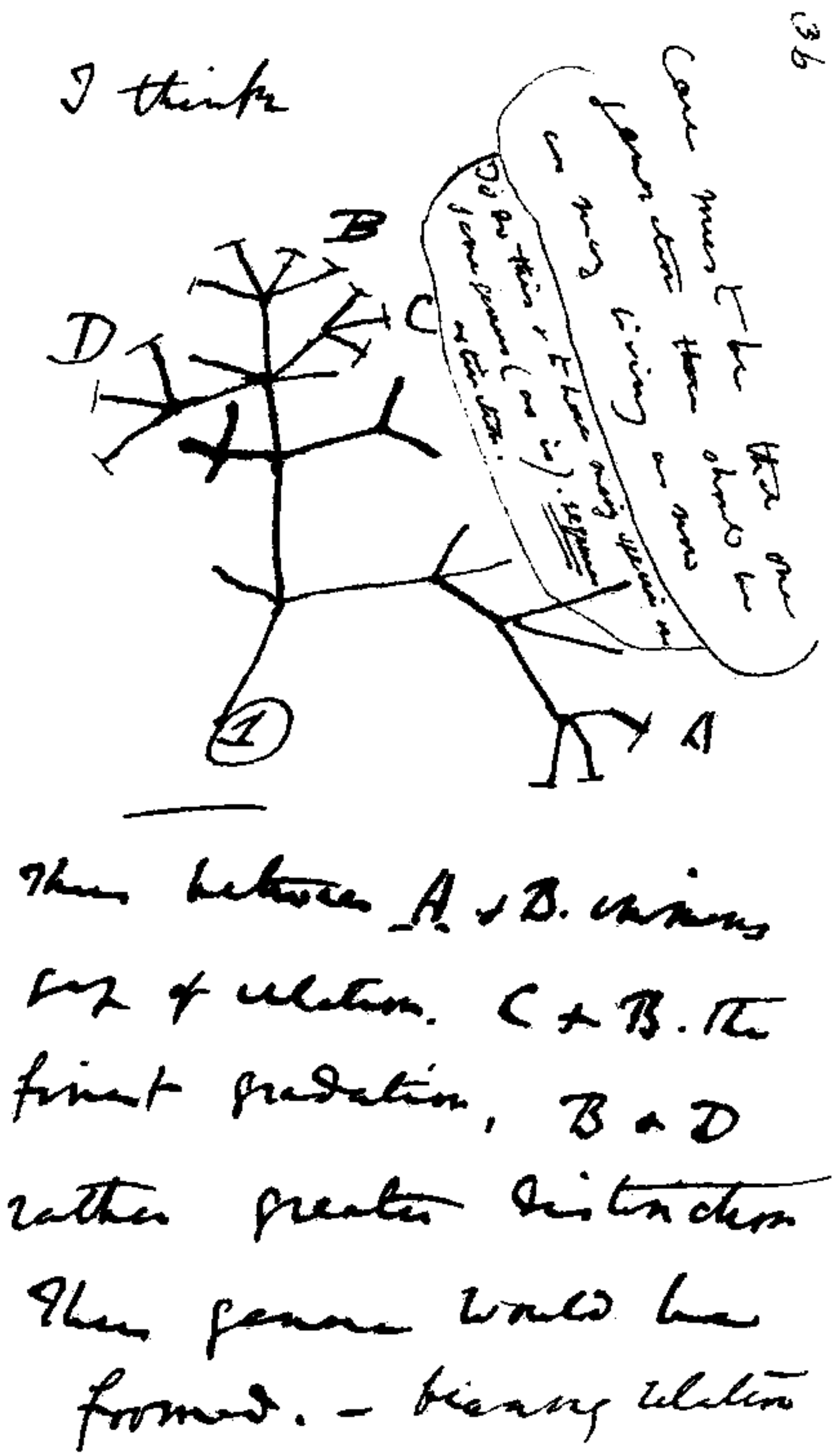
Biomedical Discovery through Data Mining and Data Science

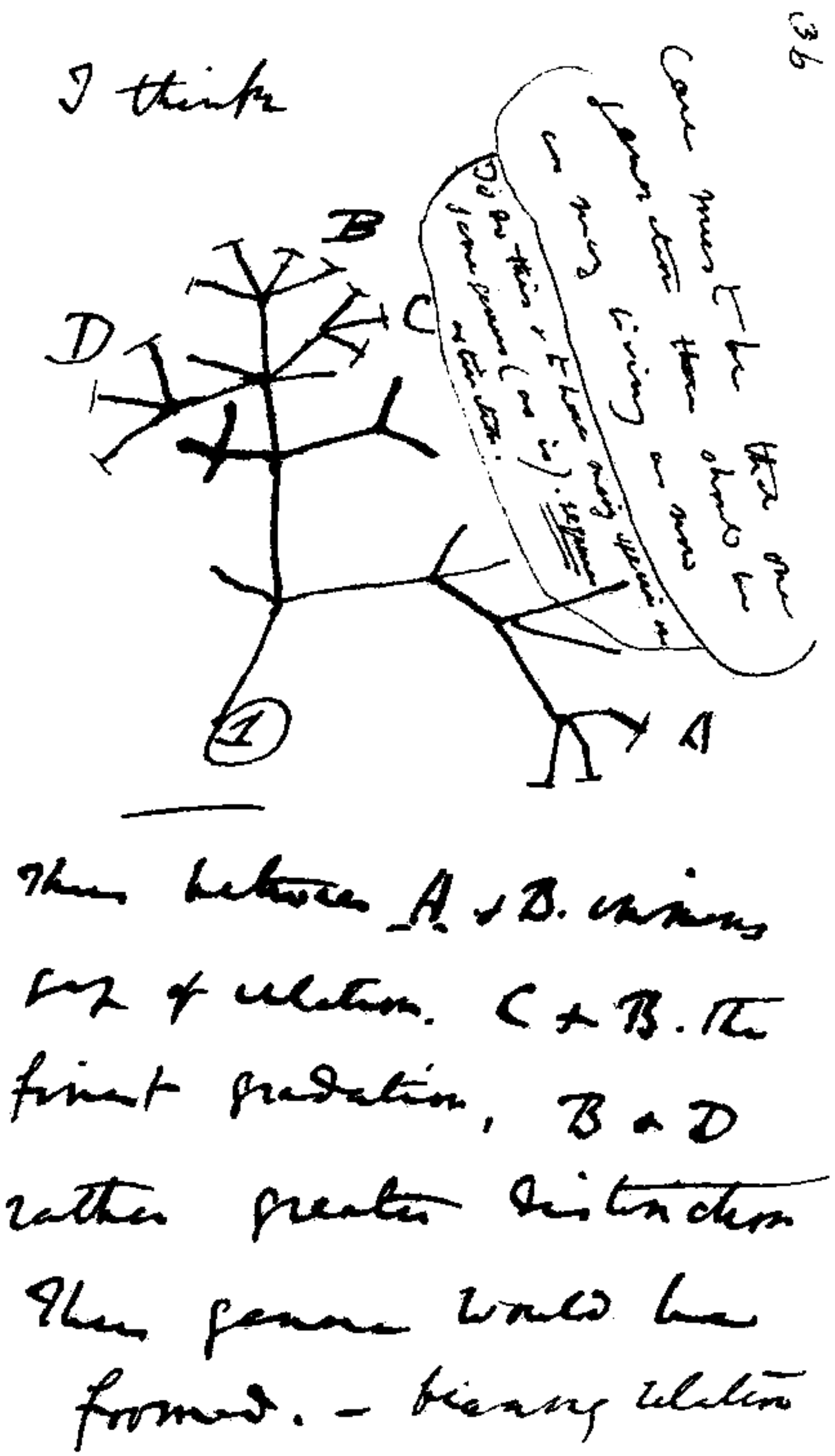
November 14th, 2016

Nicholas P. Tatonetti, PhD
Columbia University

Observation is the starting point of biological discovery

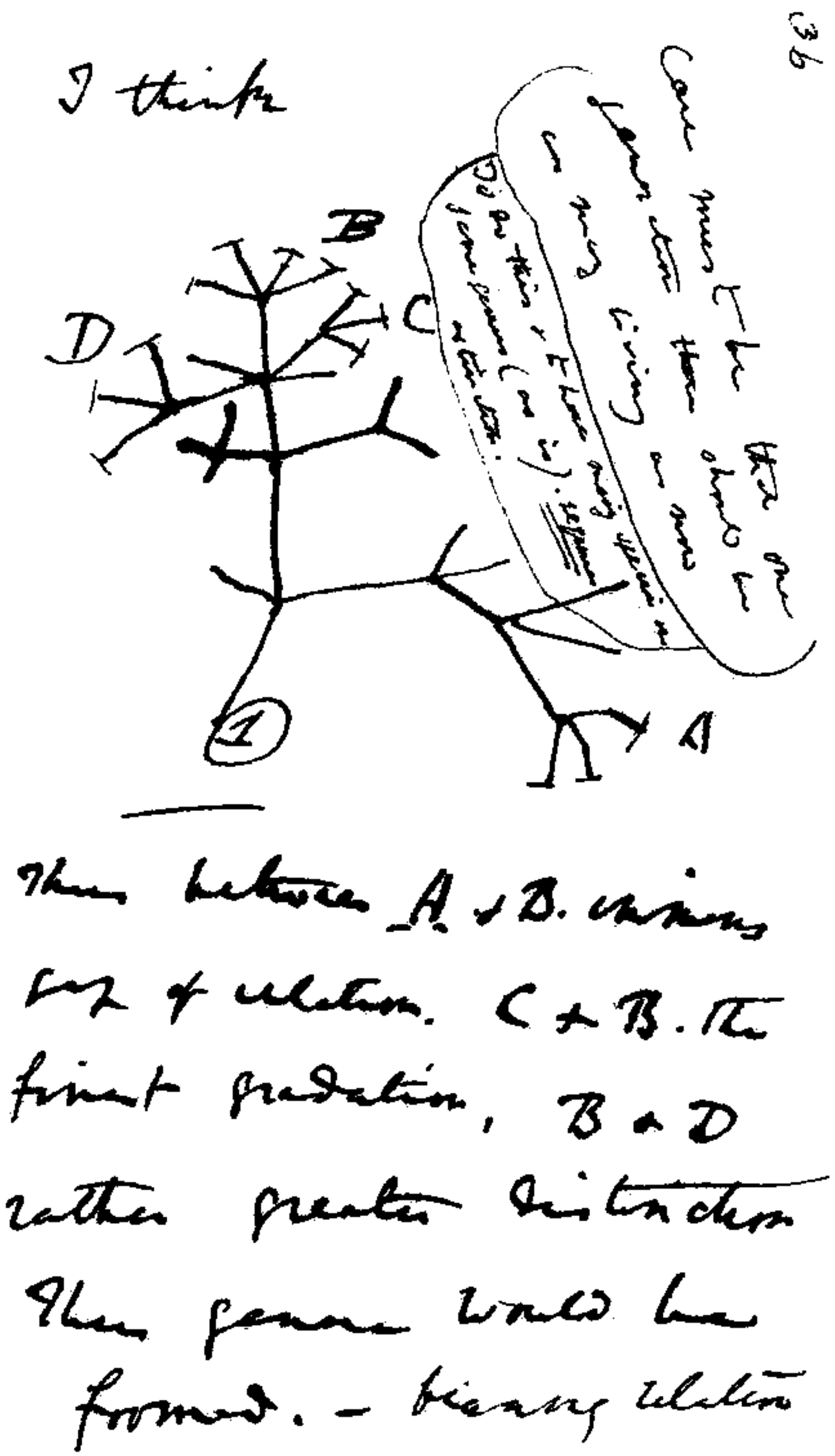


Observation is the starting point of biological discovery



- Charles Darwin observed relationship between geography and phenotype

Observation is the starting point of biological discovery



- Charles Darwin observed relationship between geography and phenotype
- William McBride & Widukind Lenz observed association between thalidamide use and birth defects

The tools of observation are advancing

The tools of observation are advancing

- Human senses

The tools of observation are advancing

- Human senses
  - sight, touch, hearing, smell, taste

The tools of observation are advancing

- Human senses
  - sight, touch, hearing, smell, taste
- Mechanical augmentation

The tools of observation are advancing

- Human senses
  - sight, touch, hearing, smell, taste
- Mechanical augmentation
  - binoculars, telescopes, microscopes, microphones

The tools of observation are advancing

- Human senses
  - sight, touch, hearing, smell, taste
- Mechanical augmentation
  - binoculars, telescopes, microscopes, microphones
- Chemical and Biological augmentations

The tools of observation are advancing

- Human senses
    - sight, touch, hearing, smell, taste
- Mechanical augmentation
    - binoculars, telescopes, microscopes, microphones
- Chemical and Biological augmentations
    - chemical screening, microarrays, high throughput sequencing technology

The tools of observation are advancing

- Human senses
    - sight, touch, hearing, smell, taste
- Mechanical augmentation
    - binoculars, telescopes, microscopes, microphones
- Chemical and Biological augmentations
    - chemical screening, microarrays, high throughput sequencing technology

Bytes to KB

Megabytes to Terabytes

The tools of observation are advancing

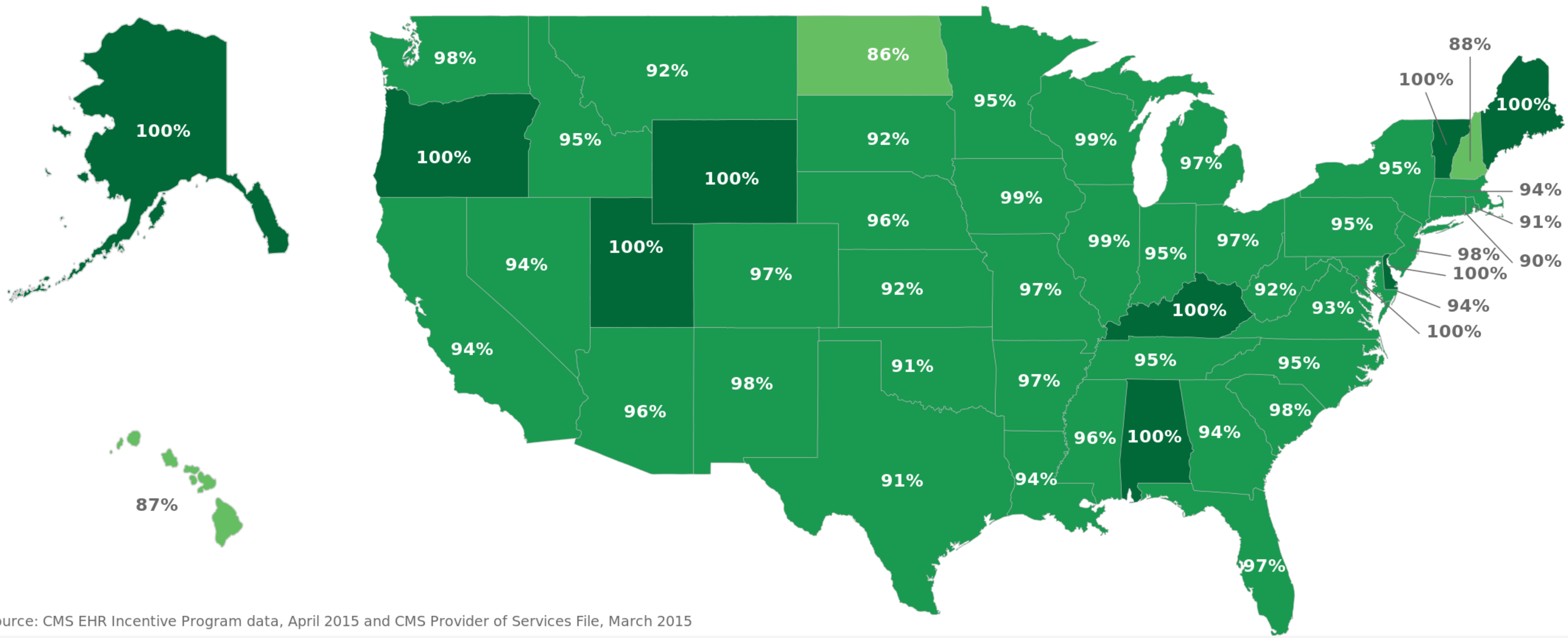
- Human senses
  - sight, touch, hearing, smell, taste
- Mechanical augmentation
  - binoculars, telescopes, microscopes, microphones
- Chemical and Biological augmentations
  - chemical screening, microarrays, high throughput sequencing technology
- What's next?

Bytes to KB

Megabytes to Terabytes

Your doctor is observing you like never before

>99% of Hospitals have Electronic Health Records



Source: CMS EHR Incentive Program data, April 2015 and CMS Provider of Services File, March 2015

Every drug order is an experiment.

Observation analysis in a *peta*byte world

Observation analysis in a *peta*byte world

- Darwin, McBride, and Lenz were working with *kilo*bytes of data

Observation analysis in a *peta*byte world

- Darwin, McBride, and Lenz were working with *kilo*bytes of data
- Today's scientists are observing *tera*bytes and *peta*bytes of data

Observation analysis in a *peta*byte world

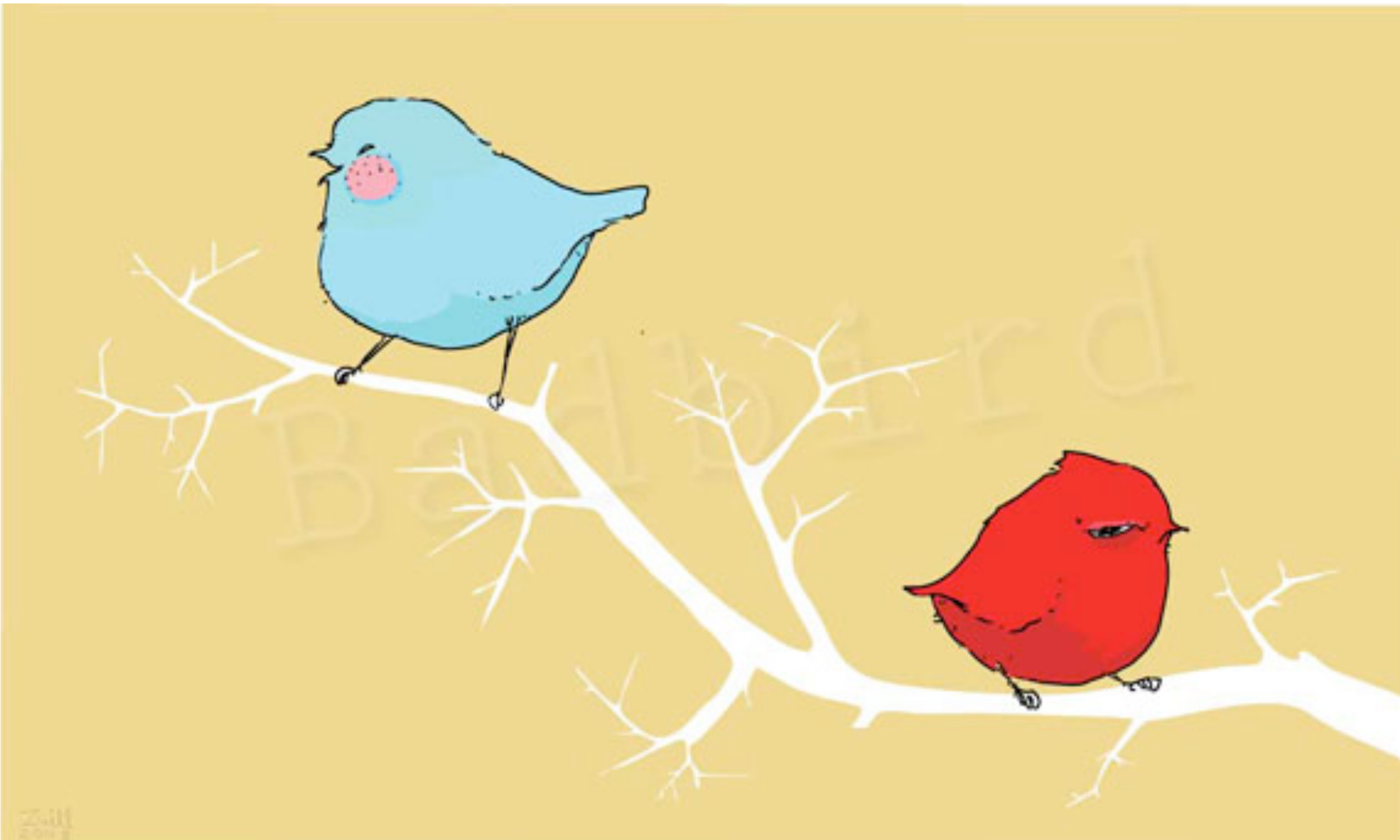
- Darwin, McBride, and Lenz were working with *kilo*bytes of data
- Today's scientists are observing *tera*bytes and *peta*bytes of data
- The human mind simply cannot make sense of that much information

Observation analysis in a *peta*byte world

- Darwin, McBride, and Lenz were working with *kilo*bytes of data
- Today's scientists are observing *tera*bytes and *peta*bytes of data
- The human mind simply cannot make sense of that much information
- Data mining is about making the tools of data analysis ("hypothesis generation") catch up to the tools of observation

But, there's a problem…

Bias confounds observations



Let’s focus on just one example...

Let's focus on just one example...

Drug-Drug Interactions

Drug-drug interactions (DDIs)

Drug-drug interactions (DDIs)

- DDIs can occur when a patient takes 2 or more drugs

Drug-drug interactions (DDIs)

- DDIs can occur when a patient takes 2 or more drugs
- DDIs cause unexpected side effects

Drug-drug interactions (DDIs)

- DDIs can occur when a patient takes 2 or more drugs
- DDIs cause unexpected side effects
  - 10-30% of adverse drug events are attributed to DDIs

Drug-drug interactions (DDIs)

- DDIs can occur when a patient takes 2 or more drugs
- DDIs cause unexpected side effects
  - 10-30% of adverse drug events are attributed to DDIs
- Understanding of DDIs may lead to better outcomes

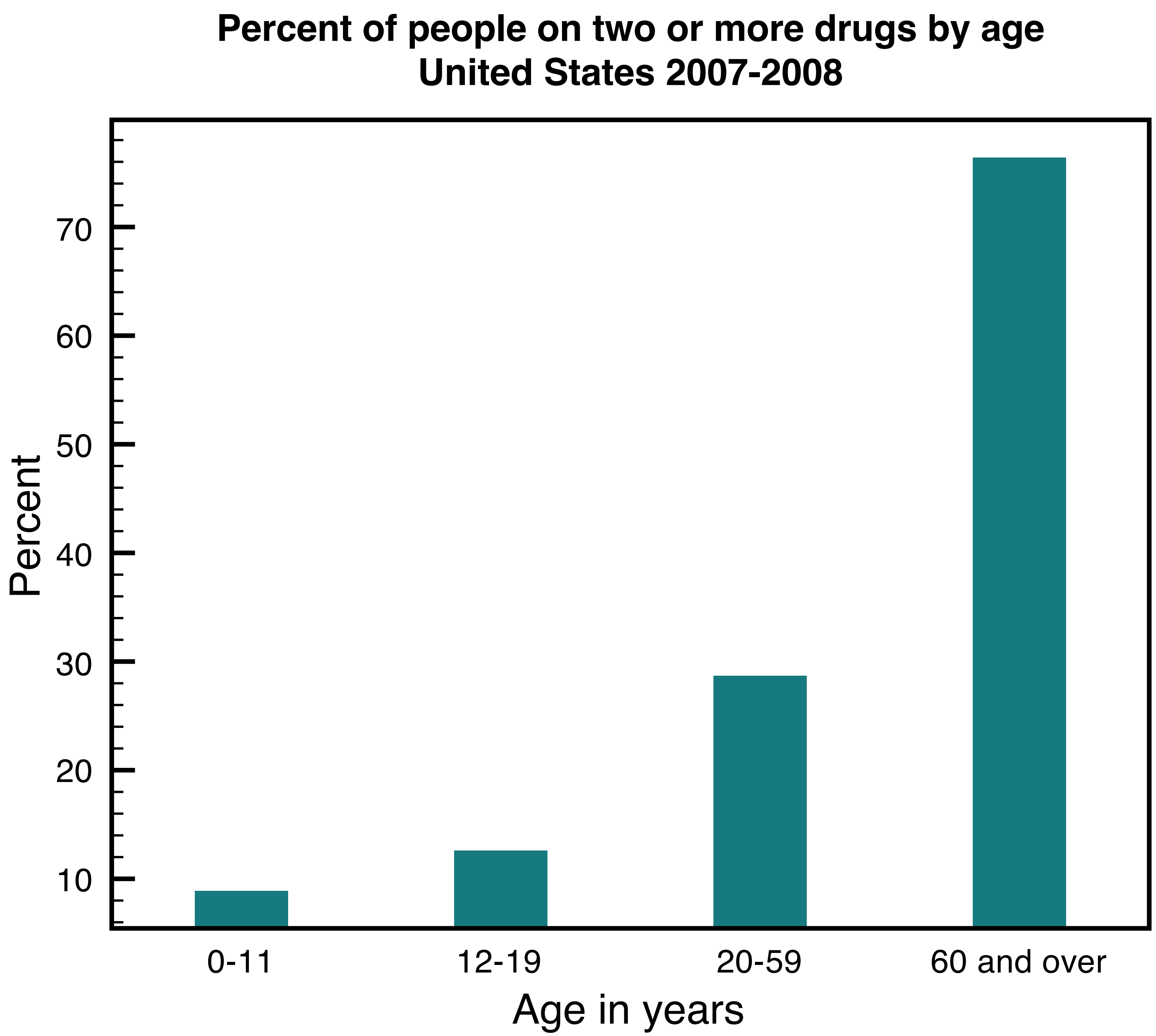
Drug-drug interactions (DDIs)

- DDIs can occur when a patient takes 2 or more drugs
- DDIs cause unexpected side effects
  - 10-30% of adverse drug events are attributed to DDIs
- Understanding of DDIs may lead to better outcomes
  - precaution in prescription

Drug-drug interactions (DDIs)

- DDIs can occur when a patient takes 2 or more drugs
- DDIs cause unexpected side effects
    - 10-30% of adverse drug events are attributed to DDIs
- Understanding of DDIs may lead to better outcomes
    - precaution in prescription
    - synergistic therapies

Polypharmacy increases with age



SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey

76% of older Americans used two or more prescription drugs

More needs to be done to understand and identify drug-drug interactions

More needs to be done to understand and identify drug-drug interactions

- Clinical trials do not typically investigate drug-drug interactions

More needs to be done to understand and identify drug-drug interactions

- Clinical trials do not typically investigate drug-drug interactions
- **Observational studies** are the only systematic way to detect drug-drug interactions

Large population databases enable DDI discovery

- Contain clinical data on millions of patients over many years
- Currently being used to establish single drug adverse events (pharmacovigilance)
- Eg. **Spontaneous Adverse Event Reporting Systems**
    - Collect adverse event reports for a patient (a snapshot in time)
    - Maintained by WHO > FDA > Health Canada

Observational data are messy

Drugs	**Adverse Events**
METFORMIN	ACUTE RESP. DISTRESS
ROSIGLITAZONE	ANEMIA
PRAVASTATIN	DECR. BLOOD PRESSURE
TACROLIMUS	CARDIAC FAILURE
PREDNISOLONE	DEHYDRATION

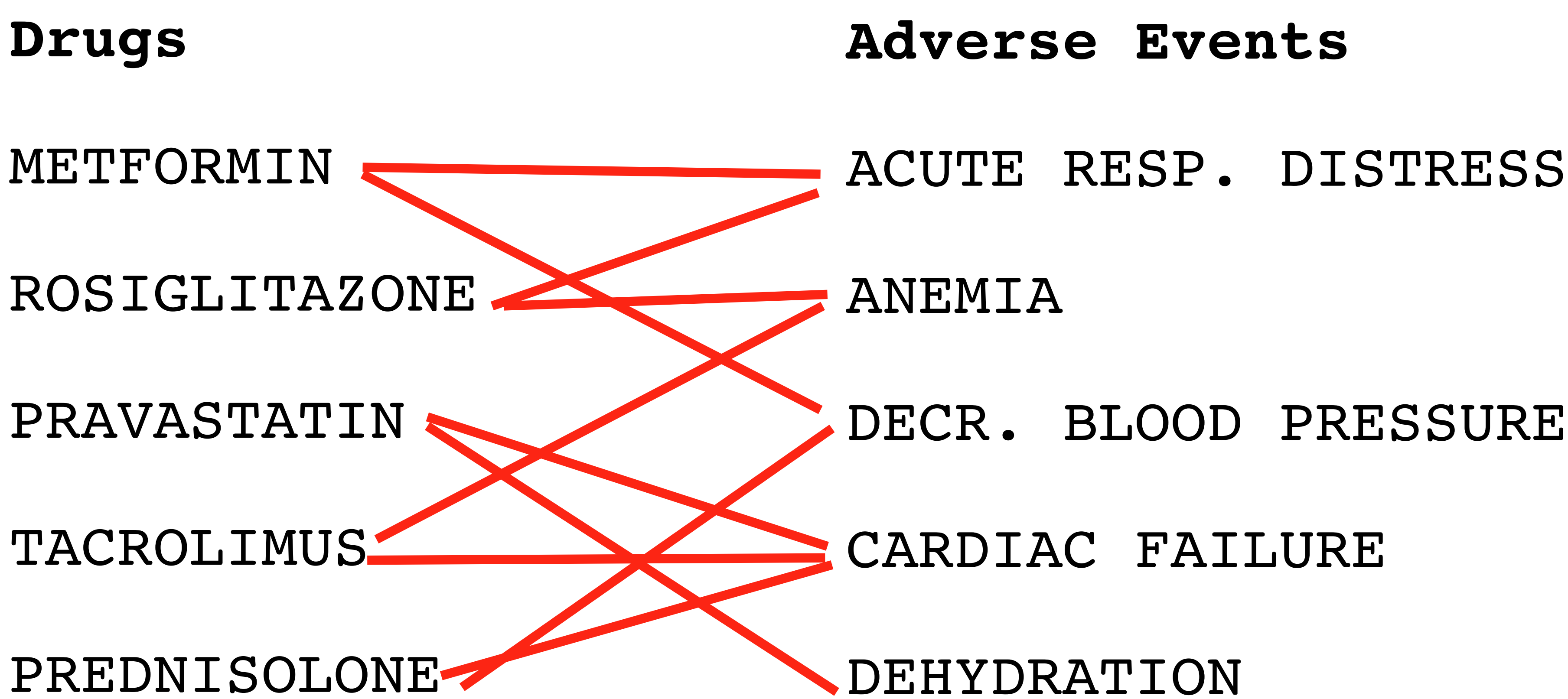
Observational data are messy

- Many drugs, many adverse events

Drugs	**Adverse Events**
METFORMIN	ACUTE RESP. DISTRESS
ROSIGLITAZONE	ANEMIA
PRAVASTATIN	DECR. BLOOD PRESSURE
TACROLIMUS	CARDIAC FAILURE
PREDNISOLONE	DEHYDRATION

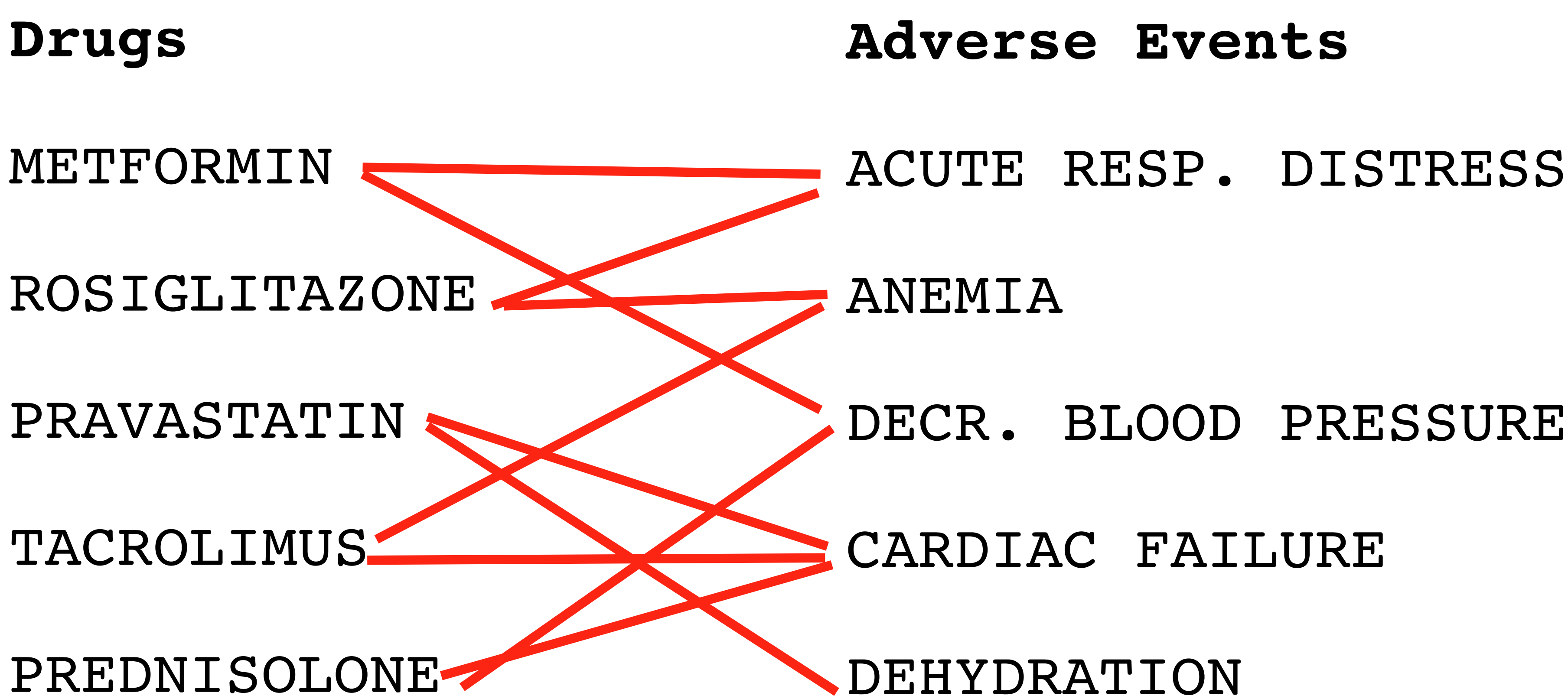
Observational data are messy

- Many drugs, many adverse events
  - what causes what?



Observational data are messy

- Many drugs, many adverse events
  - what causes what?



most of these red lines are false - which are true?

Observational data are confounded

- Spontaneous reporting systems are observational data sets (unknown biases)
- noise from concomitant drug use (***co-Rx effect***)
  - drugs co-prescribed with Vioxx more likely to be associated with heart attacks
- noise from indications (***indication-effect***)
  - drugs given to diabetics more likely to be associated with hyperglycemia

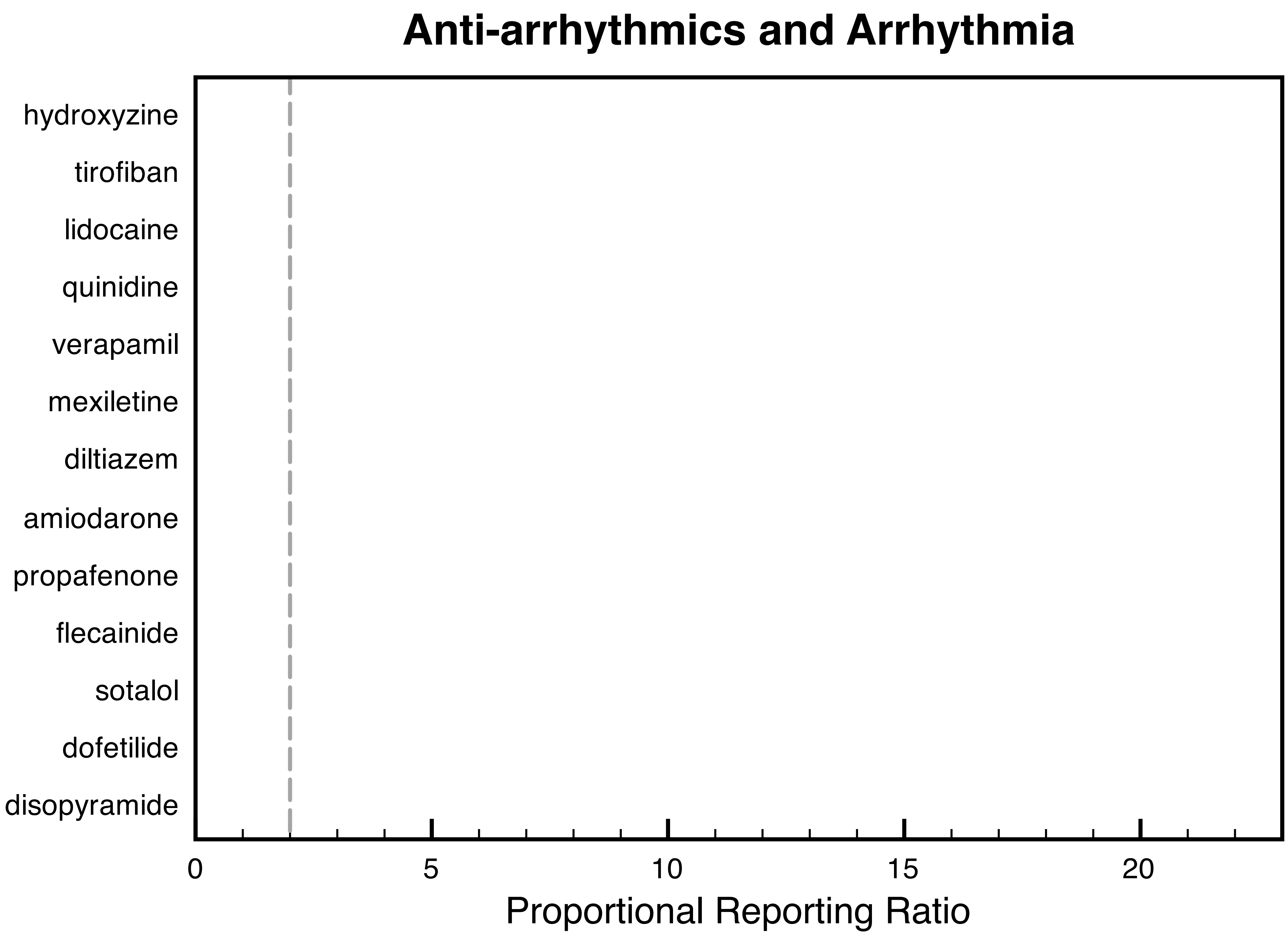
SCRUB

Statistical CorRection of Uncharacterized Bias

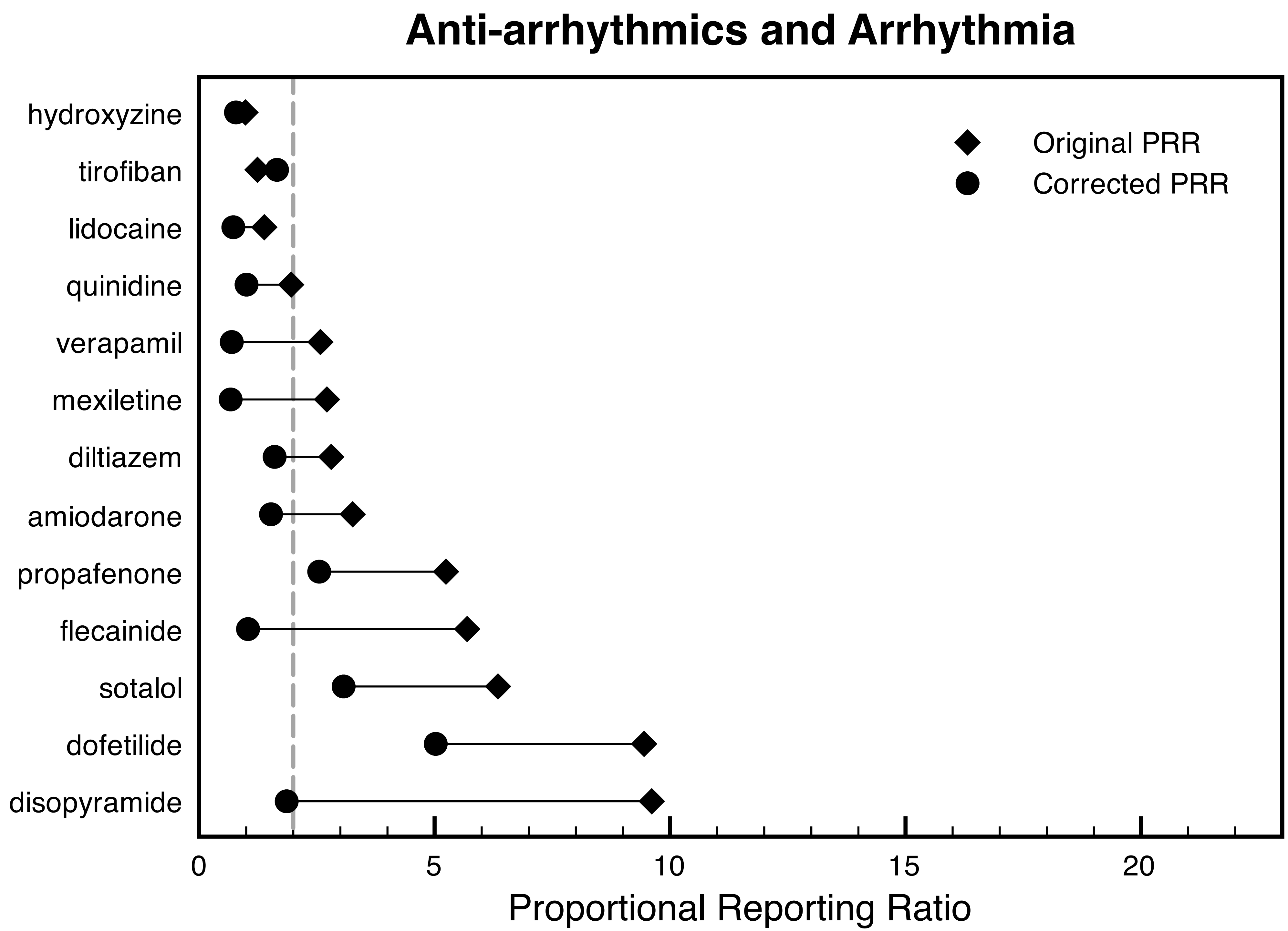
- *Implicitly* corrects for confounding of both observed and missing variables
- Assumes some combination of the **drugs** and **indications** describes the patient covariates
- Only works on very large data sets

N. Tatonetti et al., *Science Translational Medicine* (2012)

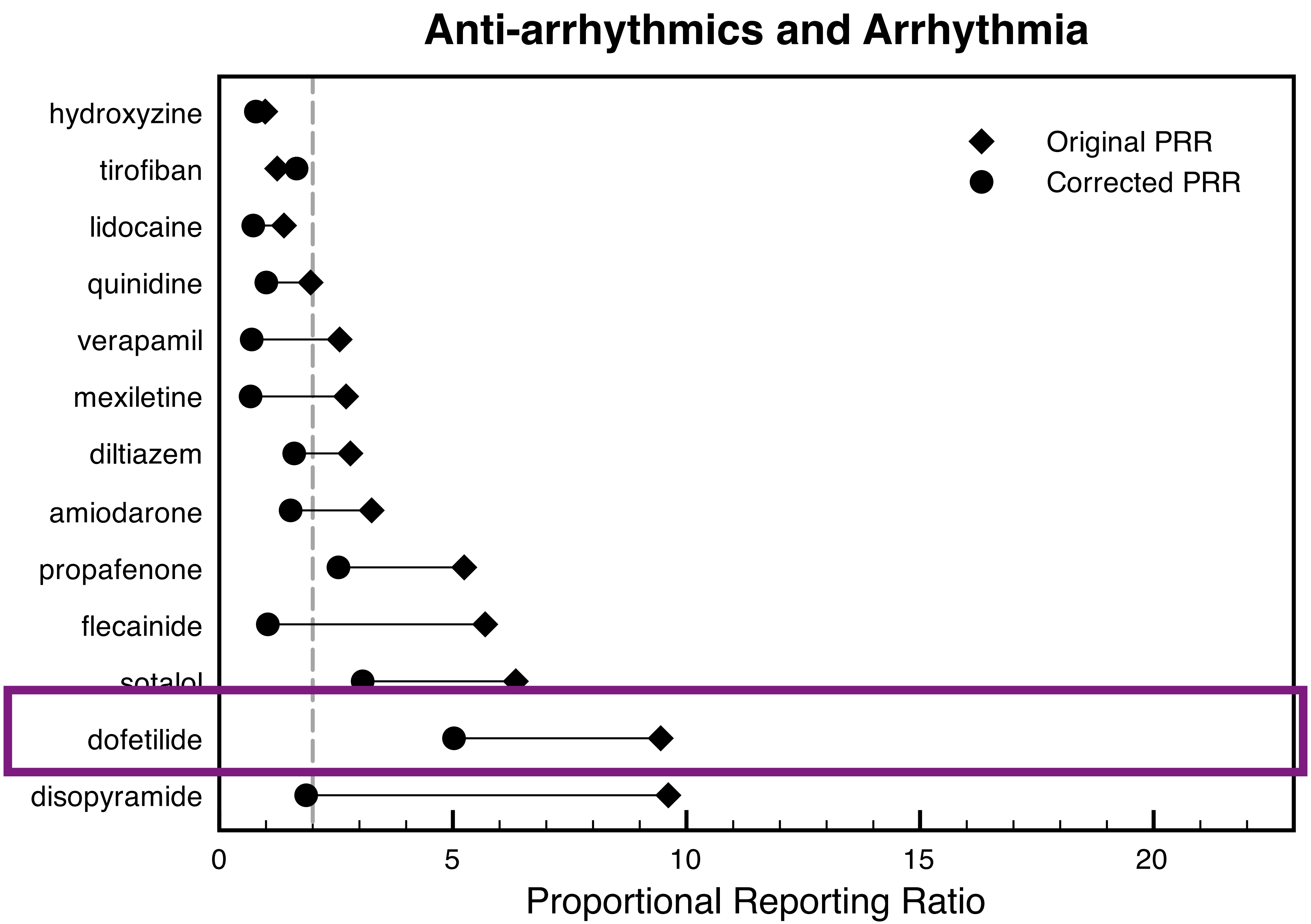
Method corrects for **indication biases**



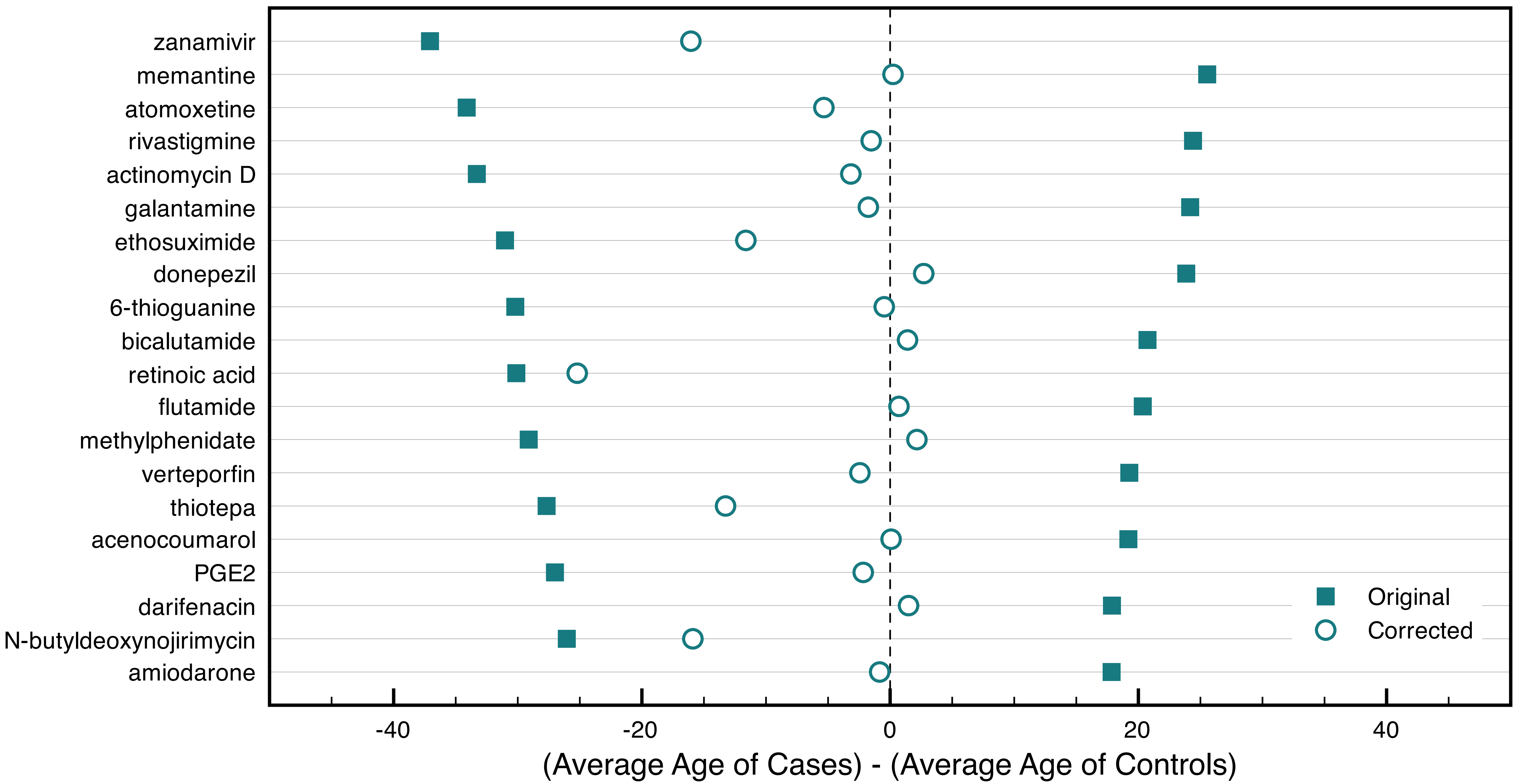
Method corrects for indication biases



Method corrects for **indication biases**



Implicit correction of age differences in exposed vs non-exposed



Bias, corrected. Missing data?

If there are no observations
then no associations can be found.

Diseases can be identified by the side effects they elicit

Diseases can be identified by the side effects they elicit

Diabetes

Diseases can be identified by the side effects they elicit

level of detection

Diabetes

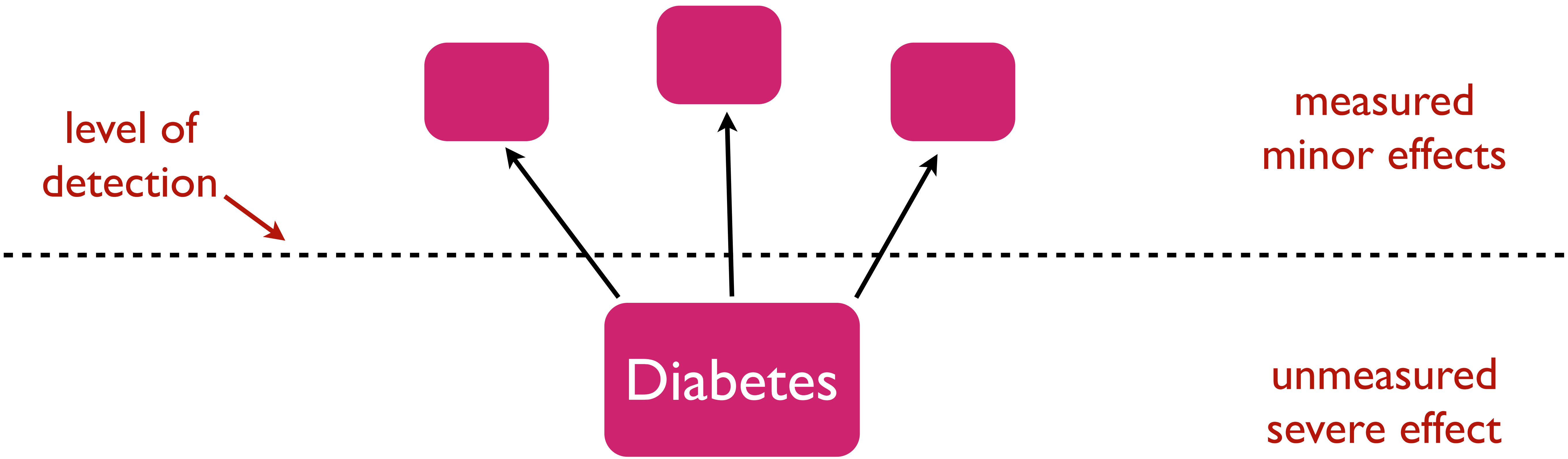
Diseases can be identified by the side effects they elicit

level of detection

Diabetes

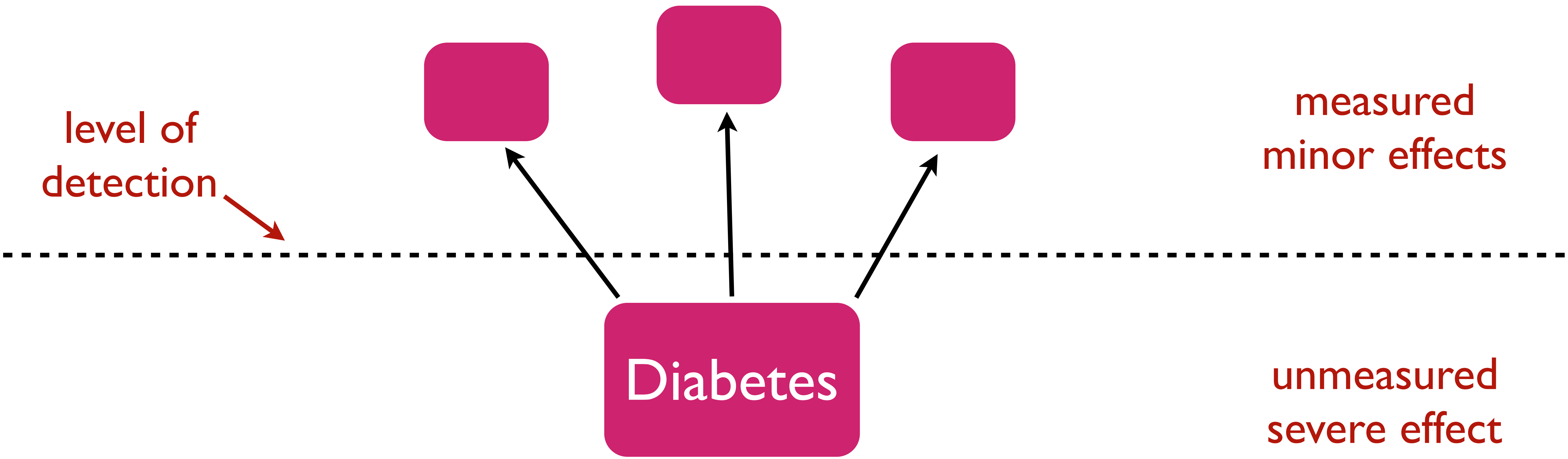
unmeasured severe effect

Diseases can be identified by the side effects they elicit



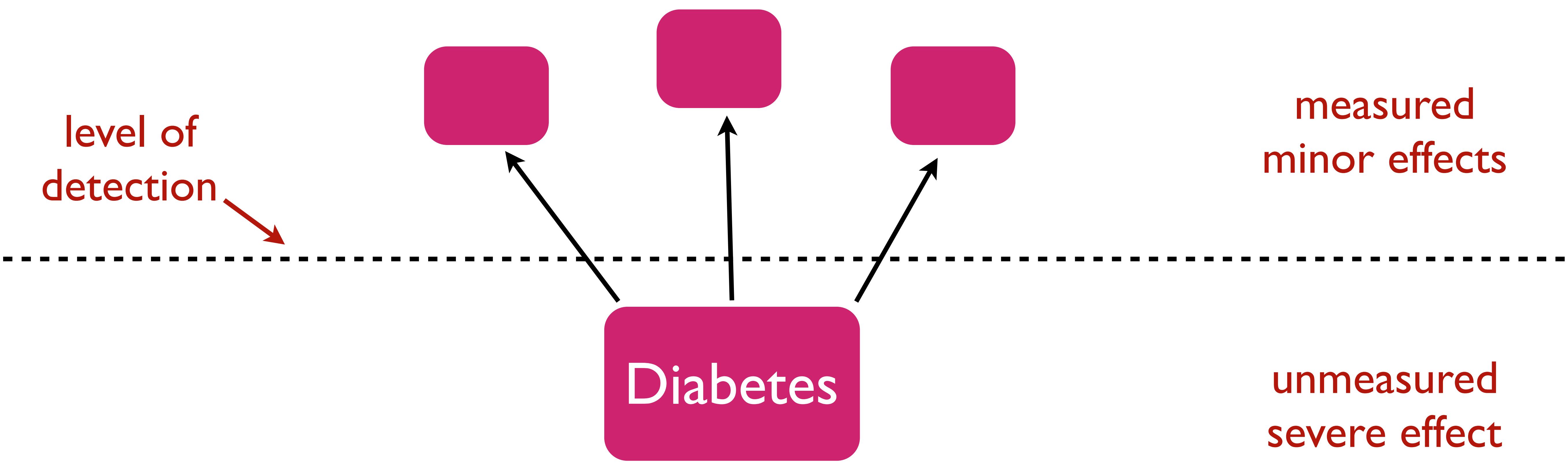
Diseases can be identified by the side effects they elicit

- physicians use observable side effects to form hypothesis about the underlying disease

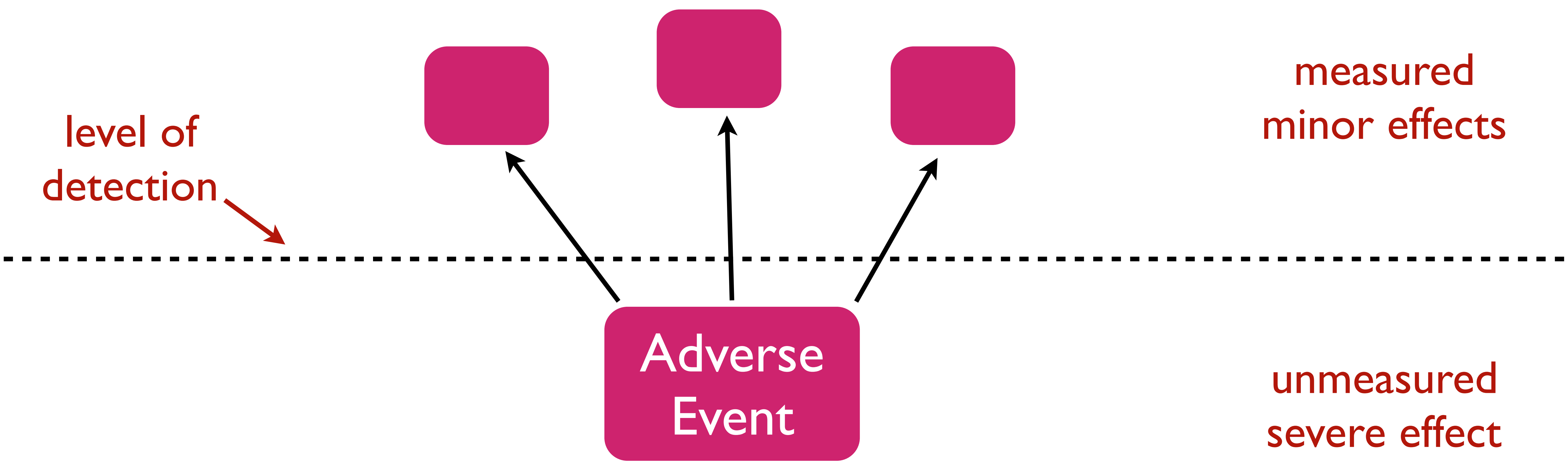


Diseases can be identified by the side effects they elicit

- physicians use observable side effects to form hypothesis about the underlying disease
- e.g. you can't *see* diabetes, but you can *measure* blood glucose

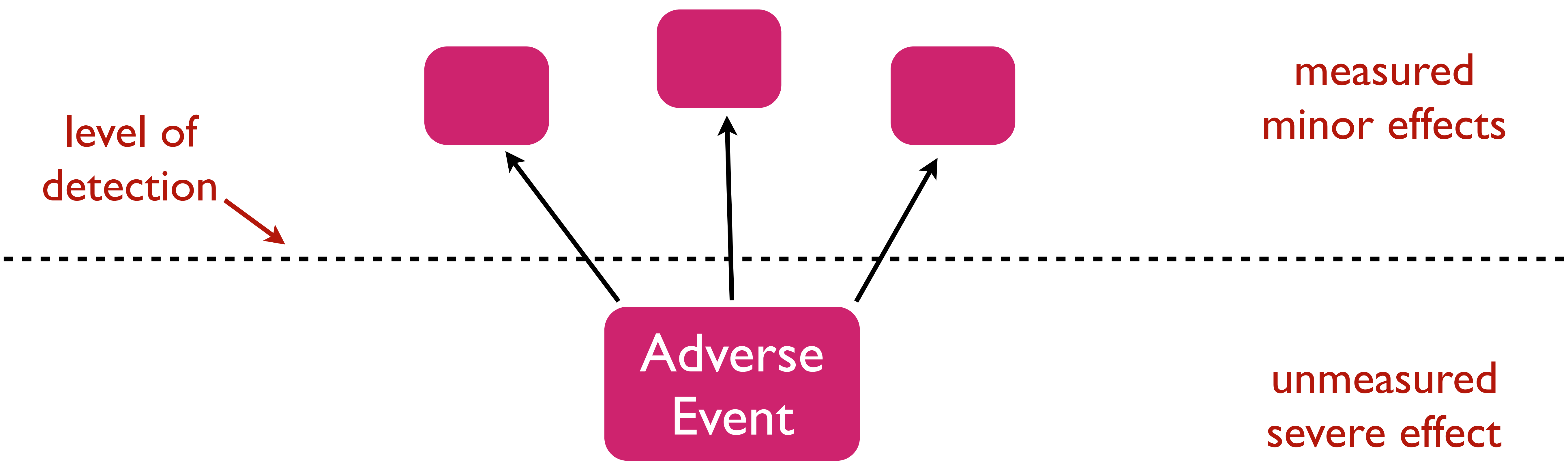


Severe ADE's can be identified by the presence of more minor (and more common) side effects



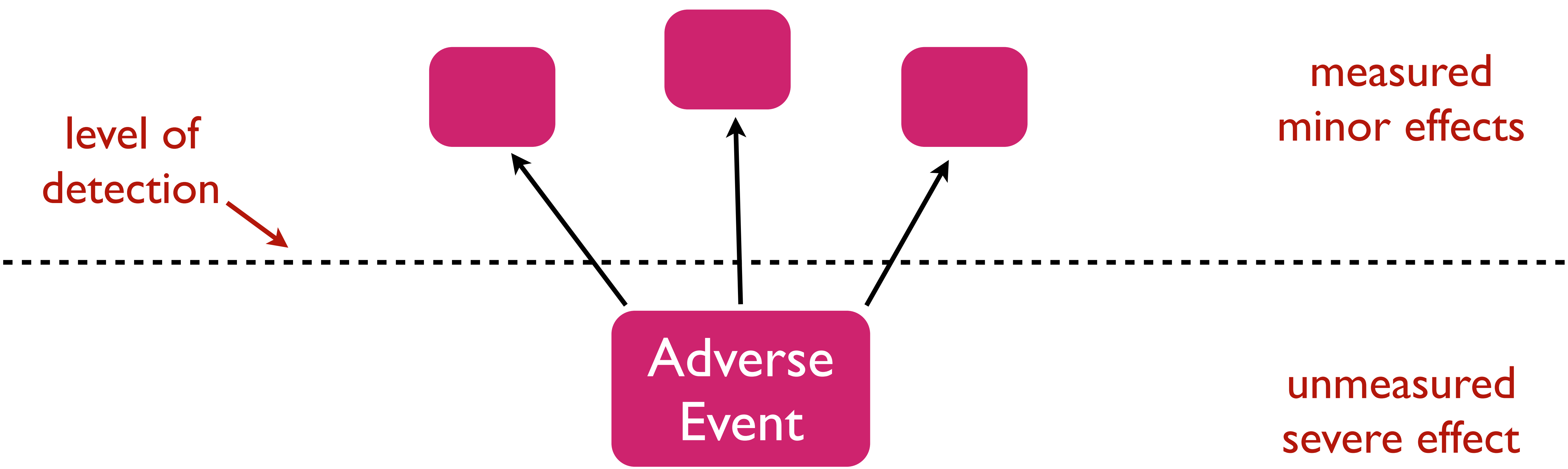
Severe ADE's can be identified by the presence of more minor (and more common) side effects

- First, identify the common side effects that are harbingers for the underlying severe AE

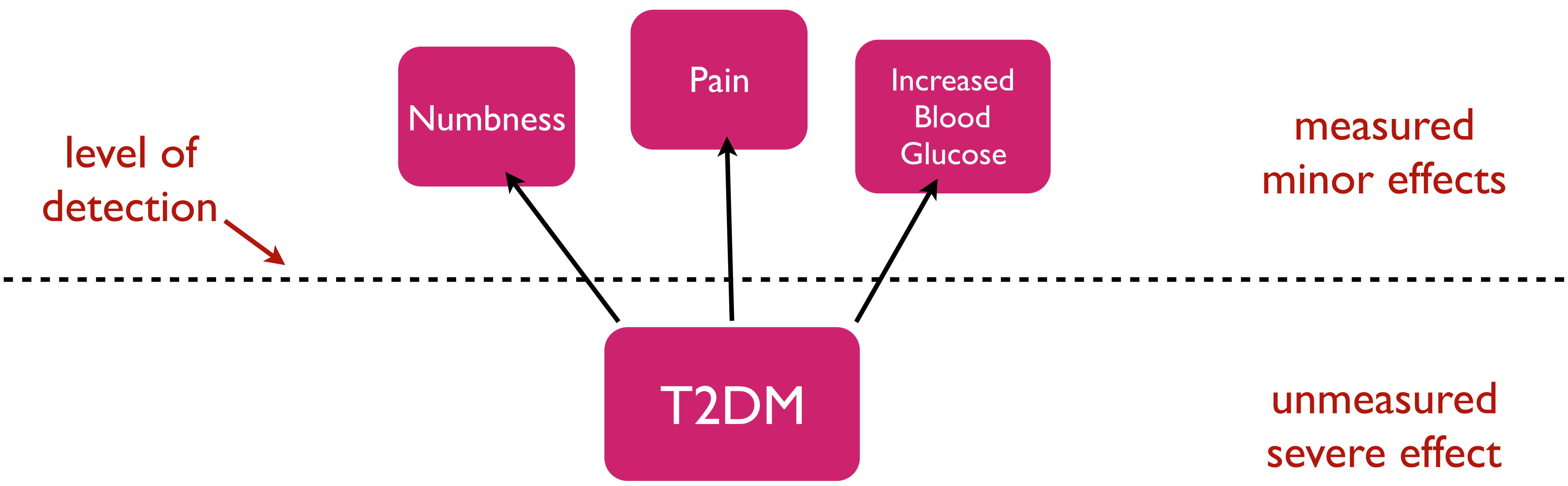


Severe ADE's can be identified by the presence of more minor (and more common) side effects

- First, identify the common side effects that are harbingers for the underlying severe AE
- Then, combine these side effects together to form an "effect profile" for an adverse event



Severe ADEs can be identified by the presence of more minor (and more common) side effects



DDI prediction validation

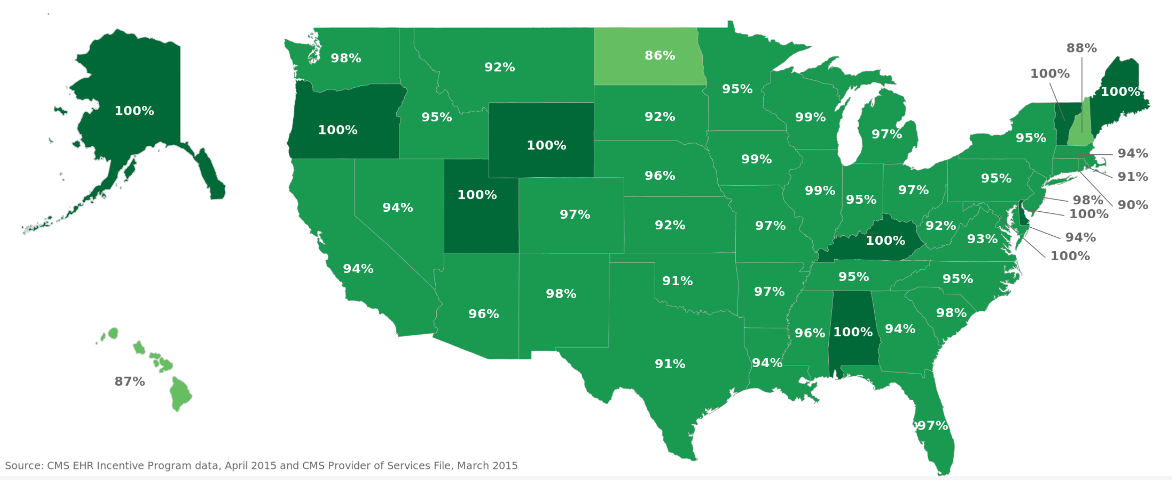
Table S3 Novel drug-drug interaction predictions for diabetes related adverse events.

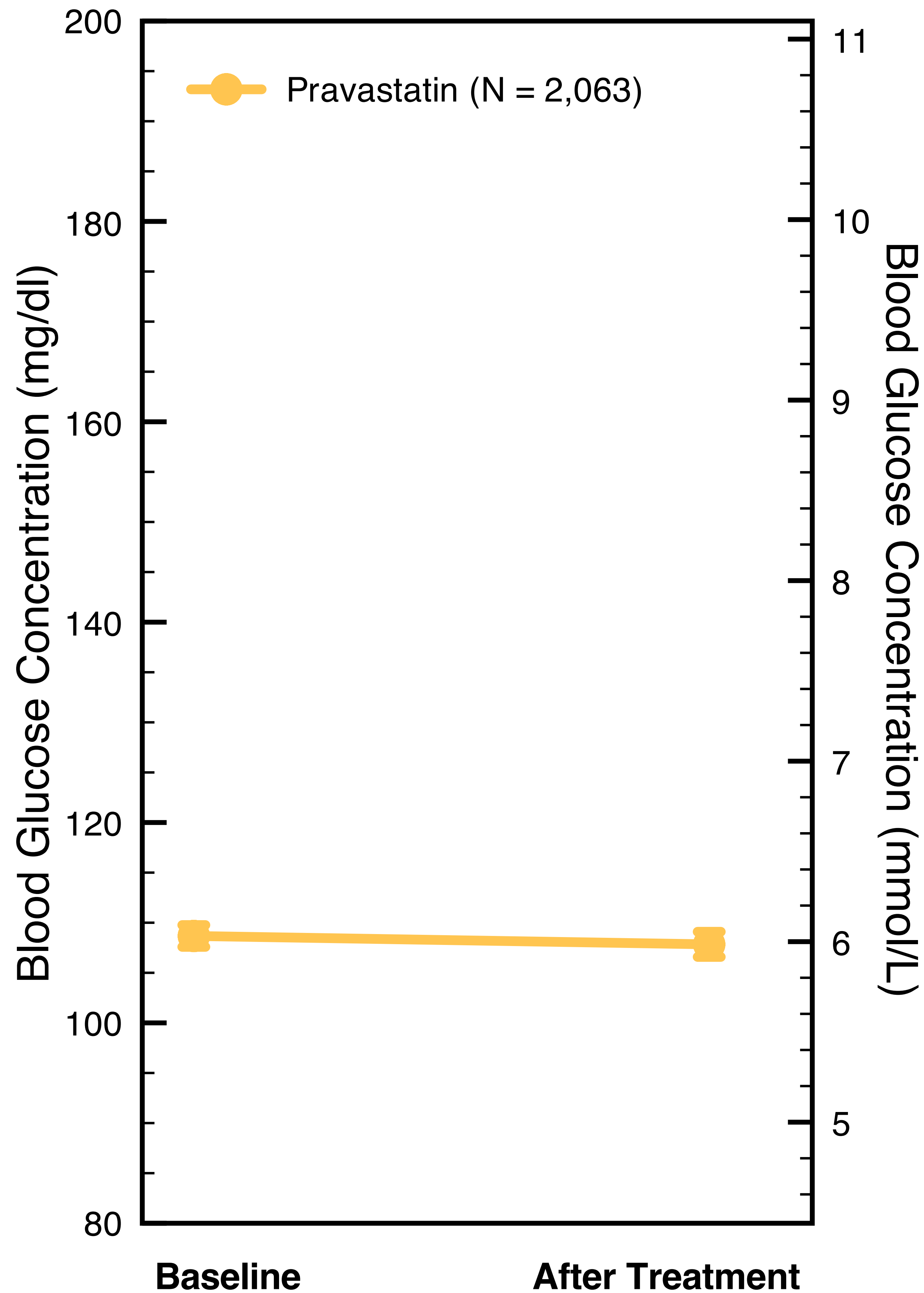
Rank	Drug A	Drug B	Score	Minimum Randomization Rank	Known DDI exists
38	PAROXETINE HCL	PRAVASTATIN SODIUM	11.35189601496	62	
72	DIOVAN HCT	HYDROCHLOROTHIAZIDE	7.1786599539	89	
94	CRESTOR	PREVACID	4.7923771645	148	
107	DESFERAL	EXJADE	3.97220625	129	
159	COUMADIN	VESICARE	0.8928376683	169	
160	DEXAMETHASONE	THALIDOMIDE	0.8928376683	168	CRITICAL
170	FOSAMAX	VOLTAREN	0.5033125	1138	
175	ALIMTA	DEXAMETHASONE	0.2442375	197	

- Focus on top hit from diabetes classifier
- paroxetine = depression drug, pravastatin = cholesterol drug
- Popular drugs, est. ~1,000,000 patients on this combination!

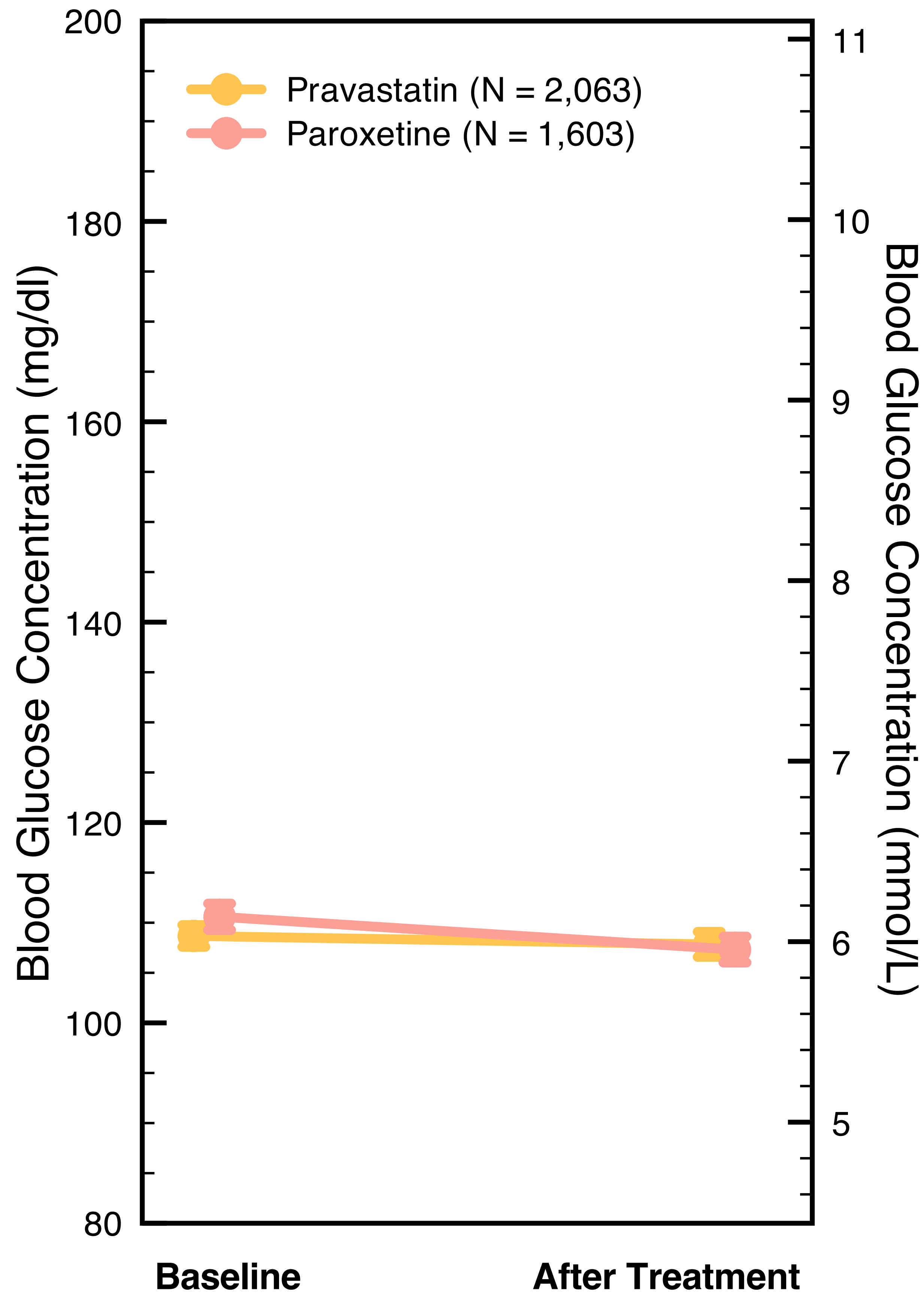
Analyzed blood glucose values for patients on either or both of these drugs

To the electronic health records…

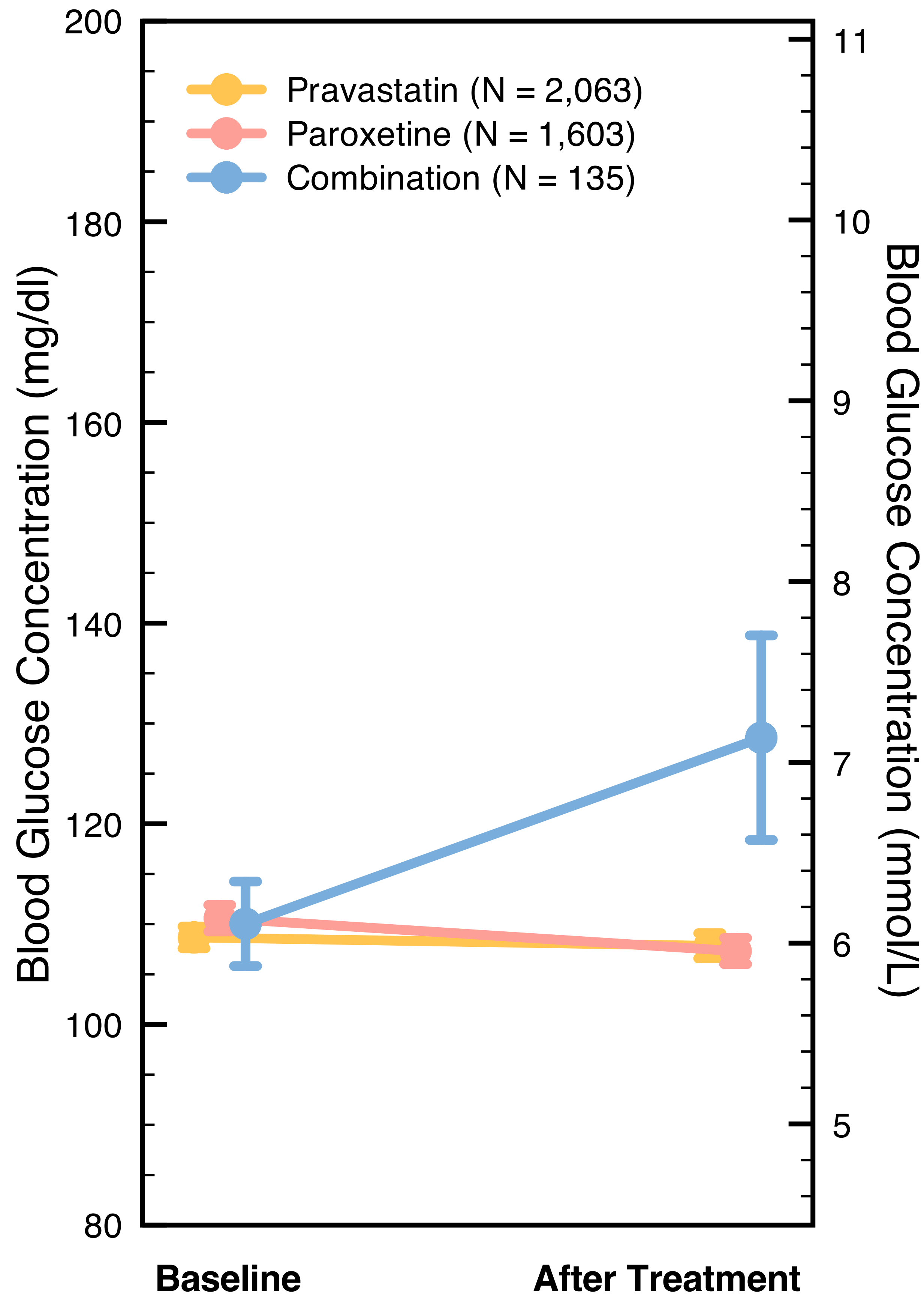




Tatonetti, et al. *Clinical Pharmacology & Therapeutics* (2011)



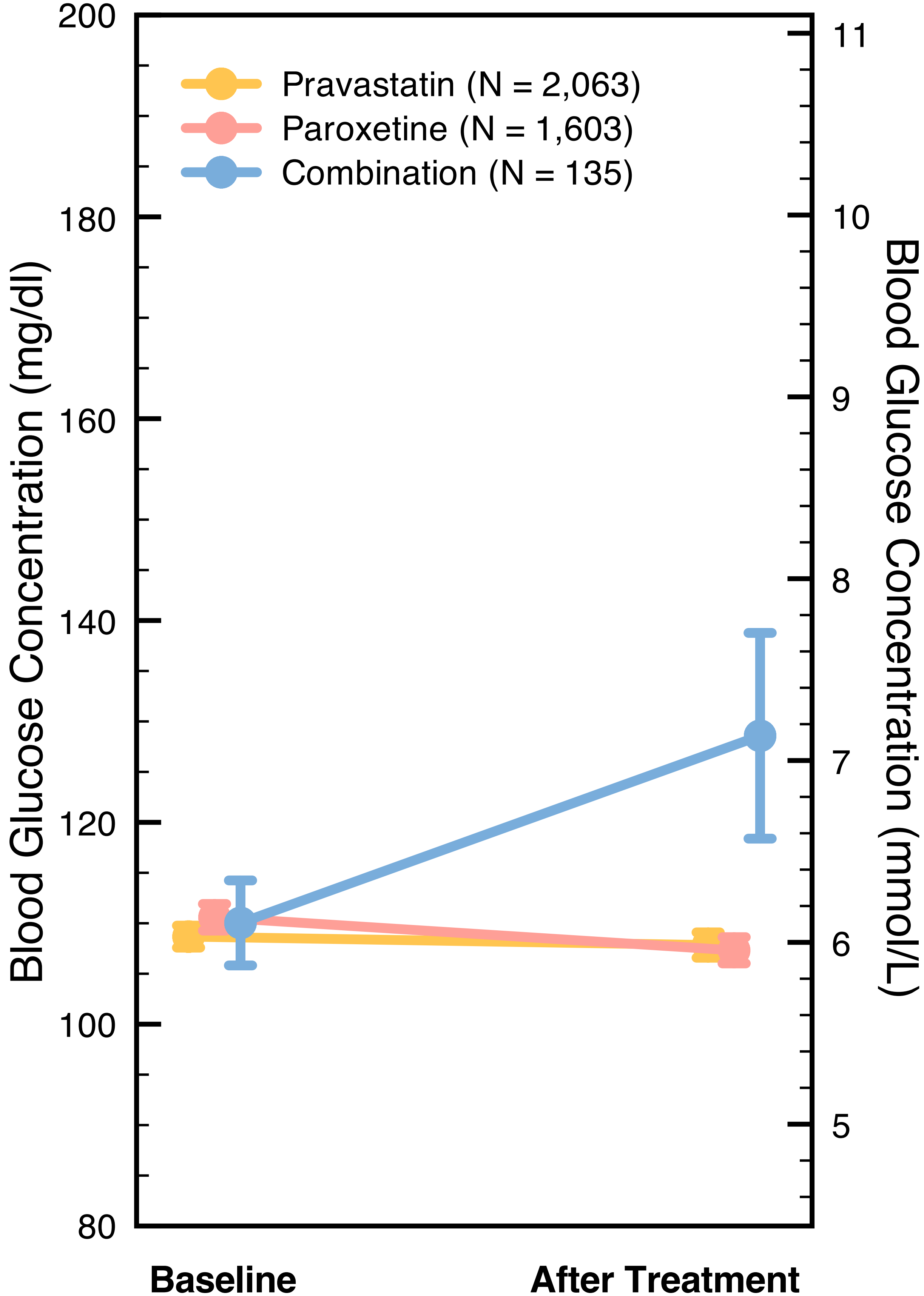
Tatonetti, et al. *Clinical Pharmacology & Therapeutics* (2011)



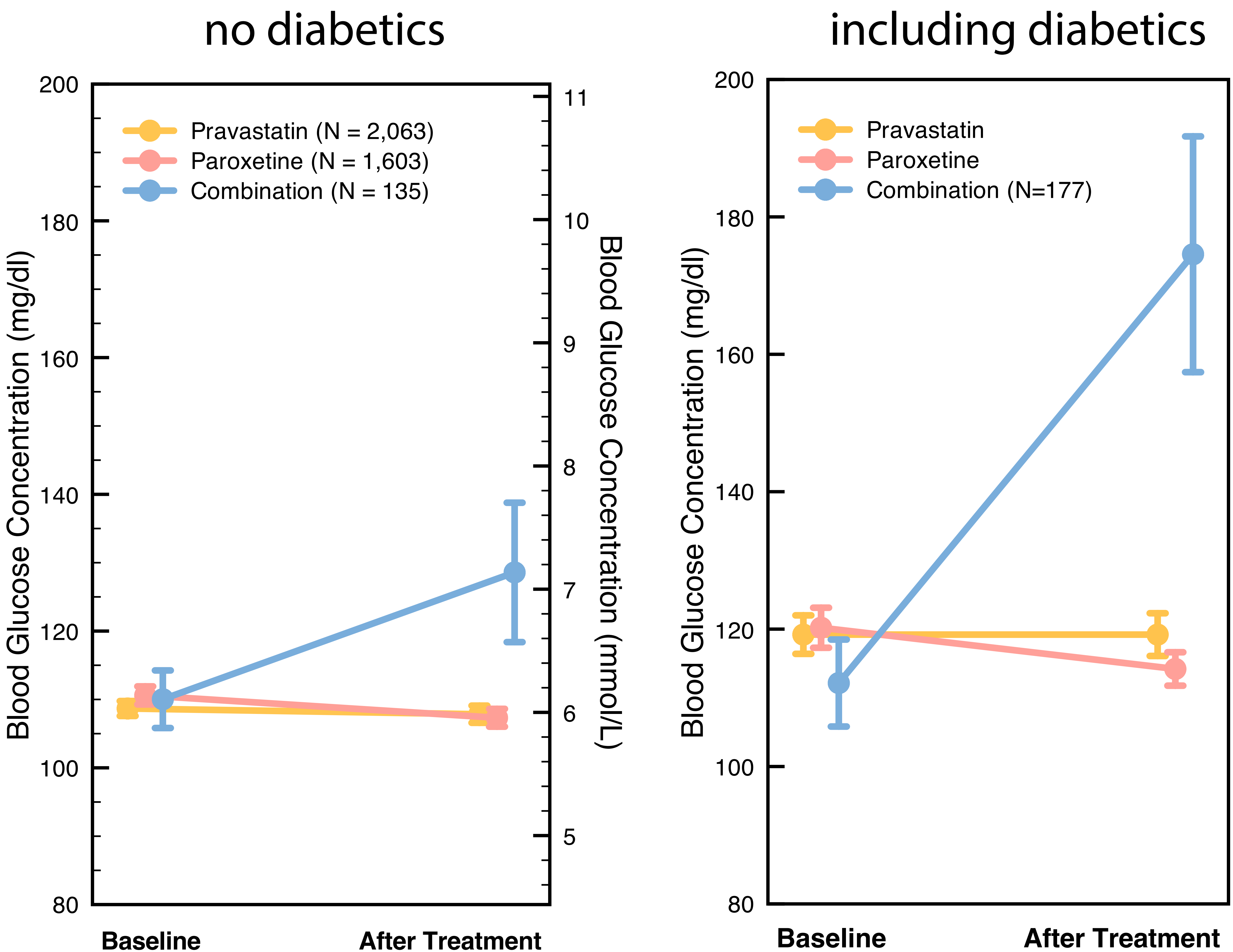
+18 mg/dl incr.
$p < 0.001$

Tatonetti, et al. *Clinical Pharmacology & Therapeutics* (2011)

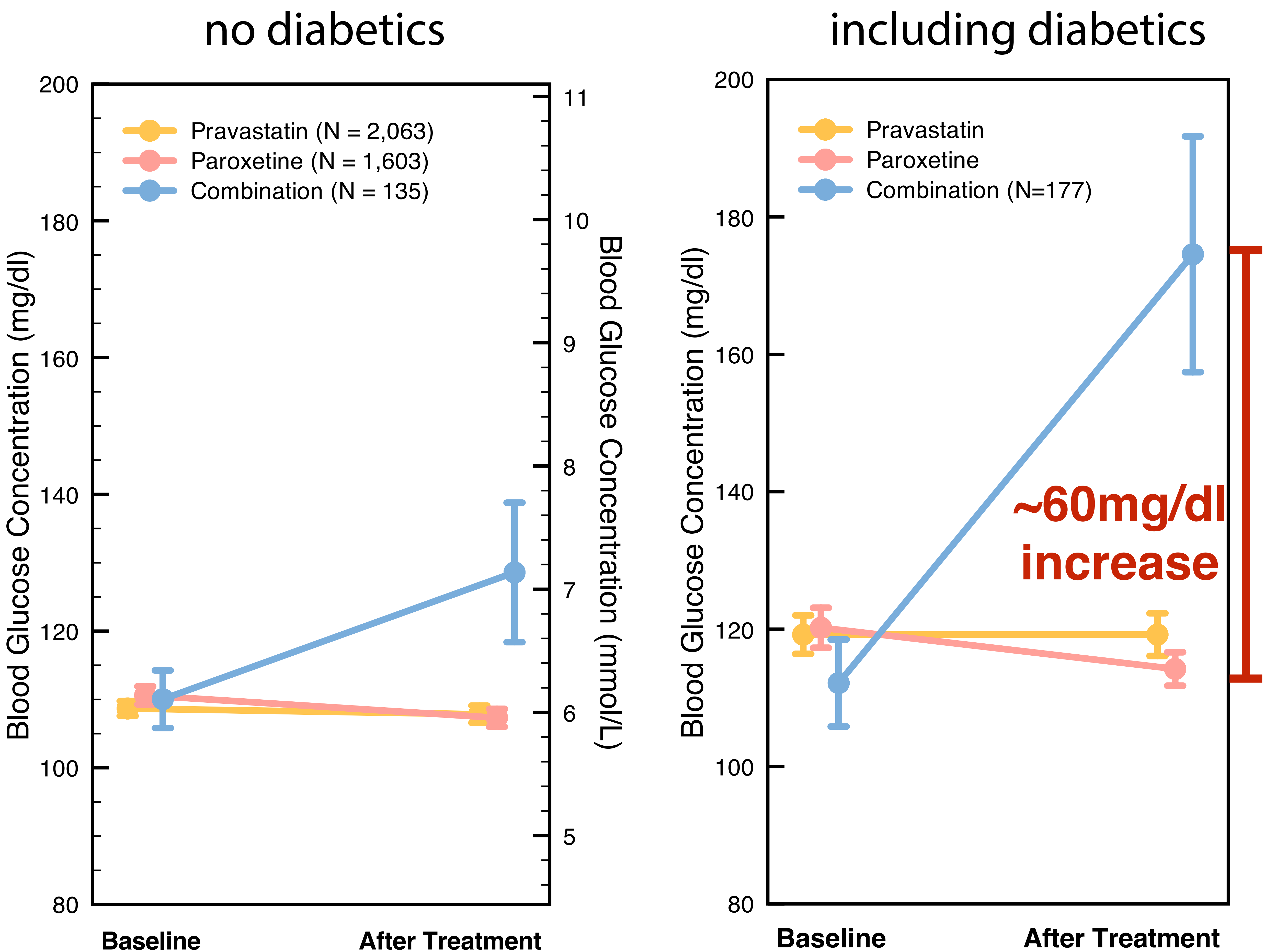
no diabetics



Tatonetti, et al. *Clinical Pharmacology & Therapeutics* (2011)



Tatonetti, et al. *Clinical Pharmacology & Therapeutics* (2011)



Tatonetti, et al. *Clinical Pharmacology & Therapeutics* (2011)

Informatics methods have taken us far, skeptics remain

Informatics methods have taken us far, skeptics remain

- Insulin Resistant Mouse Model

Informatics methods have taken us far, skeptics remain

- Insulin Resistant Mouse Model
  - 10 control mice on normal diet (Ctl Ctl)

Informatics methods have taken us far, skeptics remain

- Insulin Resistant Mouse Model
  - 10 control mice on normal diet (Ctl Ctl)
  - 10 control mice on high fat diet (HFD)

Informatics methods have taken us far, skeptics remain

- Insulin Resistant Mouse Model
    - 10 control mice on normal diet (Ctl Ctl)
    - 10 control mice on high fat diet (HFD)

Informatics methods have taken us far, skeptics remain

- Insulin Resistant Mouse Model
    - 10 control mice on normal diet (Ctl Ctl)
    - 10 control mice on $\underbrace{\text{high fat diet (HFD)}}_{\text{Simulating Pre-Diabetics}}$

Informatics methods have taken us far, skeptics remain

- Insulin Resistant Mouse Model
  - 10 control mice on normal diet (Ctl Ctl)
  - 10 control mice on $\underbrace{\text{high fat diet (HFD)}}_{\text{Simulating Pre-Diabetics}}$

Informatics methods have taken us far, skeptics remain

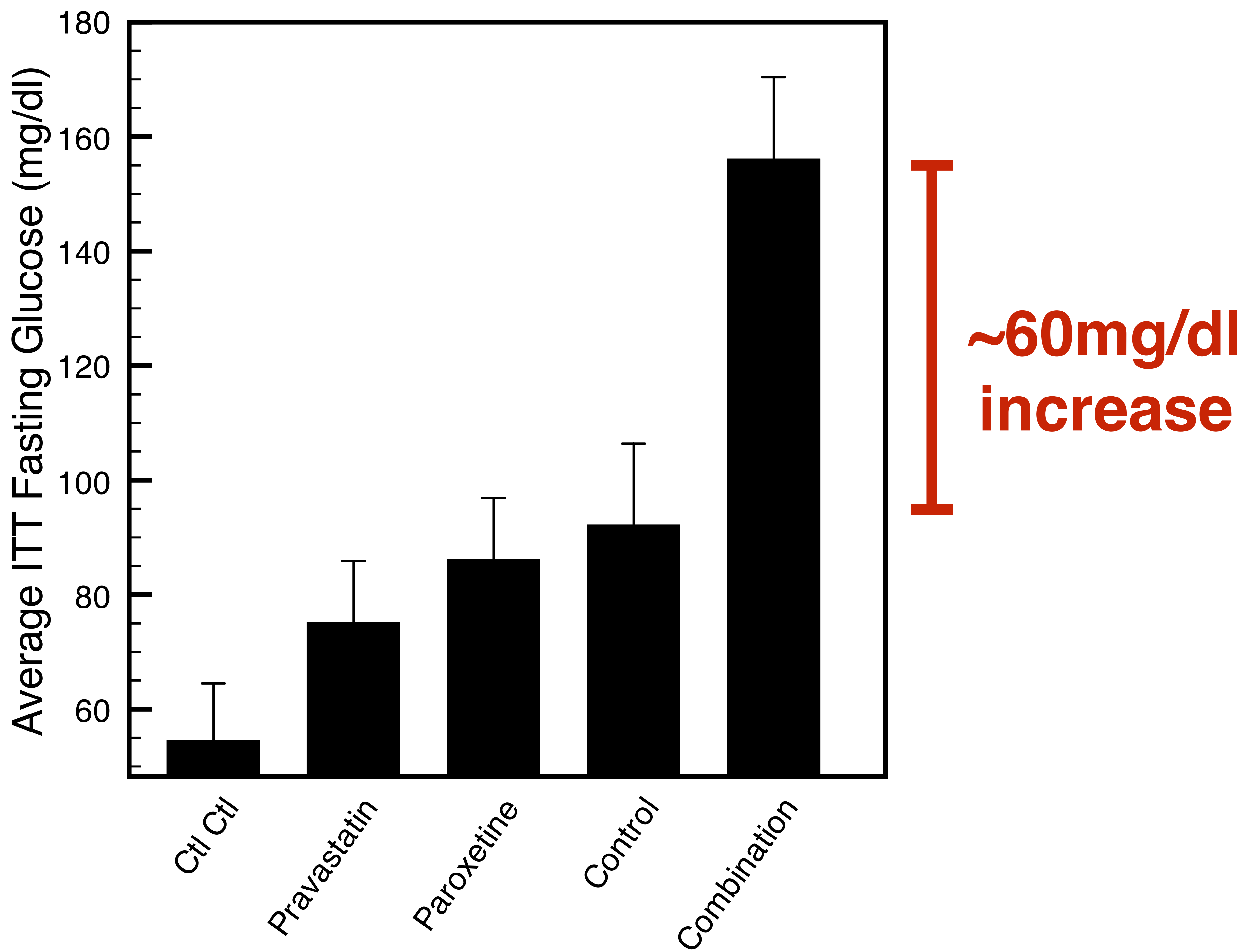
- Insulin Resistant Mouse Model
    - 10 control mice on normal diet (Ctl Ctl)
    - 10 control mice on high fat diet (HFD)

Simulating Pre-Diabetics

Informatics methods have taken us far, skeptics remain

- Insulin Resistant Mouse Model
  - 10 control mice on normal diet (Ctl Ctl)
  - 10 control mice on high fat diet (HFD)
  - 10 mice on pravastatin + HFD
  - 10 mice on paroxetine + HFD
  - 10 mice on combination + HFD

Summary of fasting glucose levels

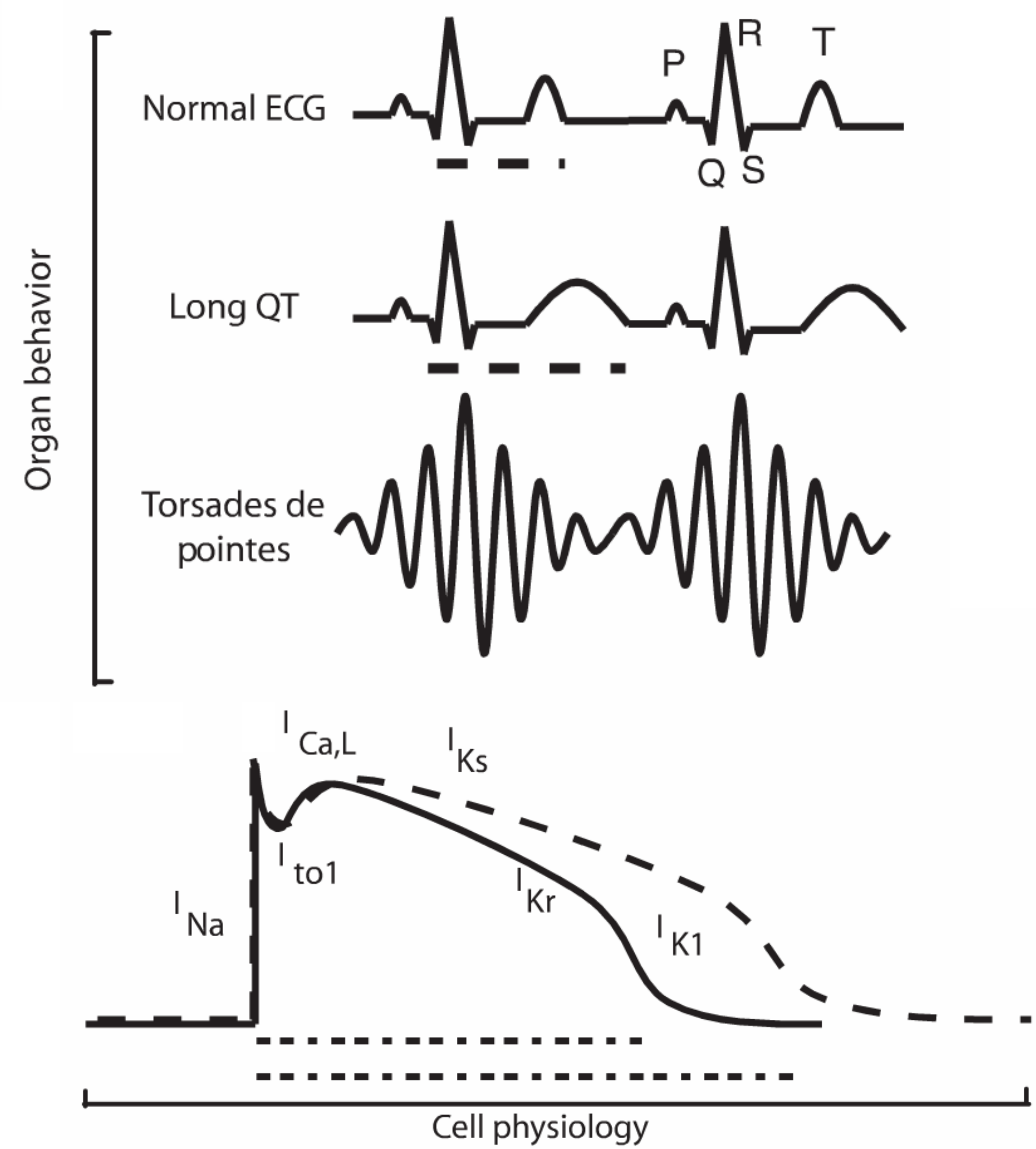


Replication is vital to science

- In biology we would never trust a result that hasn't been replicated
- Why should **algorithms** be any different?

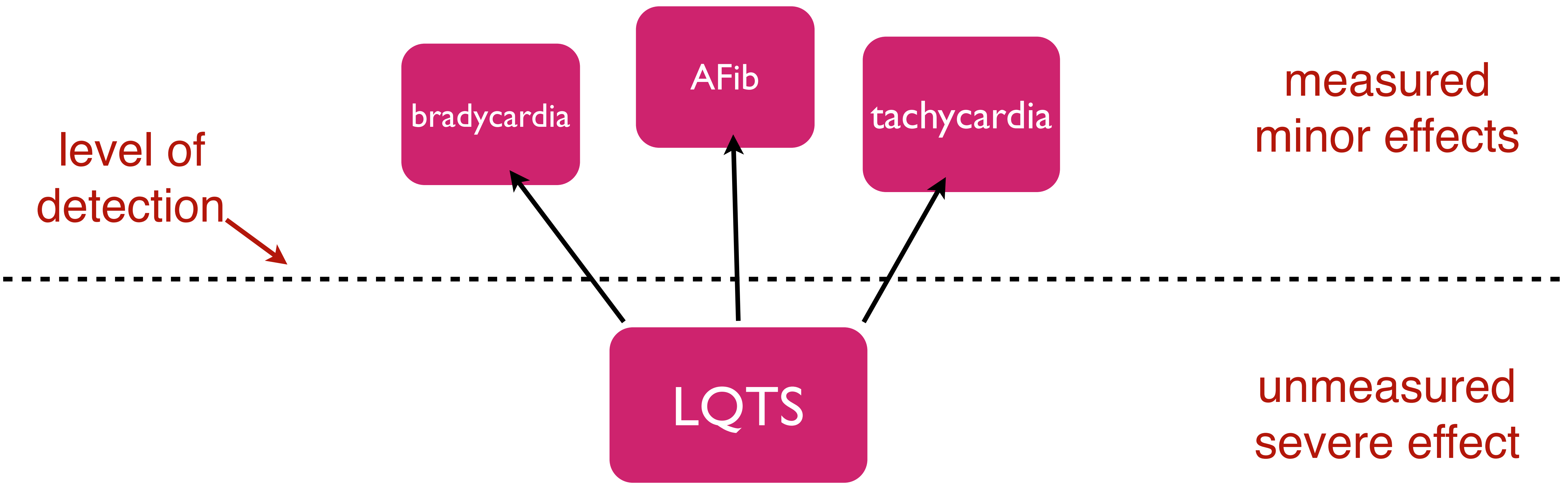
Drug-drug interactions and acquired Long QT Syndrome (LQTS)

- Long QT syndrome (LQTS): congenital or drug-induced change in electrical activity of the heart that can lead to potentially fatal arrhythmia: *torsades de pointes* (TdP)
- 13 genes associated with congenital LQTS
- Drug-induced LQTS usually caused by blocking the hERG channel (*KCNH2*)



From Berger et al., *Science Signaling* (2010)

Identify acquired LQTS drug-drug interactions using Latent Signal Detection



Lorberbaum, et al. Drug Safety (2016)

Latent Signal Detection of acquired LQTS

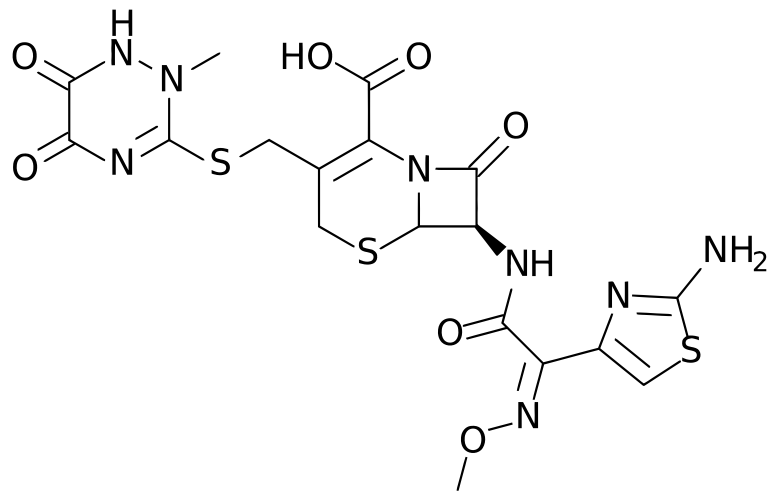
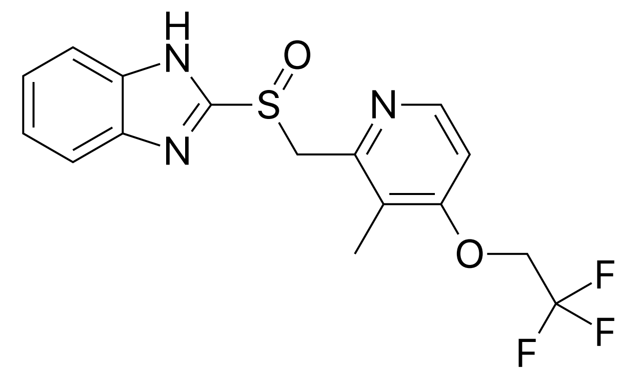
Top Prediction:
Ceftriaxone + Lansoprazole

- Ceftriaxone — common in-patient cephalosporin antibiotic
- Lansoprazole — proton-pump inhibitor used to treat GERD, one of the most commonly taken drugs in the world
- In the EHR: Patients on the combination have QT intervals 11ms longer, on average and are **1.5X as likely to have a QT interval > 500ms**

	White	Black/African American	Other, including Hispanic	Asian
Females	11.1 ± 3.1 ms** (N=220)	-1.3 ± 7.4 ms (N=91)	6.0 ± 4.9 ms (N=78)	13.2 ± 4.8 ms (N=4)
Males	15.1 ± 4.1 ms** (N=164)	0.7 ± 7.2 ms (N=53)	10.5 ± 6.6 ms (N=46)	8.3 ± 12.5 ms (N=4)

** $p < 0.01$, one sample Student's T test

Lorberbaum, et al. Drug Safety (2016)
Lorberbaum, et al. JACC (In press)

- Predicted QT-DDI: **ceftriaxone** (cephalosporin antibiotic) and **lansoprazole** (proton pump inhibitor)

- Neither drug alone has any evidence of QT prolongation/ hERG block

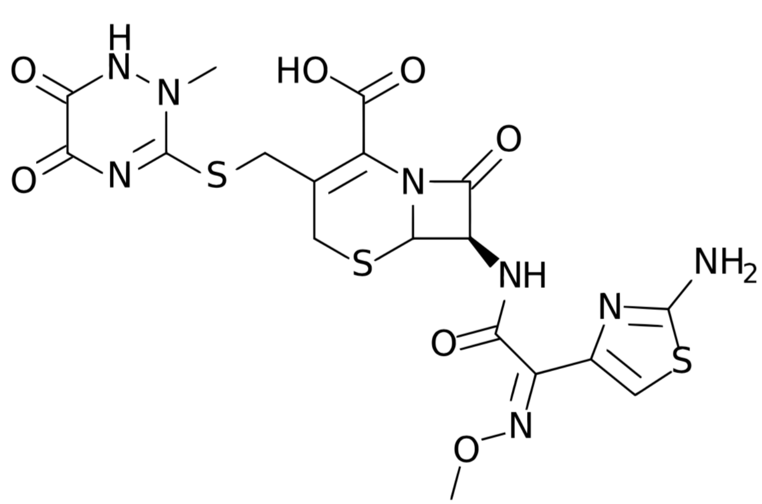
- Predicted QT-DDI: **ceftriaxone** (cephalosporin antibiotic) and **lansoprazole** (proton pump inhibitor)

- Predicted QT-DDI: **ceftriaxone** (cephalosporin antibiotic) and **lansoprazole** (proton pump inhibitor)

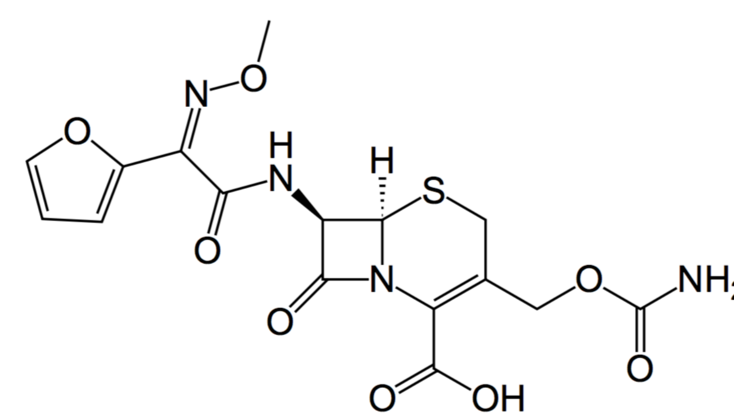
- <u>Negative control</u>: lansoprazole + **cefuroxime** (another cephalosporin) – no evidence in FAERS of an interaction

- Predicted QT-DDI: **ceftriaxone** (cephalosporin antibiotic) and **lansoprazole** (proton pump inhibitor)

- <u>Negative control</u>: lansoprazole + **cefuroxime** (another cephalosporin) – no evidence in FAERS of an interaction



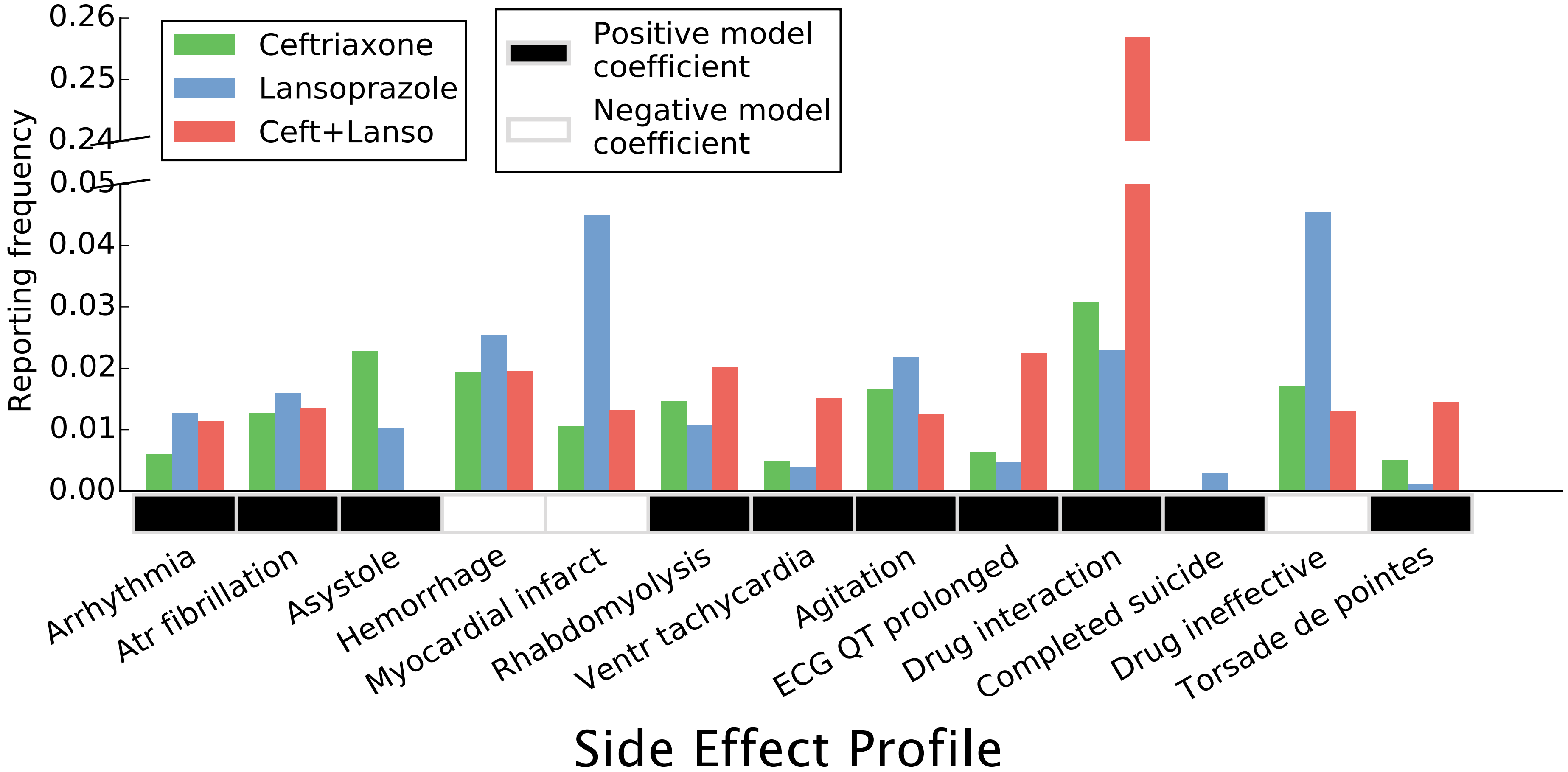
Ceftriaxone



Cefuroxime

FAERS

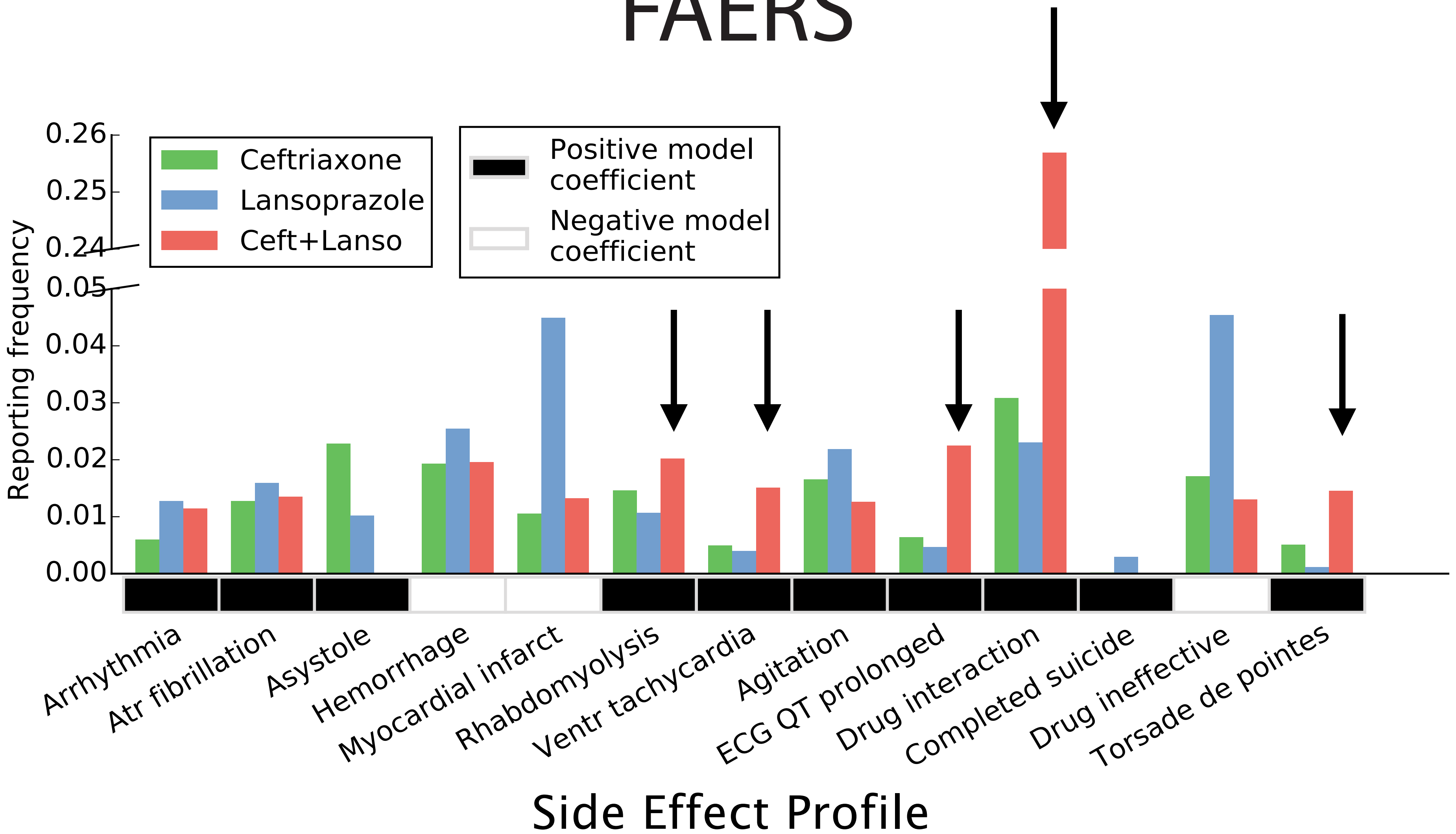
Ceftriaxone+ Lansoprazole



Lorberbaum, et al. In Revision

FAERS

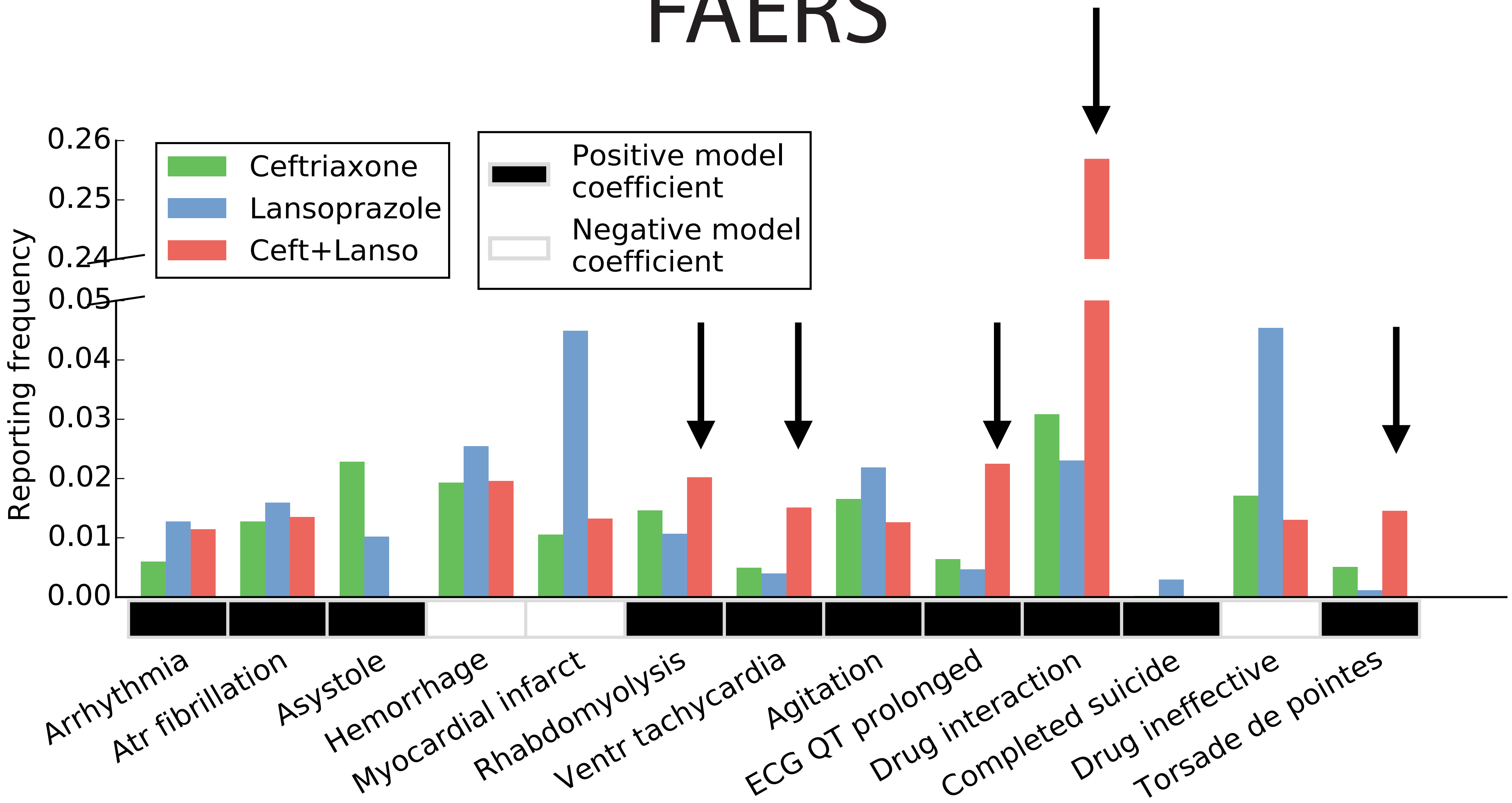
Ceftriaxone+
Lansoprazole



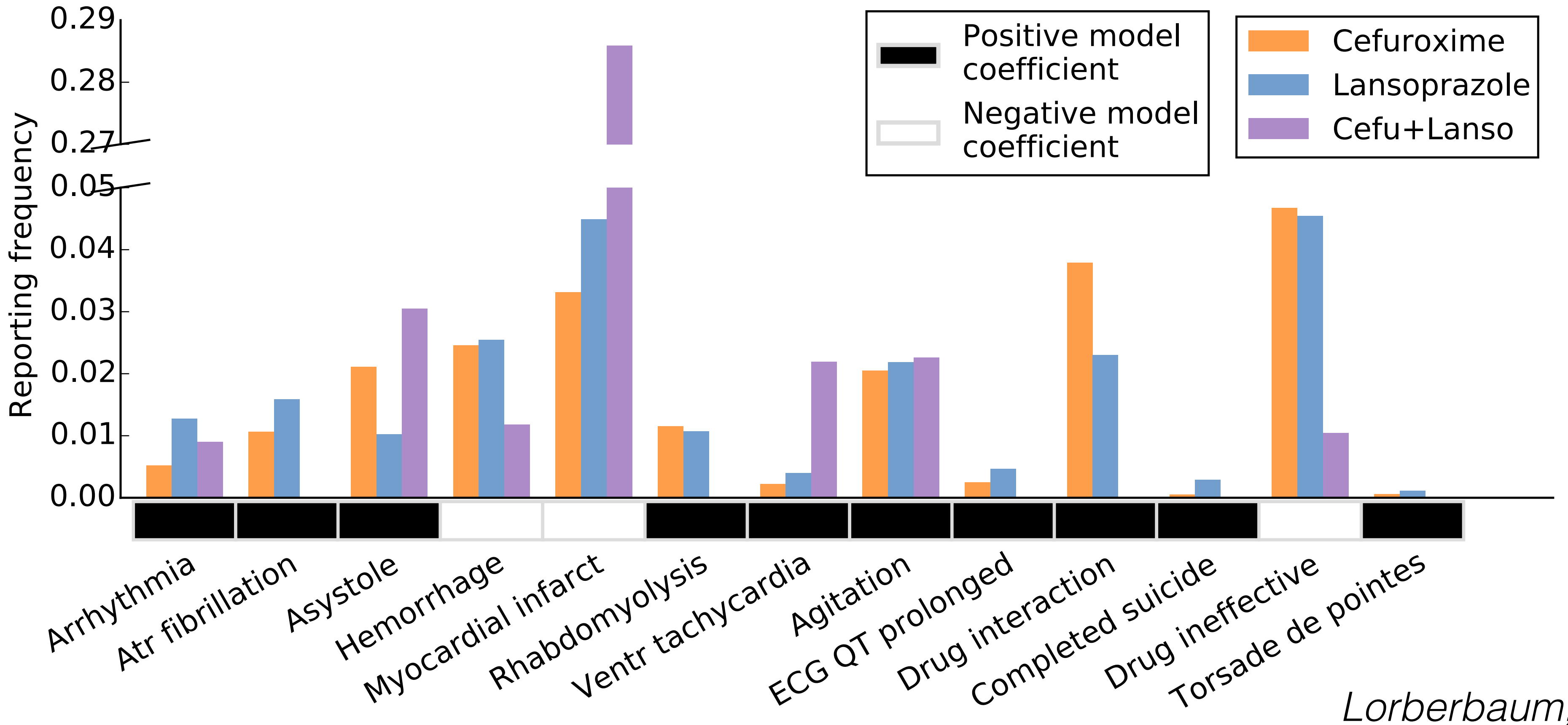
Lorberbaum, et al. In Revision

FAERS

Ceftriaxone+ Lansoprazole

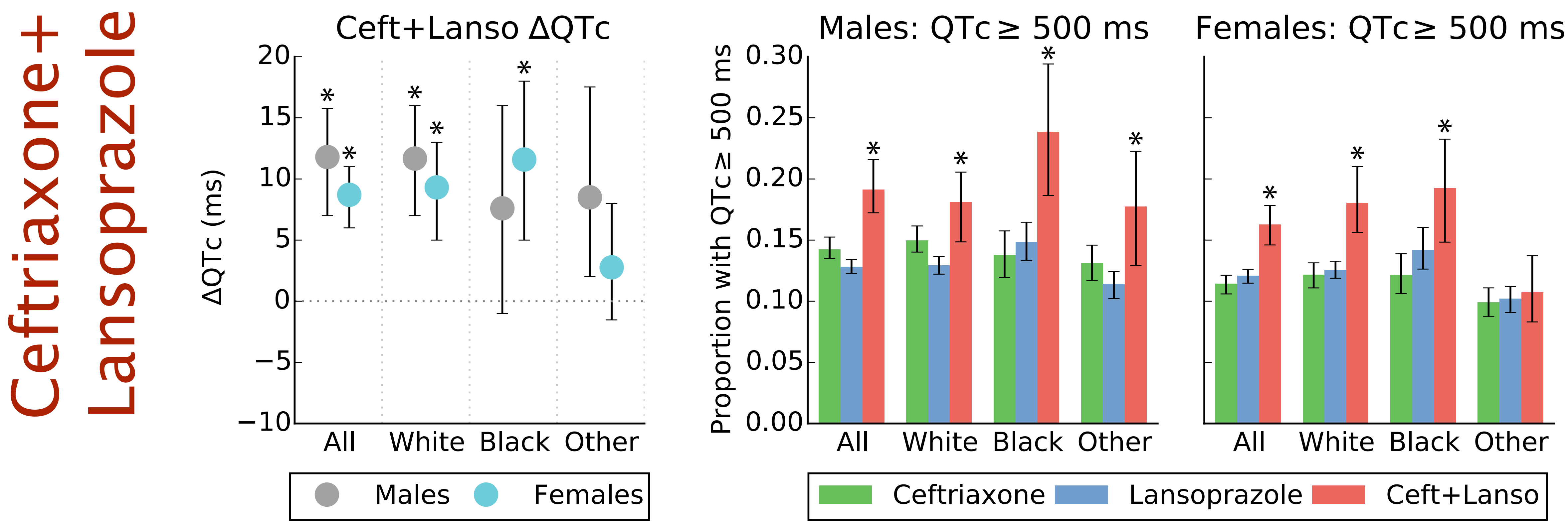


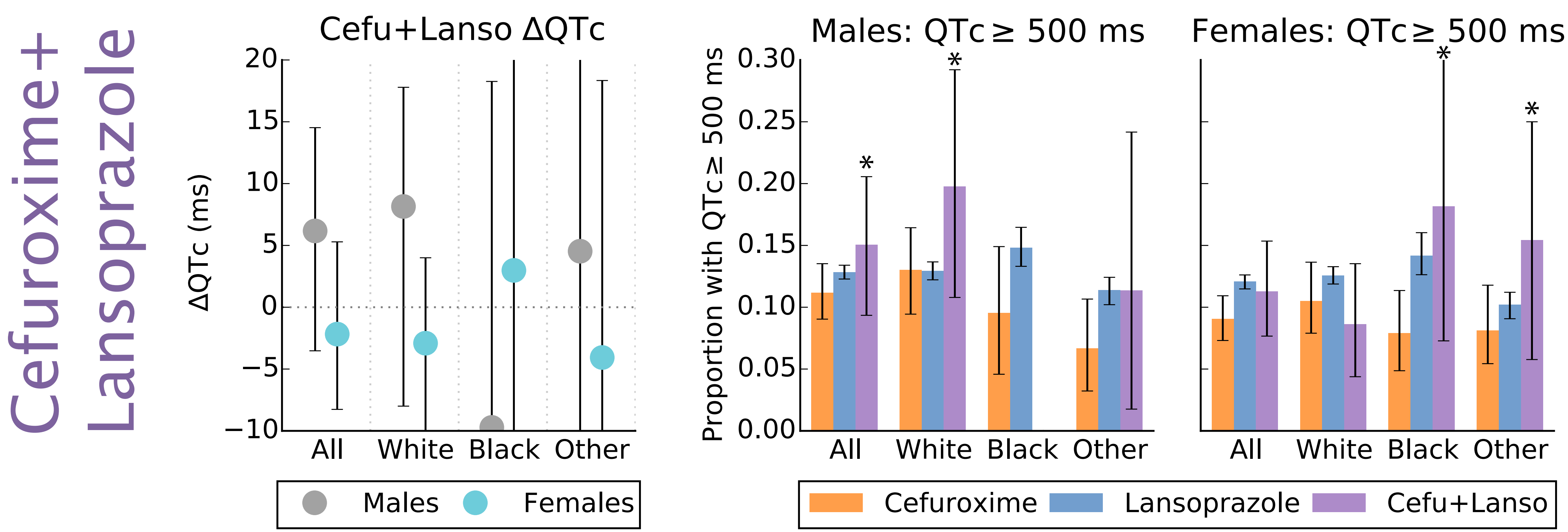
Cefuroxime+ Lansoprazole



Lorberbaum, et al. In Revision

Electronic Health Records

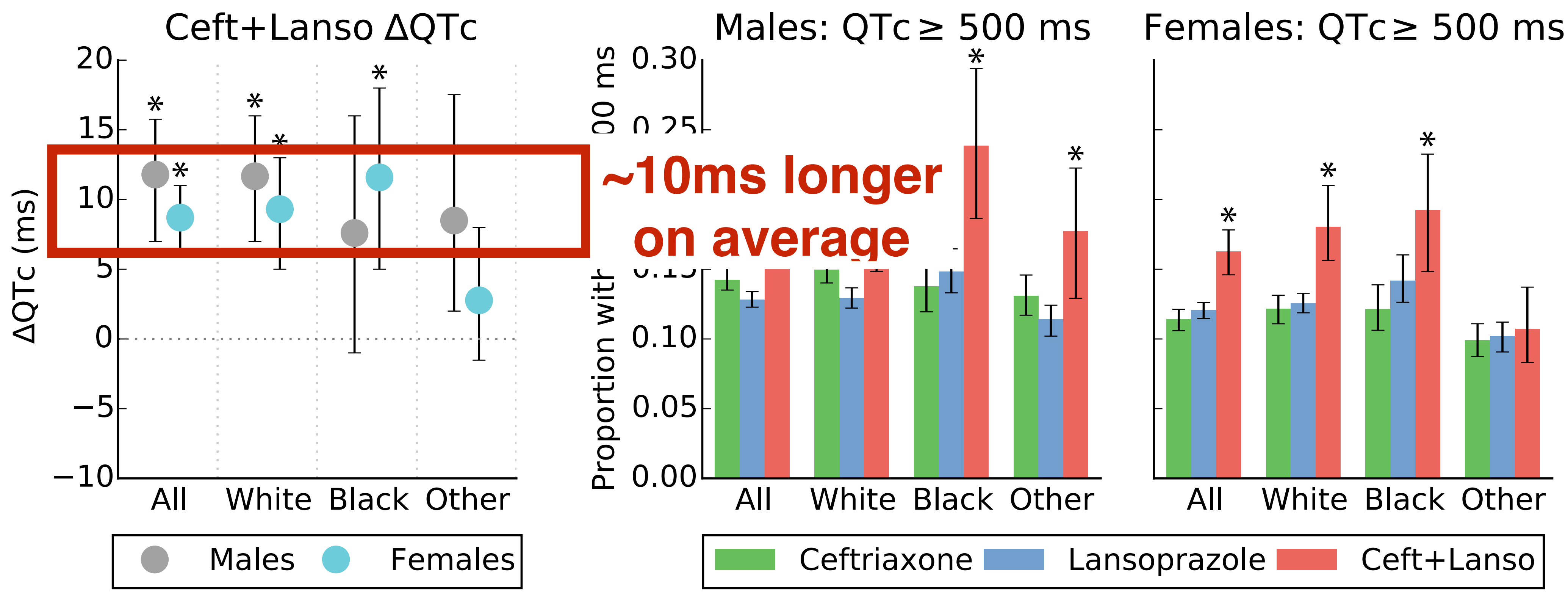




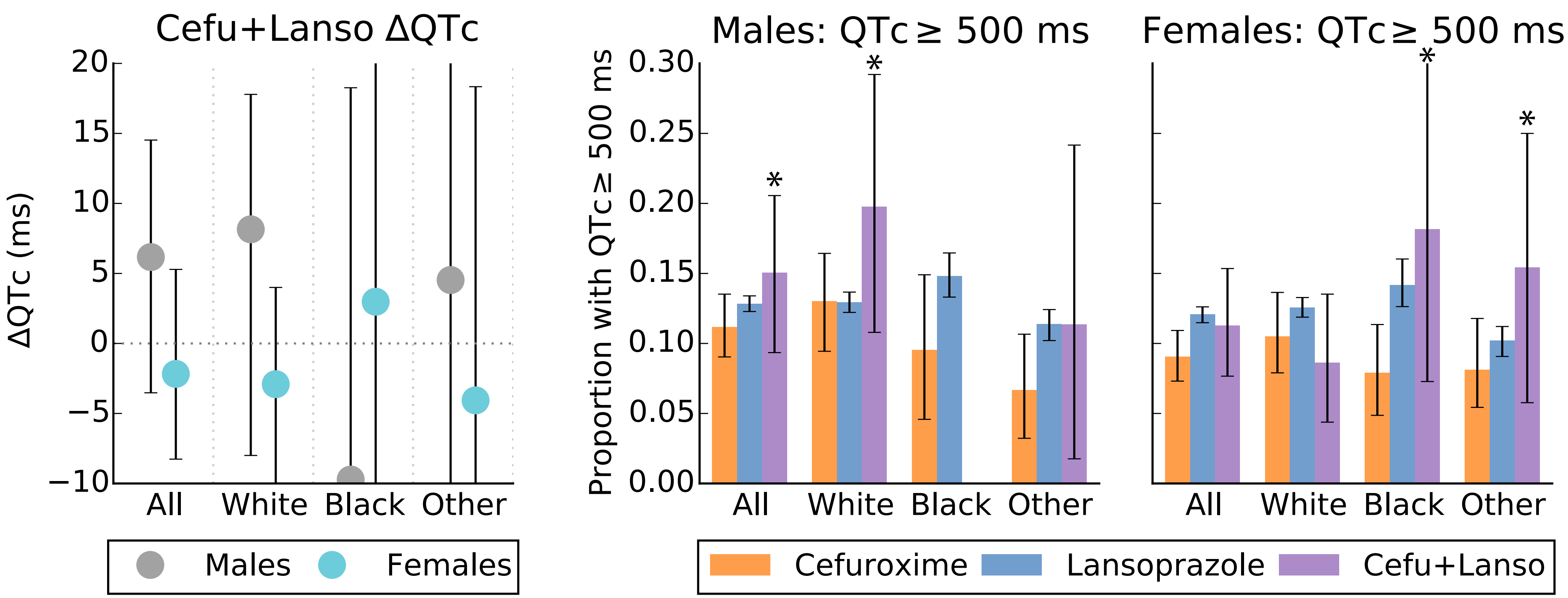
Lorberbaum, et al. In Revision

Electronic Health Records

Ceftriaxone+ Lansoprazole



Cefuroxime+ Lansoprazole



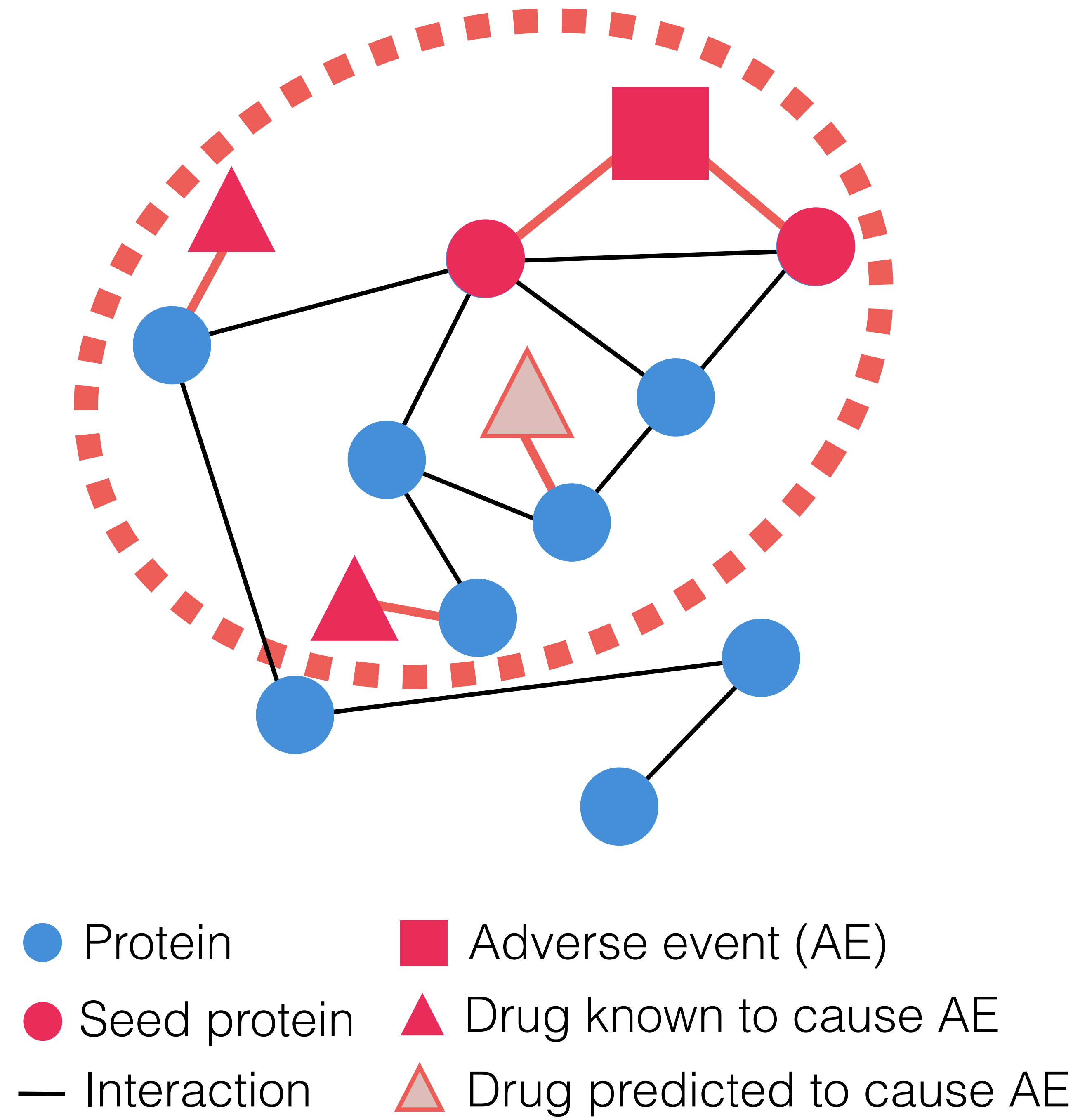
Lorberbaum, et al. In Revision

What is the mechanism?

MADSS

Modular Assembly of Drug Safety Subnetworks

- Use network analysis to build AE neighborhoods: a subset of the interactome surrounding AE "seed" proteins
- Score each protein on connectivity to seeds using:
    - Mean first passage time
    - Betweenness centrality
    - Shared neighbors
    - Inverse shortest path
- Overarching hypothesis: drugs targeting proteins within an AE neighborhood more likely to be involved in mediating that AE

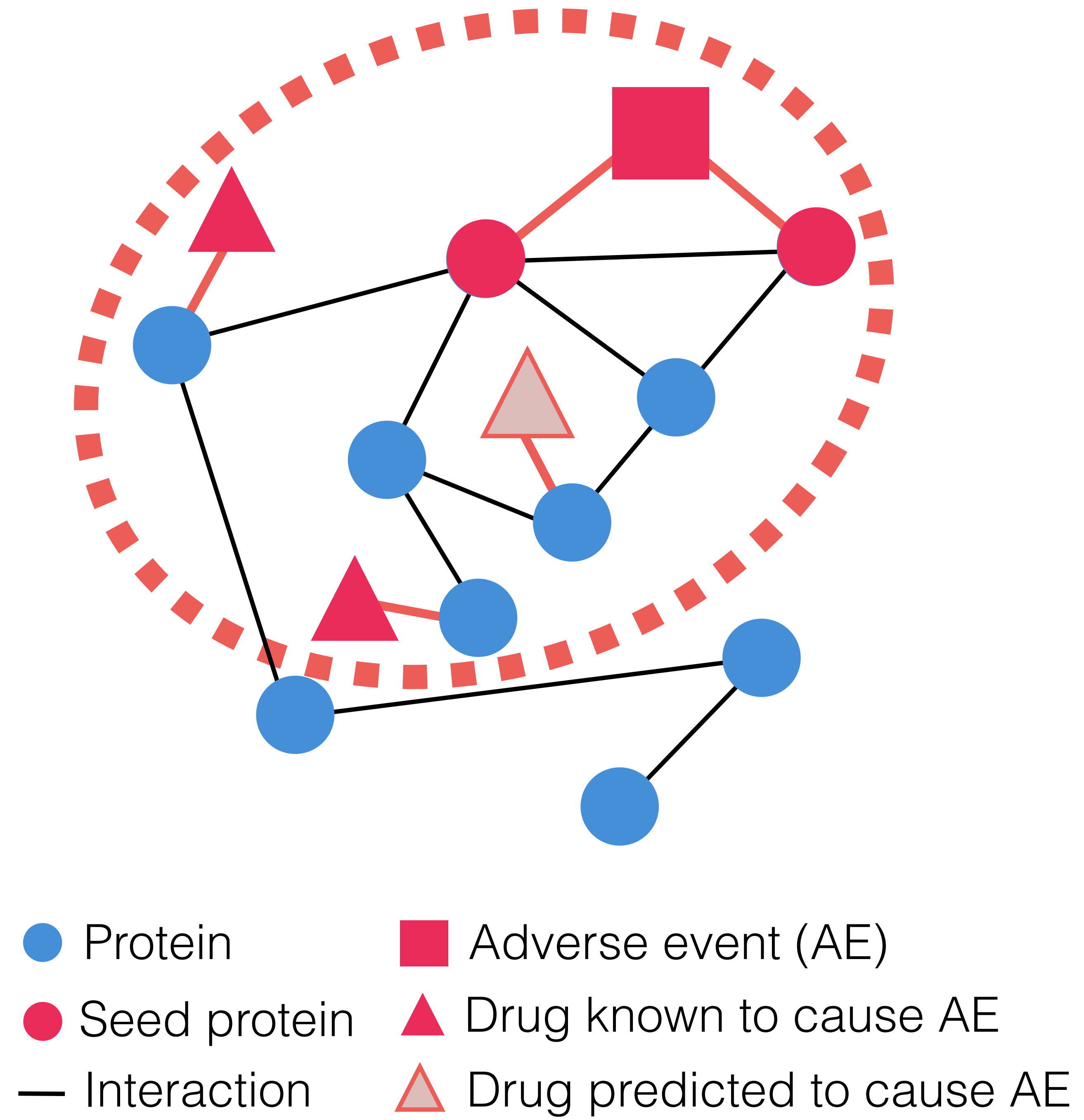


Lorberbaum, et al. *Clin. Pharmacol. Ther.* (2015)

MADSS

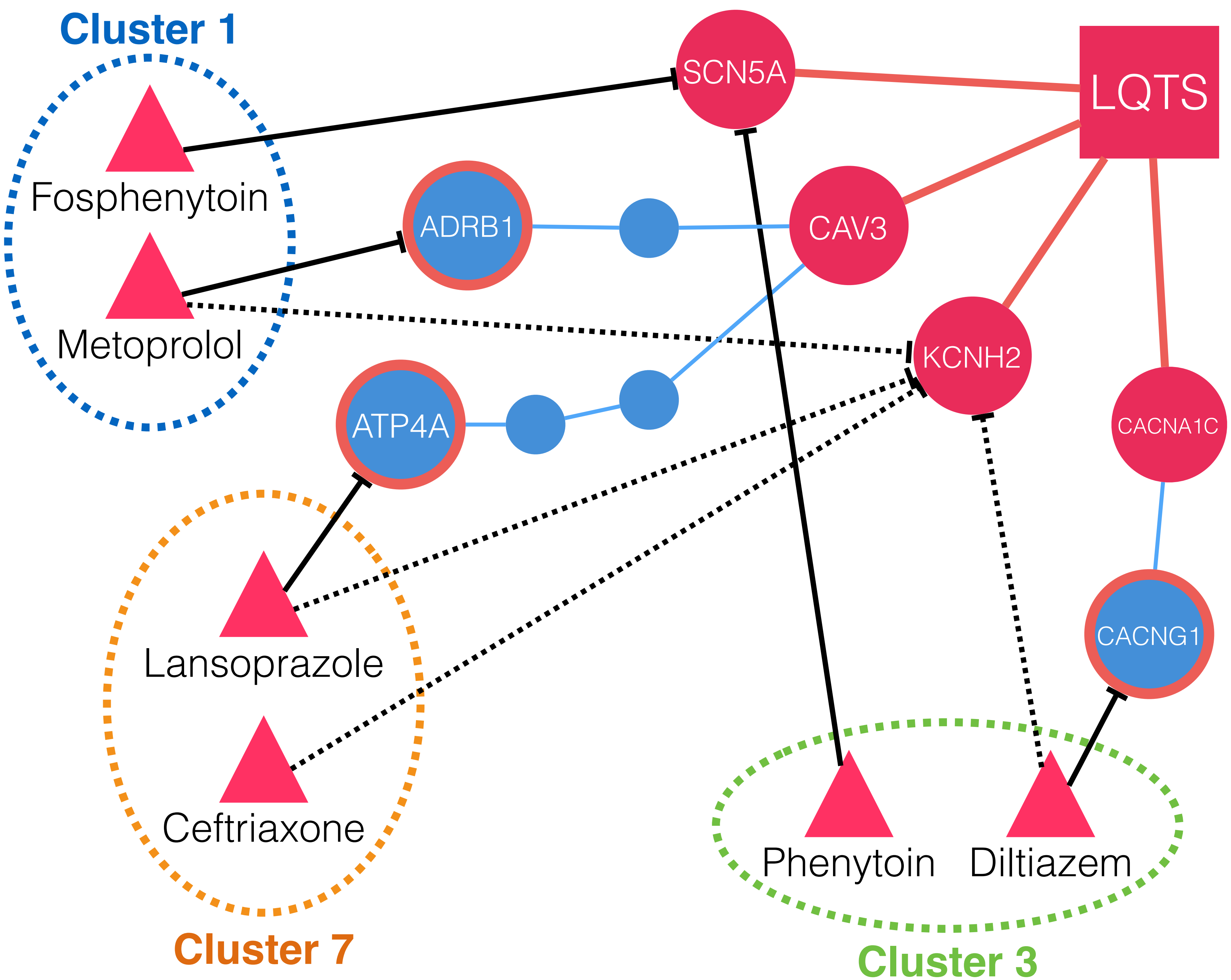
Modular Assembly of Drug Safety Subnetworks

- Use network analysis to build AE neighborhoods: a subset of the interactome surrounding AE “seed” proteins
- Score each protein on connectivity to seeds using:
  - Mean first passage time
  - Betweenness centrality
  - Shared neighbors
  - Inverse shortest path
- Overarching hypothesis: drugs targeting proteins within an AE neighborhood more likely to be involved in mediating that AE
- **Ran MADSS using 13 LQTS genes as seeds**



Lorberbaum, et al. *Clin. Pharmacol. Ther.* (2015)

Putative mechanisms of QT-DDIs

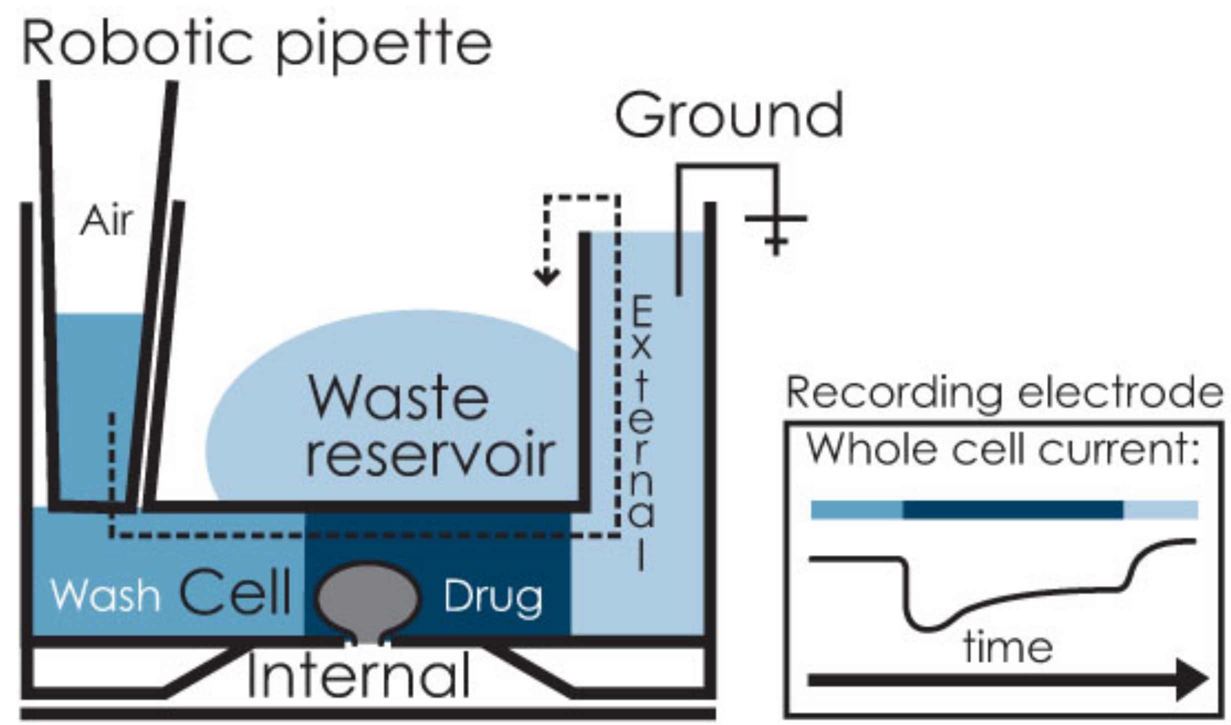


Known drug-target binding (DrugBank)

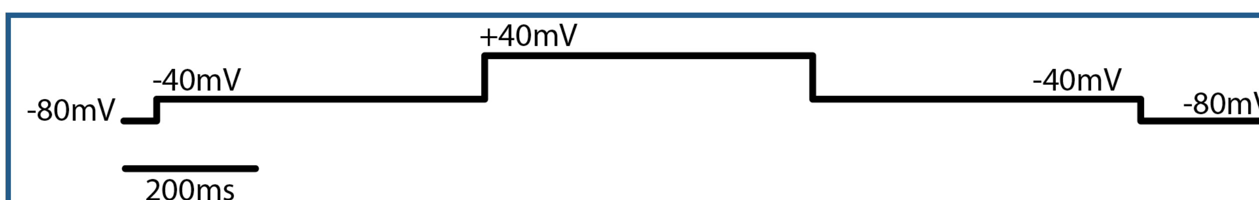
Predicted drug-hERG binding (Random Forest classifier)

Automated Patch Clamp

- Collaboration with Rocky Kass (CUMC Pharmacology Dept.)
- Take HEK293 cells over-expressing the hERG channel
- Perform a single-cell patch clamp experiment
  - control
  - ceftriaxone alone
  - lansoprazole alone
  - combination of ceftriaxone and lansoprazole



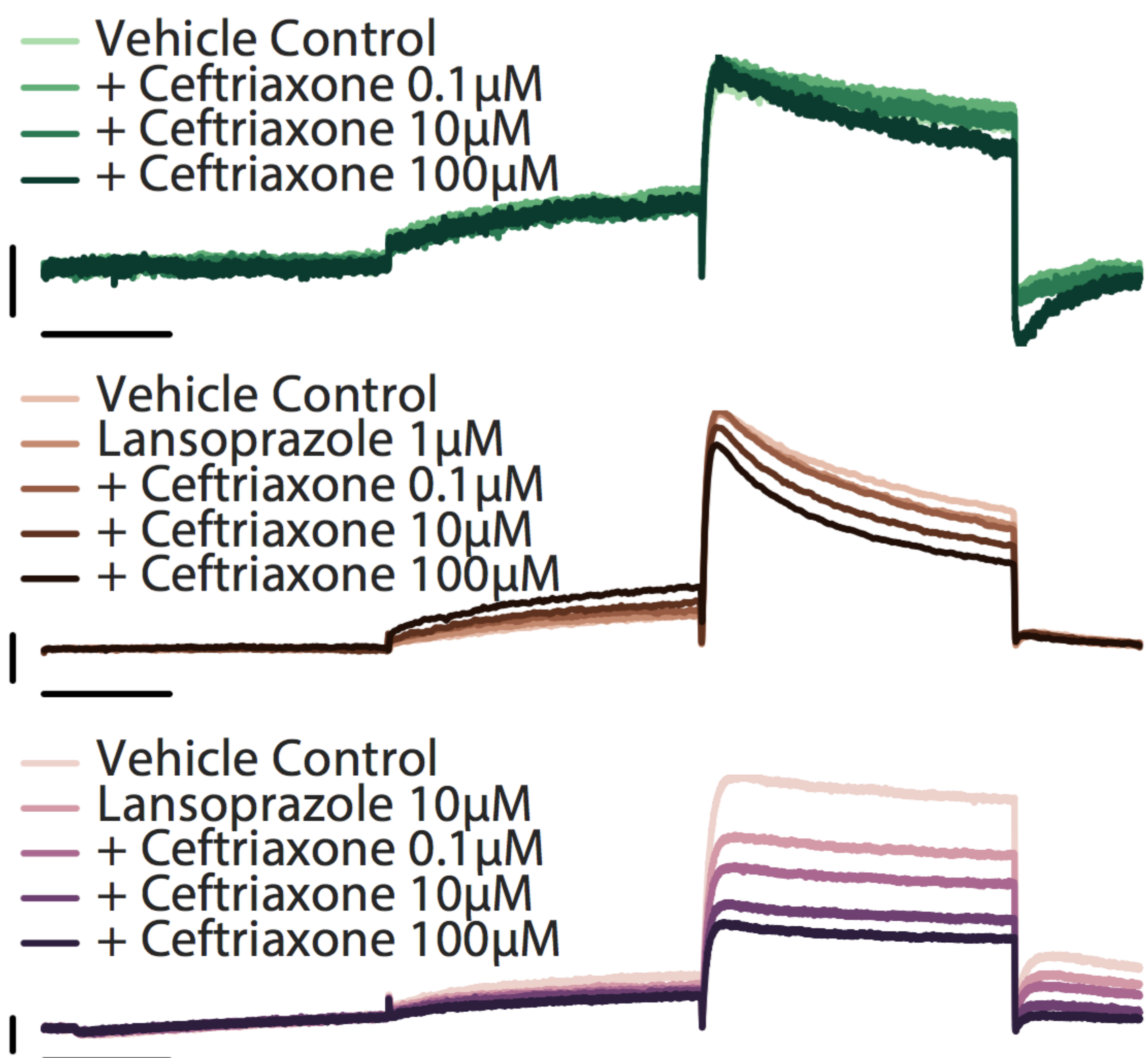
Nanion Patchliner



Voltage protocol: step to +40mV followed by a return to -40mV

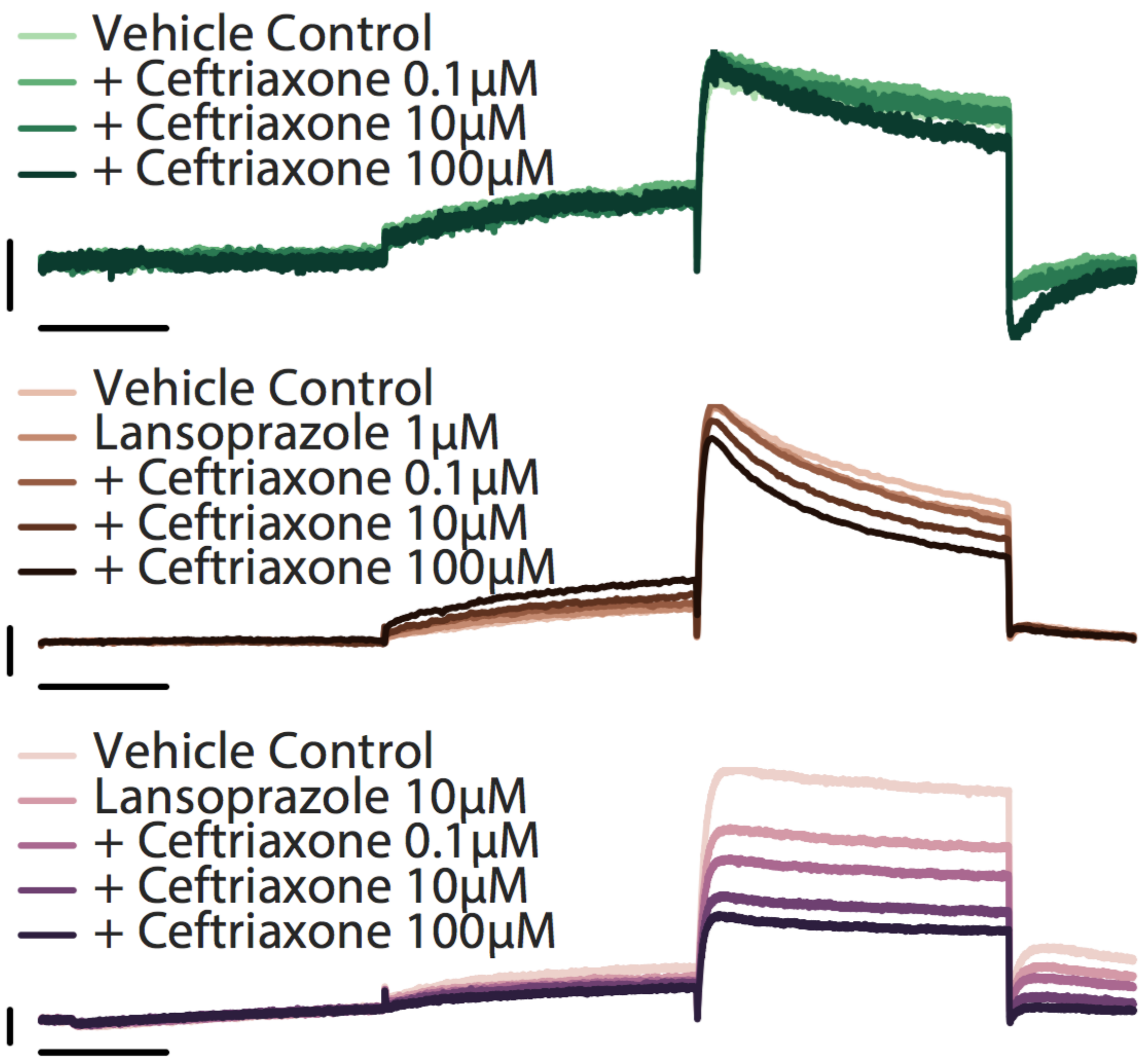
Lorberbaum, et al. JACC (In press)

Ceftriaxone+Lansoprazole

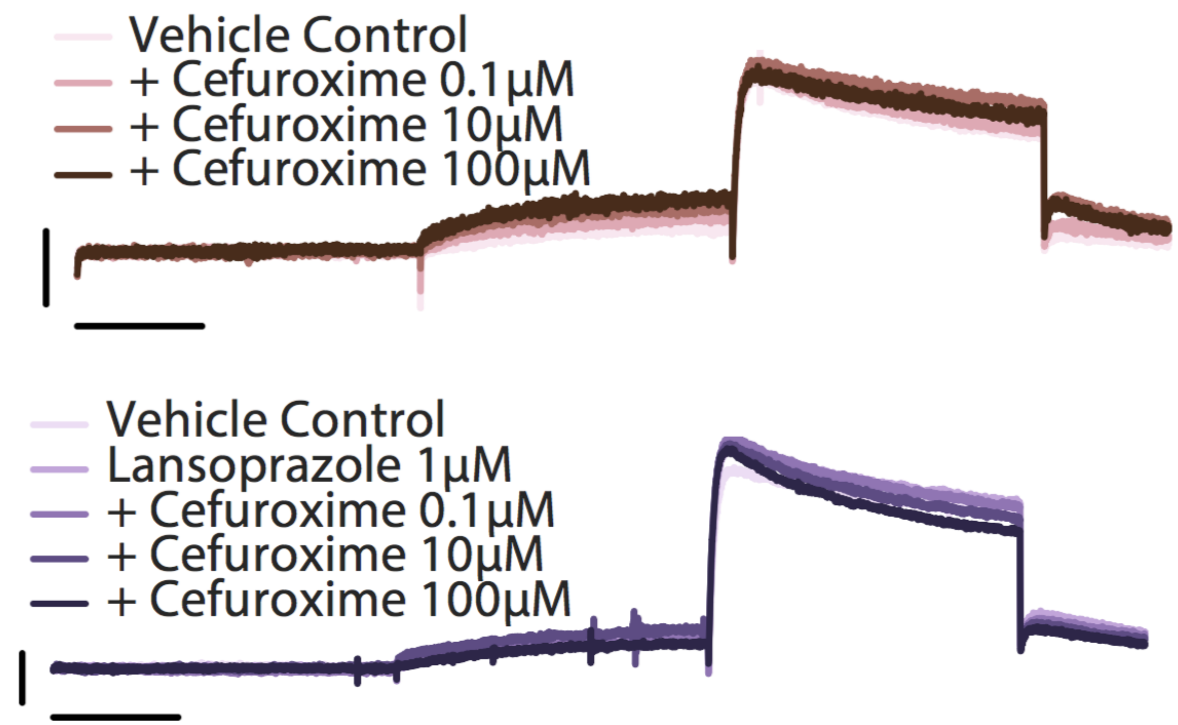


Lorberbaum, et al. JACC (In press)

Ceftriaxone+Lansoprazole

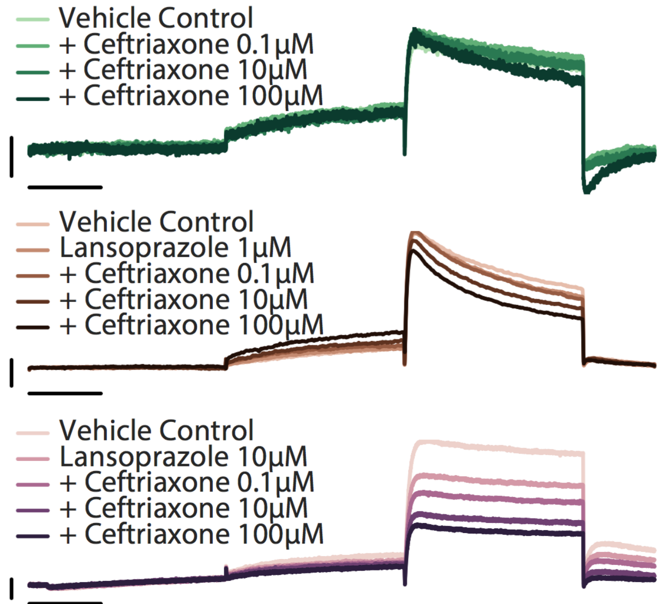


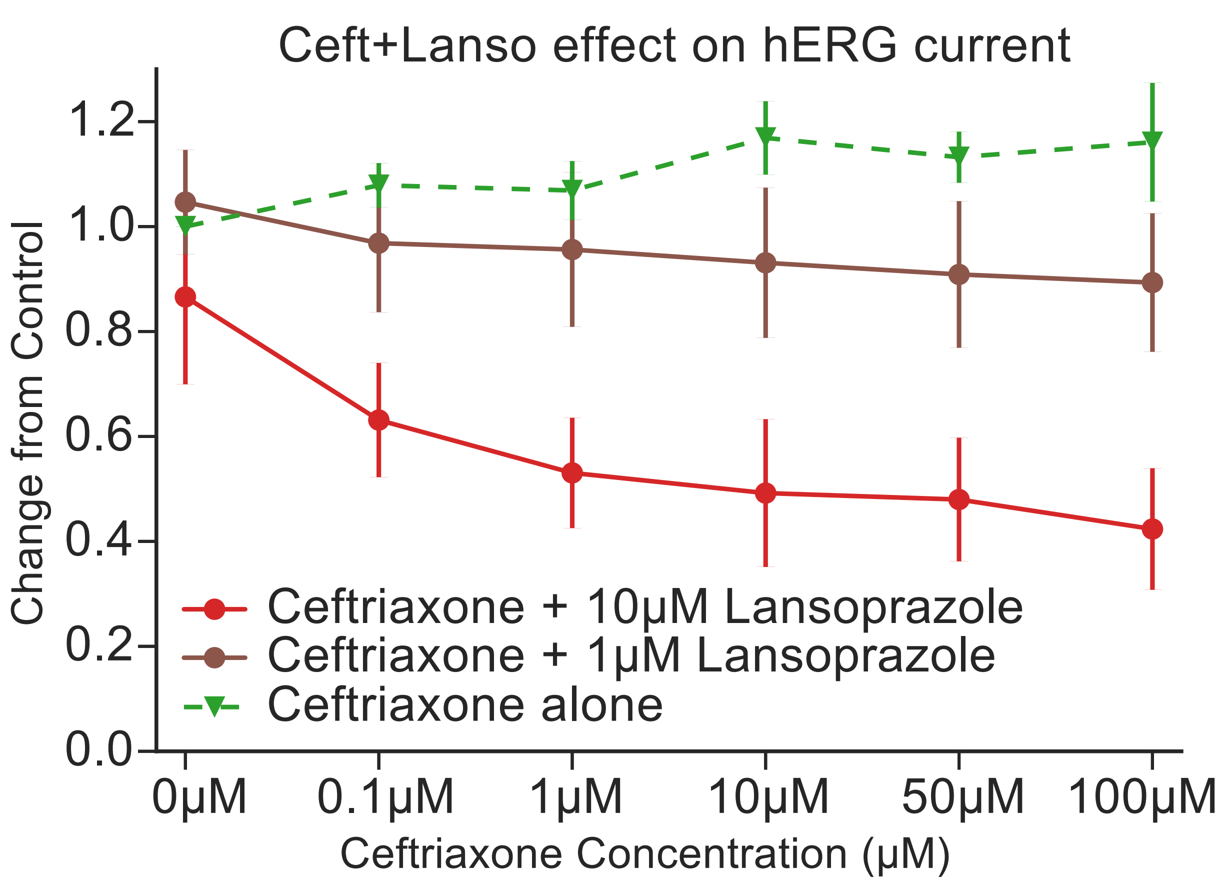
Cefuroxime+Lansoprazole



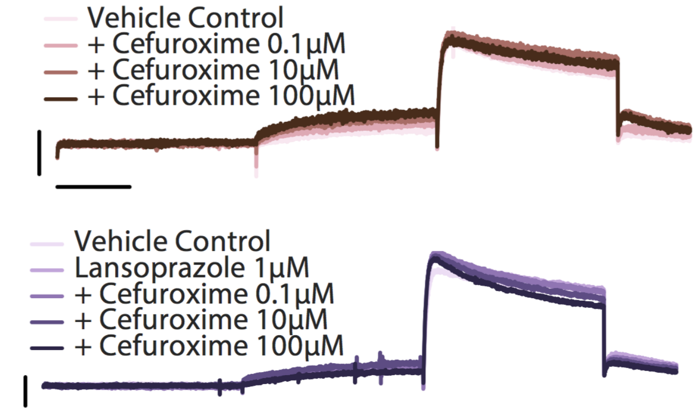
Lorberbaum, et al. JACC (In press)

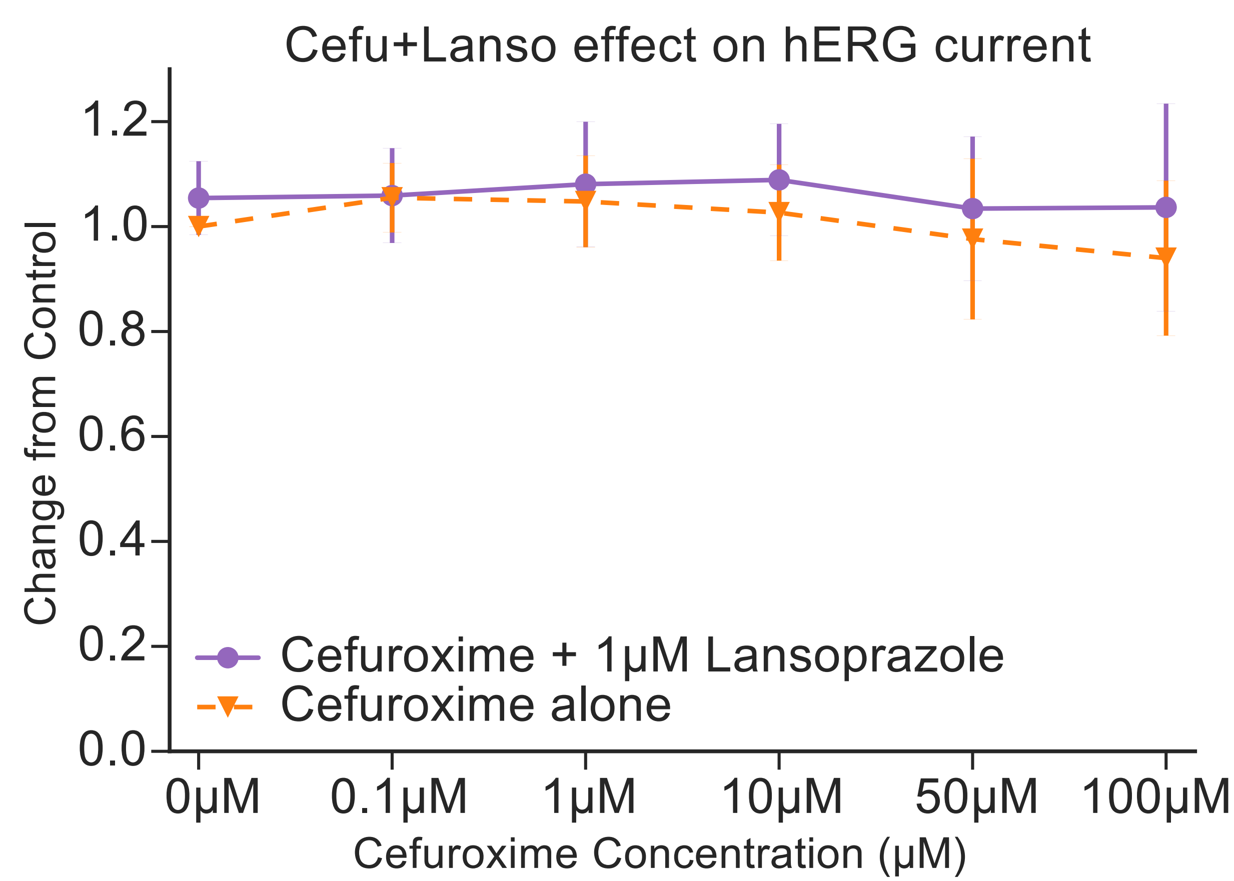
Ceftriaxone+Lansoprazole





Cefuroxime+Lansoprazole





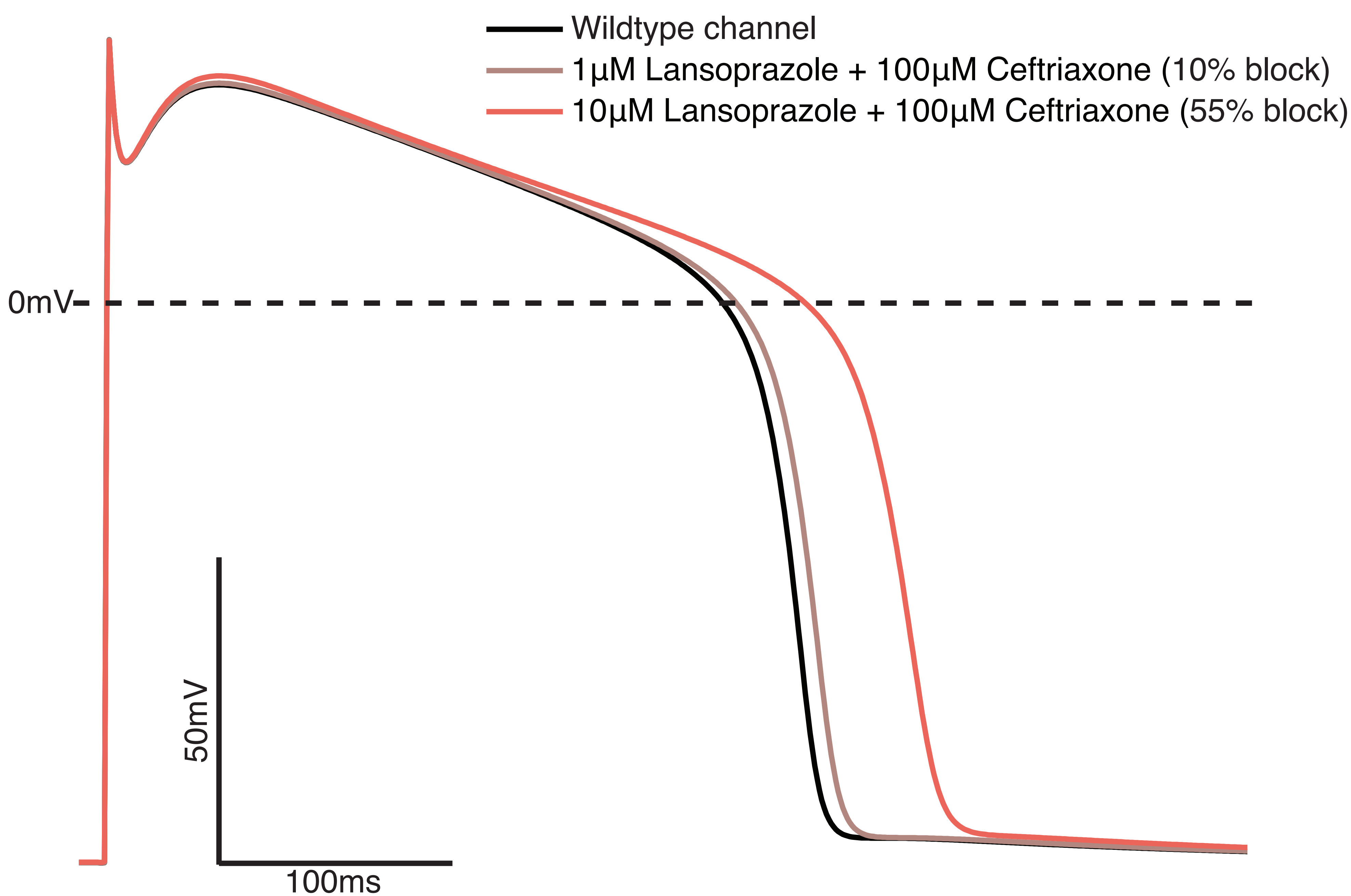
Lorberbaum, et al. JACC (In press)

Biophysical Journal Volume 87 September 2004 1507–1525

A Computational Model of the Human Left-Ventricular Epicardial Myocyte

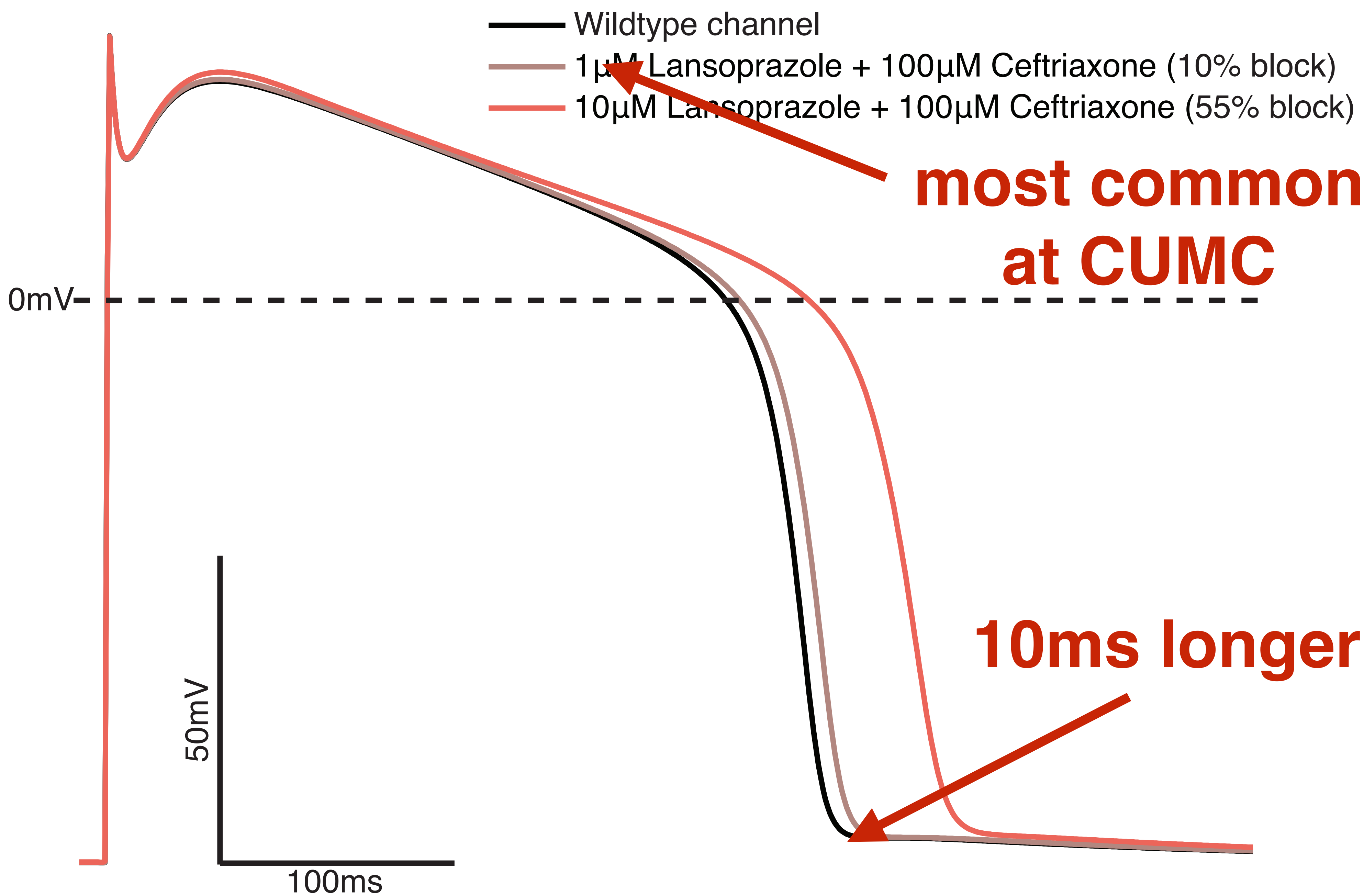
Vivek Iyer, Reza Mazhari, and Raimond L. Winslow
The Center for Cardiovascular Bioinformatics and Modeling and the Whitaker Biomedical Engineering Institute, The Johns Hopkins University School of Medicine and Whiting School of Engineering, Baltimore, Maryland

Computational model of human ventricular myocyte



Lorberbaum, et al. JACC (In press)

Computational model of human ventricular myocyte



Lorberbaum, et al. JACC (In press)

Data mining clinical information

- Drug-drug interactions can be discovered using observational data
  - **paroxetine/pravastatin**
  - **ceftriaxone/lansoprazole**
- <u>EHR data accurately predict prospective experiments</u>

Thank you

tatonettilab.org
nick.tatonetti@columbia.edu
@nicktatonetti

Current Lab Members

Rami Vanguri, PhD
Kayla Quinnies, PhD
Alexandra Jacunski
Tal Lorberbaum
Mary Boland
Joseph Romano
Yun Hao
Phyllis Thangaraj
Alexandre Yahi
Fernanda Polubriaginof, MD

Tal Lorberbaum
PhD Candidate in Cellular Physiology and Biophysics
Computational biology, systems pharmacology, protein structure modeling

Collaborators

David Goldstein, PhD
Krzysztof Kiryluk, MD, MS
David Vawdrey, PhD
Robert Kass, PhD
Kevin Sampson, PhD
Brent Stockwell, PhD
George Hripcsak, MD, MS
Ziad Ali, MD, DPhil
Ray Woosley, MD, PhD (Credible Meds)
Konrad Karczewski, PhD (Broad/MGH)
Joel Dudley, PhD (Mount Sinai)
Li Li, PhD (Mount Sinai)
Patrick Ryan, PhD (OHDSI)
Russ Altman (Stanford)
Issac Kohane (HMS)
Shawn Murphy (HMS)

Funding

NIGMS R01GM107145
Herbert Irving Fellowship
PhRMA Research Starter Grant
NCI P30CA013696
NIMH R03MH103957

Columbia University Medical Center
Discover. Educate. Care. Lead.