harpaz R, Tonning J, Szarfman A, Francis H, Bodenreider O. Leveraging MEDLINE indexing for pharmacovigilance - Inherent limitations and mitigation strategies. J Biomed Inform. 2015 Oct;57:425-35. doi: 10.1016/j.jbi.2015.08.022. Epub 2015 Sep 2. PubMed PMID: 26342964; PubMed Central PMCID: PMC4775467.

Kilicoglu H, Rosemblat G, Fiszman M, Rindflesch TC. Constructing a semantic predication gold standard from the biomedical literature. BMC Bioinformatics. 2011 Dec 20;12:486. doi: 10.1186/1471-2105-12-486. PubMed PMID: 22185221; PubMed Central PMCID: PMC3281188

Process Pipeline



CEM_UNIFIED

Column	Description
ID	Unique identifier for each row
CONCEPT_ID_1	Each row represents a pair of concepts, this is the first of the pair represented by an OMOP Concept ID
SOURCE_CODE_1	For CONCEPT_ID_1 what was the original code provided by the source
SOURCE_CODE_TYPE_1	For SOURCE_CODE_1, what was the type of code given by the source (e.g. MeSH)
CONCEPT_ID_2	Each row represents a pair of concepts, this is the second of the pair represented by an OMOP Concept ID
SOURCE_CODE_2	For CONCEPT_ID_2 what was the original code provided by the source
SOURCE_CODE_TYPE_2	For SOURCE_CODE_2, what was the type of code given by the source (e.g. MeSH)
SOURCE_ID	Identifier for the source where this row of evidence came from
EVIDENCE_TYPE	Sometimes the source will provide additional qualifiers of the evidence (e.g. a publication is of type "Clinical Trial" or the stat provided is a PRR score)
RELATIONSHIP_ID	Description of the type of relationship between the two concepts for the row
STATISTIC_VALUE	The evidence provided by the source
STATISTIC_VALUE_TYPE	Describes what the STATISTIC_VALUE is
UNIQUE_IDENTIFIER	Some rows of evidence have a unique identifier (e.g. PMID), that is listed here
UNIQUE_IDENTIFIER_TYPE	Description of what the UNIQUE_IDENTIFIER is
COUNT_HOW	Describes how to use the evidence in aggregate

CEM_UNIFIED

- Currently the evidence that is populated in the CEM_UNIFIED table includes:
  - Label data from SPLICER
  - Information from Medline using Winnenburg approach
  - Information from Medline using Avillach approach
  - Adverse event information from SemMedDB
  - European Product Label adverse event information
  - Spontaneous Report information from AEOLUS
  - Co-occurrence of terms found from PubMed

- Subset of information found in CEM_UNIFIED

concept_id_1 integer	concept_name_1 character varying(255)	concept_id_2 integer	concept_name_2 character varying(255)	source_id character varying(50)	unique_identifier character varying(2000)
1314273	rituximab	140352	Acute myeloid leukemia, disease	medline avillach	28419413
1314273	rituximab	140352	Acute myeloid leukemia, disease	medline avillach	28542862
1314273	rituximab	432574	Diffuse non-Hodgkin's lymphoma, large	medline avillach	25429725
1314273	rituximab	4344166	Adult onset Still's disease	medline avillach	29076903
1314273	rituximab	4344166	Adult onset Still's disease	medline avillach	29076903
1314273	rituximab	4098292	Antiphospholipid syndrome	medline avillach	21924936
1314273	rituximab	432310	Cryptococcal meningitis	medline avillach	21700684
1314273	rituximab	432310	Cryptococcal meningitis	medline avillach	22924391

Can I Use CEM?

- CEM is consumed similarly to the Vocabulary, nothing to run yourself, just consume:
  - Export of summary table CEM_UNIFIED
  - WebAPI Calls

- All code is shared: https://github.com/OHDSI/CommonEvidenceModel

- Working Group: Pharmacovigilance Evidence Investigation Workgroup http://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:kb-wg

Initial Use Case: Finding Negative Controls

Spoiler

CEM is a huge improvement over manually finding negative controls, but process still requires thoughtful review!

Motivation

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

Martijn J. Schuemie,[a,b*†] Patrick B. Ryan,[b,c] William DuMouchel,[b,d] Marc A. Suchard[b,e] and David Madigan[b,f]

Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. Stat Med. 2014 Jan 30;33(2):209-18. doi: 10.1002/sim.5925. Epub 2013 Jul 30. PubMed PMID: 23900808; PubMed Central PMCID: PMC4285234.

Motivation

- "observational studies are **more vulnerable than RCTs to systematic error such as bias and confounding**."

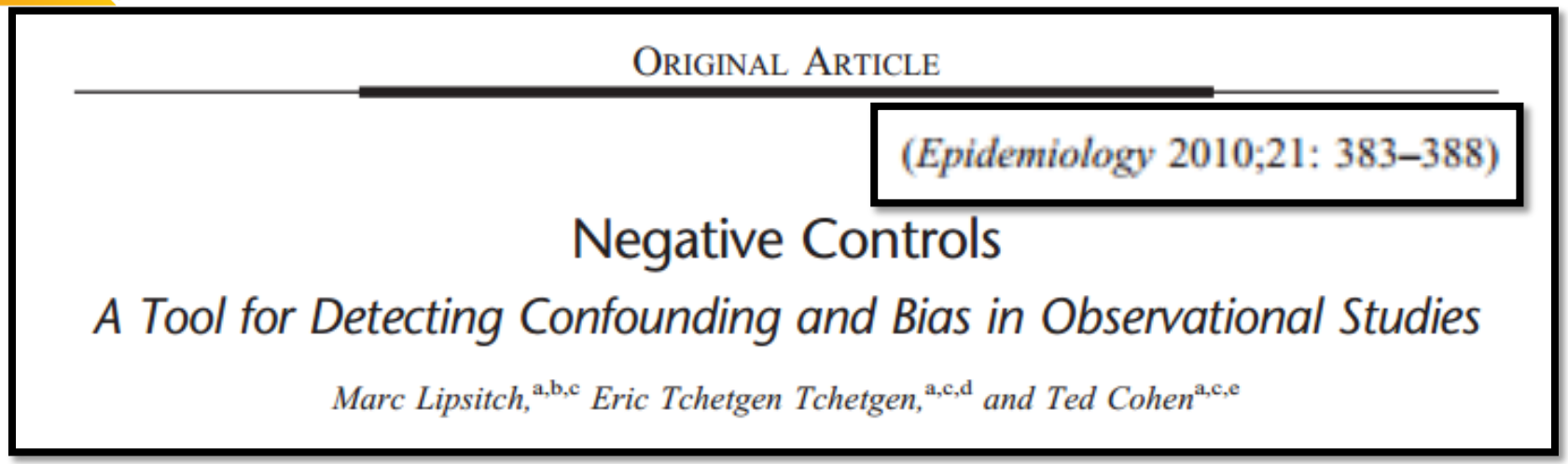
- "Although we believe that most researchers are aware of the fact that traditional p-value calculations do not adequately take systematic error into account, likely because of a lack of a better alternative, the notion of statistical significance based on the **traditional p-value is widespread in the medical literature**."

- "Using **negative controls to empirically estimate the bias** in a study provides a straightforward approach of interpreting the outcome of a study. The **observed null distribution incorporates most forms of bias**, including residual confounding, misclassification, and selection bias. The **error distribution resulting from this bias** (which does not depend on sample size) can be added to the random error distribution (which is based on sample size) **to produce a single intuitive value: the calibrated p-value**."

- "We recommend that **observational studies always include negative controls** to derive an empirical null distribution and use these to compute calibrated p-values."

Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. Stat Med. 2014 Jan 30;33(2):209-18. doi: 10.1002/sim.5925. Epub 2013 Jul 30. PubMed PMID: 23900808; PubMed Central PMCID: PMC4285234.

What are Negative Controls?



ORIGINAL ARTICLE

(*Epidemiology* 2010;21: 383–388)

Negative Controls

A Tool for Detecting Confounding and Bias in Observational Studies

Marc Lipsitch,[a,b,c] *Eric Tchetgen Tchetgen,*[a,c,d] *and Ted Cohen*[a,c,e]

- Biologists employ "negative controls" as a means of ruling out possible noncausal interpretations of their results
- Negative controls are **either exposures or outcomes** that when used to "**repeat the experiment under conditions which it is expected to produce a null result and verify that it does indeed produce a null result**"
- The essential purpose of a negative control is to reproduce a condition that cannot involve the hypothesized causal mechanism but is very likely to involve the same sources of bias that may have been present in the original association

Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology. 2010 May; 21(3): 383-8. doi: 10.1097/EDE.0b013e3181d61eeb. Erratum in: Epidemiology. 2010 Jul;21(4):589. PubMed PMID: 20335814; PubMed Central PMCID: PMC3053408.

Causal Diagrams

- In order to select negative controls, we need to understand the relationship between the exposure and outcome of interest

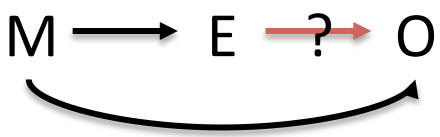
- **Causal diagrams** visually encode assumptions about causal relations

- In today's diagrams:
  - E = Exposure
  - O = Outcome
  - N = Negative Control
  - M = Measured/Unmeasured Variable

Rothman, K. J., et al. (2008). "Modern epidemiology. 3rd." Philadephia: Lippincott Williams & Wilkins.

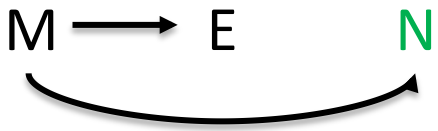
Good Negative Controls

Question of Interest



We are trying to understand the exposure of a drug on some outcome

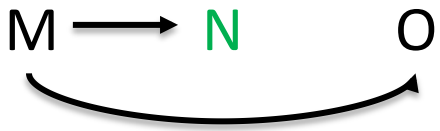
Proper Negative Control Outcome



When finding negative control outcomes, we want something that:

- E does not cause
- has the same causal associations to other variables either measured or not

Proper Negative Control Exposure



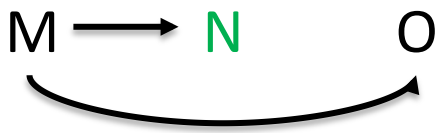
When finding negative control exposures, we want something that:

- does not cause O
- has the same causal associations to other variables either measured or not

E = Exposure; 0 = Outcome; N = Negative Control, M = Measured/Unmeasured Variable

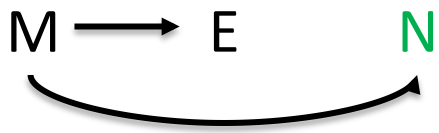
Good Negative Controls

Proper
Negative Control Exposure



Exposure Does Not Cause the Outcome

Proper
Negative Control Outcome

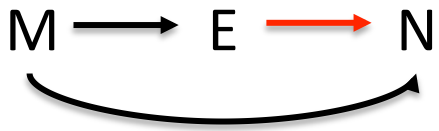


Ideally, Similar Confounding Relationships to Measured and Unmeasured Variables

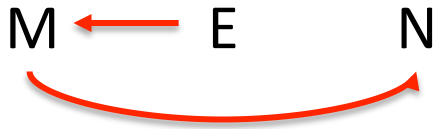
Let's focus on Negative Control Outcomes

What Relationships Make a Bad Negative Control Outcomes?

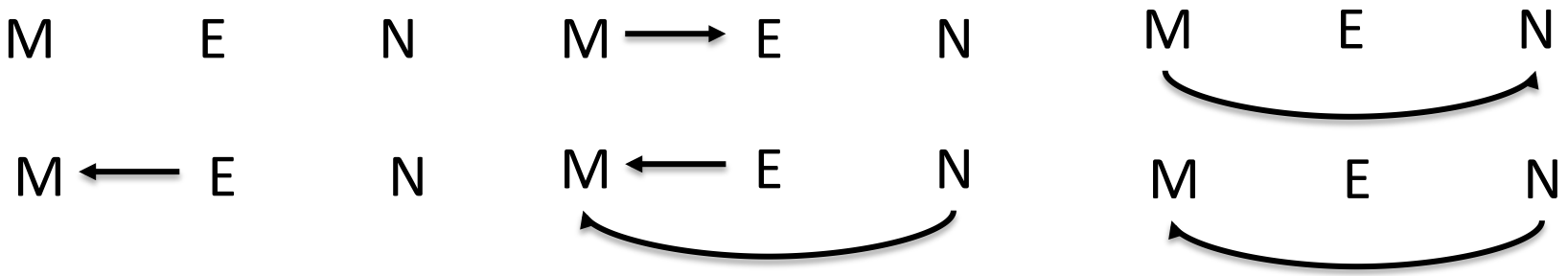
Exposure Causes/Prevents the Outcome



Exposure Causes/Prevents a variable that causes the Outcome



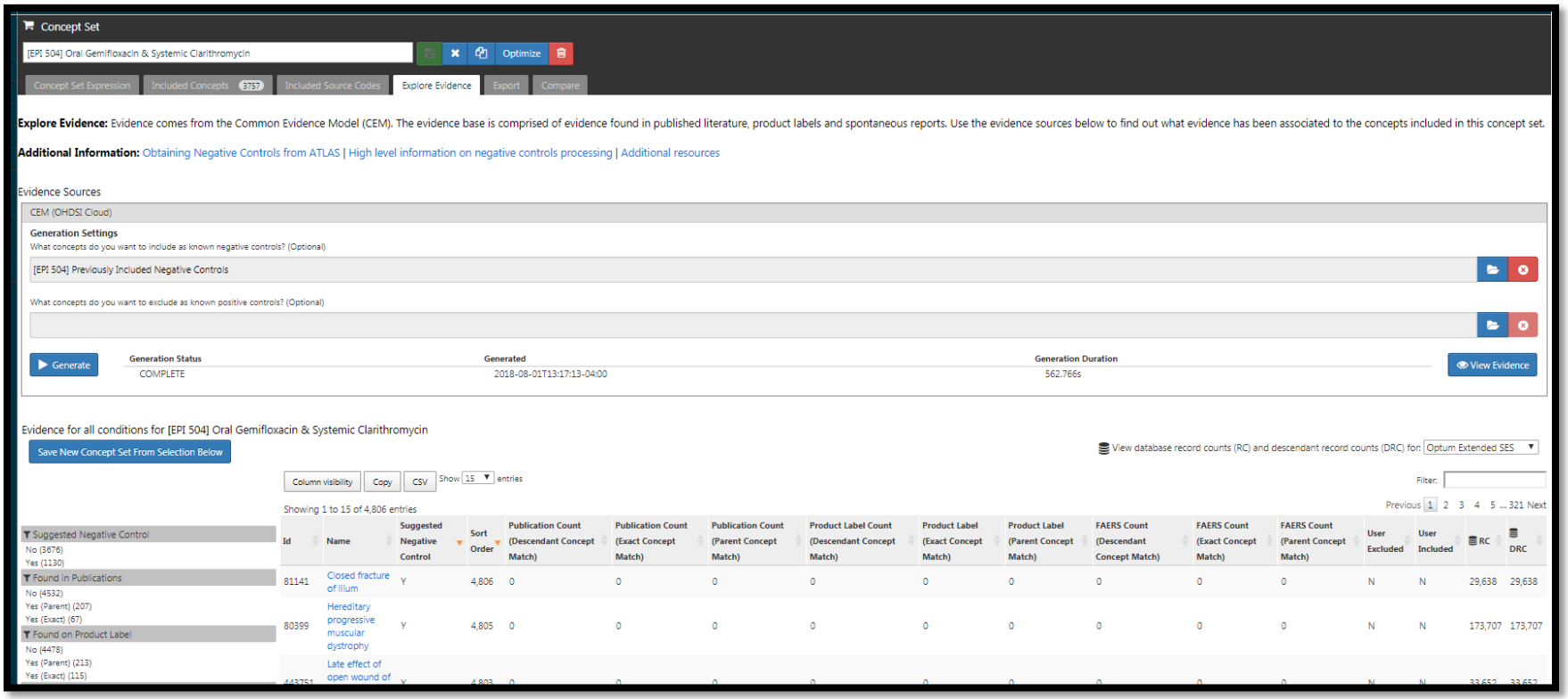
No Confounding Relationships



How to Find Negative Controls?

- Finding negative controls manually:
  - **FDA product labels** does not associate the drug and outcome
  - There are now spontaneous reports signals in the **U.S. FDA's Adverse Event Reporting System (FAERS)**
  - Manual review in PubMed that there is **no studies** showing the drug causes the condition
- However OHDSI has developed tools in ATLAS to make this easier . . .

Use ATLAS to Build Negative Control Outcomes



Generating Negative Controls in ATLAS

- ATLAS can be used to generate and suggest negative controls for a study.

- Suggested negative controls:
  - have no published literature adverse event association
  - not existing on the product label
  - not considered a FAERS signal
  - have no indication or contraindication listed in the OMOP Vocabulary for the pair
  - are not considered a broad concepts
  - are not considered a drug induced concept
  - not considered a pregnancy related concept
  - was not suggested to be excluded by the user
  - was not optimized out, meaning another parent concept existed that was also considered a good negative, so the lower level concept was excluded

- All information known about the drug-condition pair is provided so if the user wishes to loosen this criteria they can (i.e. ignore evidence from ancestor restrictions).

- https://github.com/OHDSI/CommonEvidenceModel/wiki/Negative-Controls-In-ATLAS

CEM Next Steps

- CEM is under development to help improve its use:
  - Working on a publication to describe what **is a good negative control** and comparing different negative control selection methods
  - Martijn, Rave Harpaz (Oracle), and I are working on another method for **parsing of US Product Labels**
  - Martijn, Anthony Sena, and I are working on an app for non-OHDSI members to go from **source codes to source codes** (e.g. given a set of NDC (drug source codes) receive a set of ICD10s (outcome source codes) to be used as negative controls)

BACKUP SLIDES

Finding Good Negative Control Outcomes

Example Study

BMJ Open

Risk of death among users of Proton Pump Inhibitors: a longitudinal observational cohort study of United States veterans

Yan Xie,[1] Benjamin Bowe,[1] Tingting Li,[1,2] Hong Xian,[1,3] Yan Yan,[1,4] Ziyad Al-Aly[1,2,5,6]

Objective Proton pump inhibitors (PPIs) are widely used, and their use is associated with increased risk of adverse events. However, whether PPI use is associated with excess risk of death is unknown. We aimed to examine the association between PPI use and risk of all-cause mortality.

Conclusions The results suggest excess risk of death among PPI users; risk is also increased among those without gastrointestinal conditions and with prolonged duration of use. Limiting PPI use and duration to instances where it is medically indicated may be warranted.

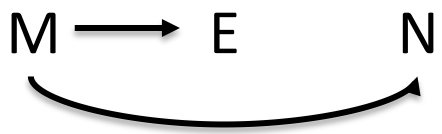
Study Set Up

Comparative Cohort Analysis			
Inputs	Target Cohort (T)		*New Users of a Proton Pump Inhibitor (PPI)*
	Comparator Cohort (C)		*New Users of Histamine H2 Receptor Antagonist (H2 Blockers)*
	Outcome Cohort (O)		*Death*
	Database		*Dept. Of Veterans Affairs Database*
	Effect Estimate Parameters	Model Type	*Cox*
		Time at Risk Start and End	*Start: Index (exposure to drug)* *End: Until End of Observation Periods (Intend to Treat Analysis)*
		Washout Period	*N/A*
		Minimum Days at Risk	*N/A*
		Remove subjects with prior outcomes?	*N/A*
		Methods to adjust for bias	*1:1 propensity score matching*

While the authors discussed how they tried to control for bias, empirical calibration was not one of the methods

Let us find negative controls for this study . . .

Good Negative Control Outcome



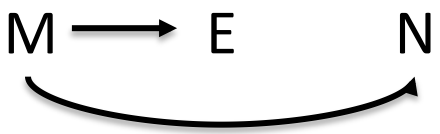
- E = PPI or H2 Blocker
- N = Rosacea
  (CEM Suggested)

Exposure Does Not Cause the Outcome

Ideally, Similar Confounding Relationships to Measured and Unmeasured Variables

Good Negative Control Outcome

Exposure Does Not Cause the Outcome



- E = PPI or H2 Blocker
- N = Rosacea
  (CEM Suggested)

- Do PPIs/H2 Blockers cause Rosacea?

- What evidence do we find?
  - **Spontaneous reports** there is no disproportionality analysis signal (PRR <2) (Evans et al.)
  - Not existing on the **US product label** in the "Adverse Drug Reactions" or "Postmarketing" section
  - Found no **publications** suggested drug-outcome pair were in adverse event relationship (Winnenburg et al.)

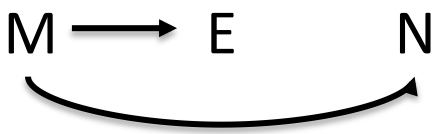
- There are some mention of rash and skin inflammation if you review the label however no mention of acne

Winnenburg R, Sorbello A, Ripple A, Harpaz R, Tonning J, Szarfman A, Francis H, Bodenreider O. Leveraging MEDLINE indexing for pharmacovigilance - Inherent limitations and mitigation strategies. J Biomed Inform. 2015 Oct;57:425-35. doi: 10.1016/j.jbi.2015.08.022. Epub 2015 Sep 2. PubMed PMID: 26342964; PubMed Central PMCID: PMC4775467.

Evans, S.J., P.C. Waller, and S. Davis, Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiol Drug Saf, 2001. 10(6): p. 483-6.

Good Negative Control Outcome

Exposure Does Not Cause the Outcome

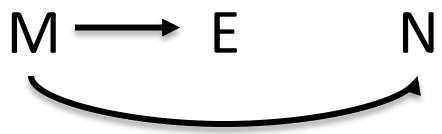


- E = PPI or H2 Blocker
- N = Rosacea
  (CEM Suggested)

CEM is helpful in finding these relationships!

Good Negative Control Outcome

Ideally, Similar Confounding Relationships to Measured and Unmeasured Variables



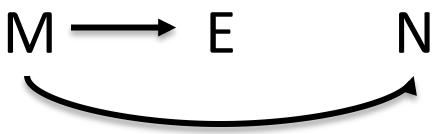
- E = PPI or H2 Blocker
- N = Rosacea
  (CEM Suggested)

- Influences that both affect getting rosacea and death?
  - Age
    - Rosacea is most common in years 30 to 50
    - Risk of death increases with age
  - Gender
    - Menopause can cause Rosacea
    - Women on average live longer
  - Smoking Status
    - Smokers have an increased risk of developing rosacea
    - Smoking influences death

Good Negative Control Outcome

Ideally, Similar Confounding Relationships to Measured and Unmeasured Variables

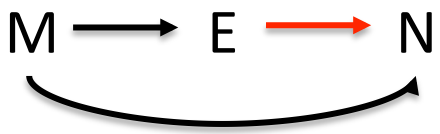
- E = PPI or H2 Blocker
- N = Rosacea
  (CEM Suggested)



CEM cannot determine which concepts have similar confounding relationships.

Finding Bad Negative Control Outcomes

Bad: Exposure Causes Outcome



- E = PPI or H2 Blocker
- N = Fever

OMEPRAZOLE- omeprazole capsule, delayed release
Aidarex Pharmaceuticals LLC

Body As
bronchos

Drug Saf. 1990 Nov-Dec;5(6):455-7.

IgE-mediated drug fever due to histamine H2-receptor blockers.

Ann Pharmacother. 2005 Apr;39(4):757-60. Epub 2005 Mar 1.

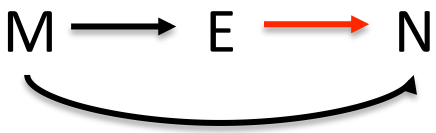
Esomeprazole-induced central fever with severe myalgia.

Pharmacology. 2014;94(1-2):78-9. doi: 10.1159/000364781. Epub 2014 Sep 13.

Drug fever due to a single dose of pantoprazole.

Schiller D[1], Maieron A, Schöfl R, Donnerer J.

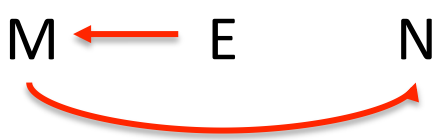
Bad: Exposure Causes Outcome



- E = PPI or H2 Blocker
- N = Fever

CEM is helpful in finding these relationships!

Bad: Exposure Causes a Variable that Cause Outcome



- E = PPI or H2 Blocker
- N = Falls
  (CEM Suggested)

OMEPRAZOLE- omeprazole capsule, delayed release
Aidarex Pharmaceuticals LLC

---

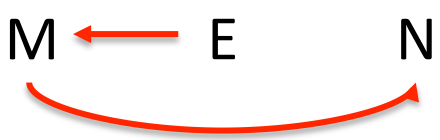
5 WARNINGS AND PRECAUTIONS

5.3 Bone Fracture

Several published observational studies suggest that proton pump inhibitor (PPI) therapy may be associated with an increased risk for osteoporosis-related fractures of the hip, wrist, or spine. The risk of fracture was increased in patients who received high-dose, defined as multiple daily doses, and long-term PPI therapy (a year or longer). Patients should use the lowest dose and shortest duration of PPI therapy appropriate to the condition being treated. Patients at risk for osteoporosis-related fractures should be managed according to established treatment guidelines. [see **Dosage and Administration (2)** and **Adverse Reactions (6.3)**]

- PPIs can cause fractures, which may lead to falls

Bad: Exposure Causes a Variable that Cause Outcome



- E = PPI or H2 Blocker
- N = Falls
  (CEM Suggested)

CEM cannot determine which concepts that lead to other concepts.

Bad: No Confounding Relationships

M E N

- E = PPI or H2 Blocker
- N = Effects of lightning
  (CEM Suggested)

- Does the outcome effects of lightning have similar confounding relationships to other measured variables
- Hard to imagine the confounding relationships

Bad: No Confounding Relationships

M E N

- E = PPI or H2 Blocker
- N = Effects of lightning
  (CEM Suggested)

CEM cannot determine which concepts lack similar confounding relationships.

Bad: Outcome not Prevalent

- E = PPI or H2 Blocker
- N = Absent nipple
  (CEM Suggested)

- If an outcome is not prevalent does it make a good negative control?

- CCAE only has 104 people with an absent nipple

- Concepts that are so rare, you are unlikely to produce an effect estimate with enough precision to be informative

Bad: Outcome not Prevalent

- E = PPI or H2 Blocker
- N = Absent nipple
  (CEM Suggested)

CEM ATLAS tool orders results from most common to least common.

Reviewing Negative Controls in CEM

- Exposure causes outcome

- Drug indications
- Orders concepts in prevalence order

- Concepts have similar confounding?
- Outcome causes another outcome or exposure

CEM is a huge improvement over manually finding negative controls, but still requires thoughtful review!