Future directions of OHDSI

George Hripcsak MD MS
Director, Columbia University OHDSI Coordinating Center
Professor and Chair of Biomedical Informatics
Columbia University Irving Medical Center
NewYork-Presbyterian Hospital

Patrick Ryan PhD
Johnson & Johnson
Columbia University Irving Medical Center

25Oct2022

Idea #1: HowOften.org

- Opportunity: Provide evidence to understand the absolute risk of adverse events
- Solution: Large-scale characterization of incidence of outcomes following drug exposure
  - Targets: New users of ingredient, for all ingredients
  - Outcomes: Event starts, for all adverse events
  - Time-at-risk: 30-day, on-treatment, intent-to-treat?
  - Results: Incidence proportion and rates, per database and prediction interval via meta-analysis
  - Dissemination: Interactive dashboard to allow user to search for drug and outcome
- Open questions:
  - Targets: Nested within indications?
  - Outcomes: 1st occurrence vs. all occurrence of outcomes? Phenotypes vs. codes?
  - Results: Stratify by age/sex/year?
  - Dissemination: How to show failures from objective database/cohort diagnostics?

25Oct2022

OHDSI's journey in incidence rates

thebmj
RESEARCH: SPECIAL PAPER
OPEN ACCESS
Check for updates
FAST TRACK

Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study

Xintong Li,[1] Anna Ostropolets,[2] Rupa Makadia,[3] Azza Shoaibi,[3] Gowtham Rao,[3] Anthony G Sena,[3,6] Eugenia Martinez-Hernandez,[4] Antonella Delmestri,[1] Katia Verhamme,[6,7] Peter R Rijnbeek,[6] Talita Duarte-Salles,[5] Marc A Suchard,[8,9] Patrick B Ryan,[2,3] George Hripcsak,[2]

frontiers | Frontiers in Pharmacology
ORIGINAL RESEARCH
published: 26 April 2022
doi: 10.3389/fphar.2022.814198
Check for updates

eClinicalMedicine
Part of THE LANCET Discovery Science
Articles

Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study

Erica A. Voss,[a,b,c,*] Azza Shoaibi,[a,c] Lana Yin Hui Lai,[a,d] Clair Blacketer,[a,b,c] Thamir Alshammari,[a,e] Rupa Makadia,[a,c] Kevin Haynes,[c] Anthony G. Sena,[a,b,c] Gowtham Rao,[a,c] Sebastiaan van Sandijk,[a,f] Clement Fraboulet,[g] Laurent Boyer,[g] Tanguy Le Carrour,[h] Scott Horban,[i] Daniel R. Morales,[j,k] Jordi Martínez Roldán,[l] Juan Manuel Ramírez-Anguita,[m,n] Miguel A. Mayer,[m,n] Marcel de Wilde,[a,b] Luis H. John,[a,b] Talita Duarte-Salles,[a,o] Elena Roel,[o] Andrea Pistillo,[o] Raivo Kolde,[p] Filip Maljković,[q] Spiros Denaxas,[r,s,t] Vaclav Papez,[r,s] Michael G. Kahn,[a,u] Karthik Natarajan,[a,v,w] Christian Reich,[a,x] Alex Secora,[x] Evan P. Minty,[a,y] Nigam H. Shah,[a,z] Jose D. Posada,[a,aa] Maria Teresa Garcia Morales,[ab] Diego Bosca,[ao] Honorio Cadenas Juanino,[ac] Antonio Diaz Holgado,[ac] Miguel Pedrera Jiménez,[ap] Pablo Serrano Balazote,[ap] Noelia García Barrio,[ap] Selçuk Şen,[ad] Ali Yağız Üresin,[ad] Baris Erdogan,[ae] Luc Belmans,[af] Geert Byttebier,[af] Manu L. N. G. Malbrain,[af,ag] Daniel J. Dedman,[ah] Zara Cuccu,[ah] Rohit Vashisht,[a,ai] Atul J. Butte,[a,ai,aj] Ayan Patel,[a,ai] Lisa Dahm,[a,aj] Cora Han,[a,aj] Fan Bu,[ak] Faaizah Arshad,[a,ak] Anna Ostropolets,[a,v] Fredrik Nyberg,[al] George Hripcsak,[a,v,w] Marc A. Suchard,[a,ak,am] Dani Prieto-Alhambra,[a,b,an] Peter R. Rijnbeek,[a,b] Martijn J. Schuemie,[a,c,ak] and Patrick B. Ryan[a,c,v]

Factors Influencing Background Incidence Rate Calculation: Systematic Empirical Evaluation Across an International Network of Observational Databases

*Anna Ostropolets[1†], Xintong Li[2†], Rupa Makadia[3], Gowtham Rao[3], Peter R. Rijnbeek[4], Talita Duarte-Salles[5], Anthony G. Sena[3,4], Azza Shaoibi[3], Marc A. Suchard[6,7], Patrick B. Ryan[1,3], Daniel Prieto-Alhambra[2] and George Hripcsak[1,8]**

[1]Columbia University Medical Center, New York, NY, United States, [2]Centre for Statistics in Medicine, NDORMS, University of Oxford, Oxford, United Kingdom, [3]Janssen Research and Development, Titusville, NJ, United States, [4]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands, [5]Fundacio Institut Universitari per a la Recerca a L'Atencio Primaria de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain, [6]Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States, [7]Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, United States, [8]New York-Presbyterian Hospital, New York, NY, United States

OHDSI's journey in incidence rates

ohdsi.github.io/CohortIncidence/

CohortIncidence 3.0.0 Reference Articles Changelog

CohortIncidence

Introduction

An R package and Java library for calculating incidence rates on the OMOP CDM.

Features

- Handles specifications of T-O-TAR-Subgroup pairs, and performs the calculation on the cross-product of the elements.
- Specify clean windows to account for immortal time after outcome.
- Allows multiple exposure and multiple outcomes per person accounting for time a
  paramaters.

Technology

CohortIncidence is an R package which wraps a Java library that implements most of t
package.

Links

Browse source code

Report a bug

License

Apache License 2.0

Citation

Citing CohortIncidence

Developers

Christopher Knoll
Maintainer

IncidencePrevalence: Estimate Incidence and Prevalence using the OMOP Common Data Model

Calculate incidence and prevalence using data mapped to the Observational Medical Outcomes Partnership (OMOP) common data model. Incidence and prevalence can be estimated for the total population in a database or for a stratification cohort.

Version:	0.4.0
Depends:	R (≥ 4.0)
Imports:	CDMConnector (≥ 1.0.0), checkmate (≥ 2.0.0), cli (≥ 3.0.0), DBI (≥ 1.0.0), dbplyr (≥ 2.0.0), dplyr (≥ 1.1.0), glue (≥ 1.5.0), ggplot2 (≥ 3.4.0), scales (≥ 1.1.0), lubridate (≥ 1.0.0), magrittr (≥ 2.0.0), purrr (≥ 0.3.5), rlang (≥ 1.0.0), stringr (≥ 1.5.0), tidyr (≥ 1.2.0), tidyselect (≥ 1.2.0), zip (≥ 2.2.0)
Suggests:	knitr, rmarkdown, RPostgres, tibble, duckdb, odbc, here, Hmisc, epitools, tictoc, testthat (≥ 0.3.1), spelling, PaRe
Published:	2023-06-18
Author:	Edward Burn [aut, cre], Berta Raventos [aut], Marti Catala [aut], Mike Du [ctb], Yuchen Guo [ctb], Adam Black [ctb], Ger Inberg [ctb], Kim Lopez [ctb]
Maintainer:	Edward Burn <edward.burn at ndorms.ox.ac.uk>
License:	Apache License (≥ 2)
URL:	https://darwin-eu.github.io/IncidencePrevalence/
NeedsCompilation:	no
Language:	en-US
Materials:	README
CRAN checks:	IncidencePrevalence results

HowOften:
Incidence of all effects in all drugs

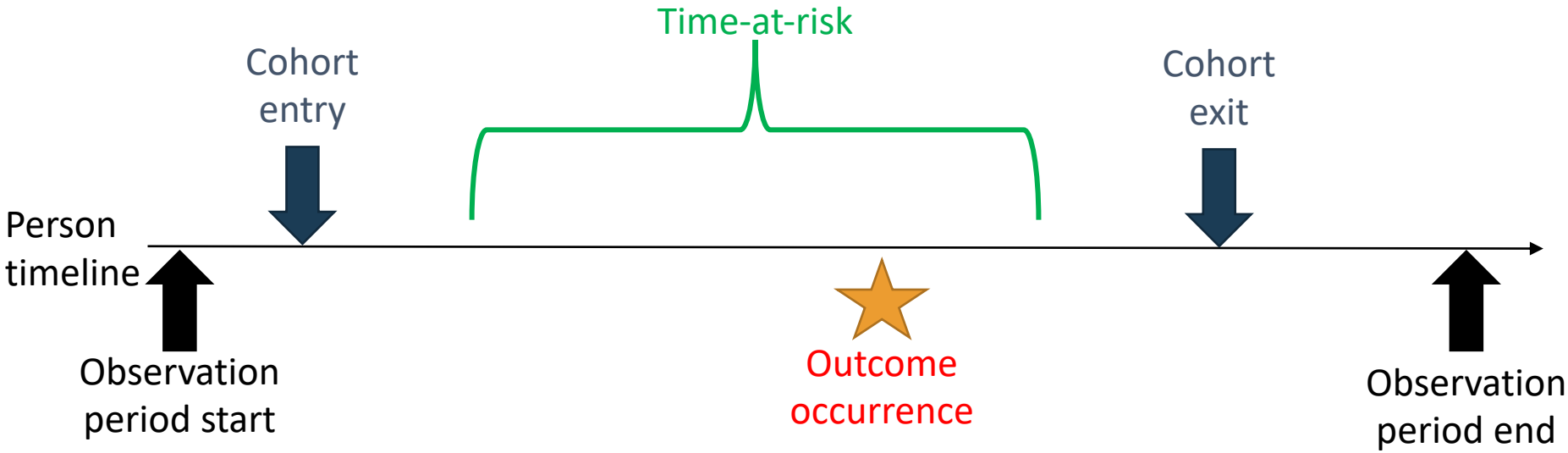
- Ask a doctor important side effects of a drug
- Then ask the incidence of that side effect
  - Many side effects are well known, but most clinicians have no idea of the incidence
  - The evidence is sparse
- Start simple
  - Characterization = non-causal rates
  - Tally how often conditions occur in drug therapy

Why start simple?

- If incidence is low, then I am set
- If incidence is high, then need to look out for it even if not caused by drug
- Feasible to execute all-by-all
- Fewer assumptions than causal
- More complicated than it looks, so need to get this one right first

"When I start this drug, what is the chance that I'll experience a condition in the next year?"

Dissecting the anatomy of incidence

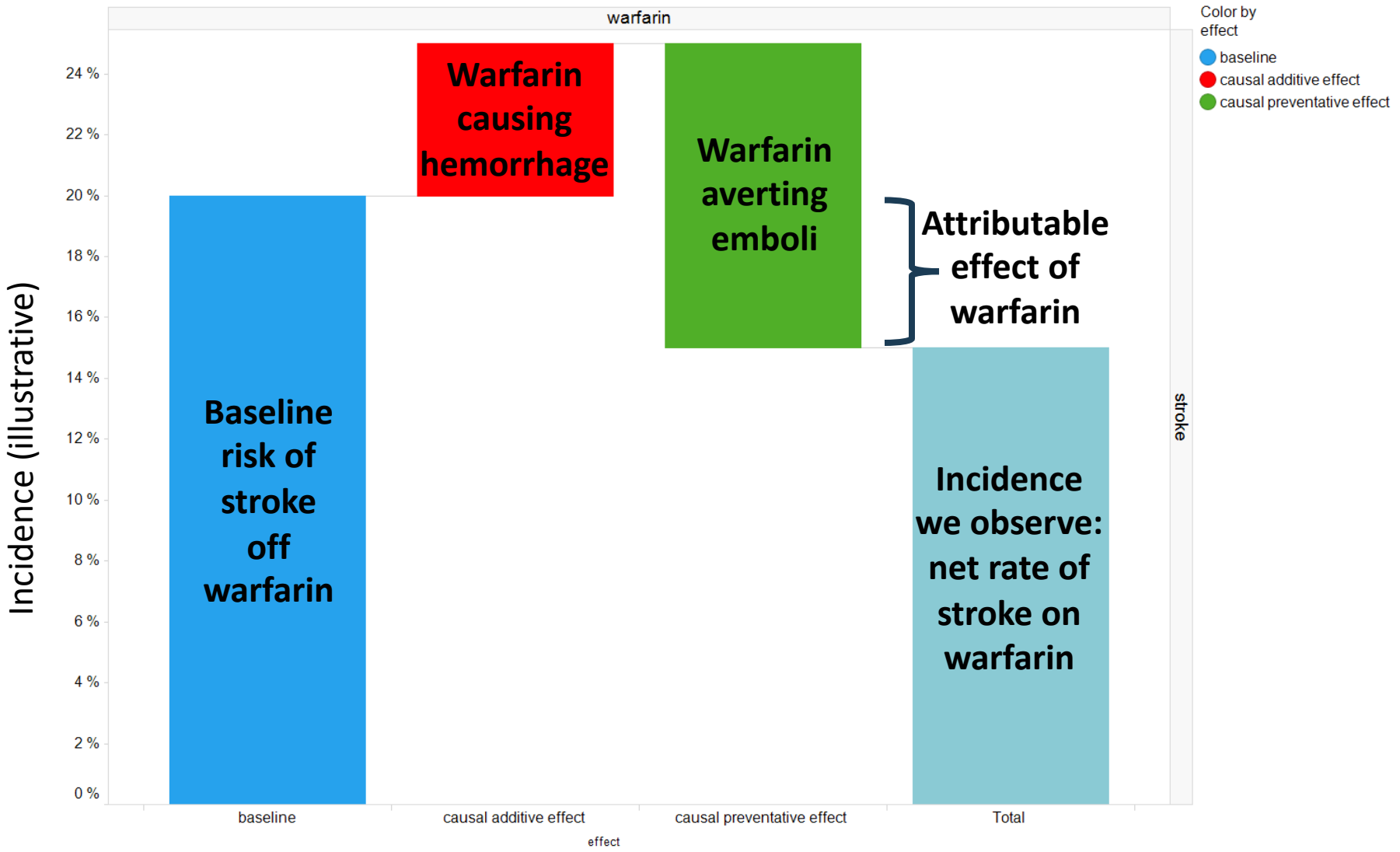


Incidence metrics:

$$\text{Incidence proportion} = \frac{\text{\# persons in the target cohort who have new outcome occurrence during the time-at-risk}}{\text{\# persons in the target cohort with time-at-risk*}}$$

$$\text{Incidence rate} = \frac{\text{\# persons in the target cohort who have new outcome occurrence during the time-at-risk}}{\text{person-time at-risk for persons in the target cohort with time-at-risk*}}$$

Incidence rates do not tell causal effect



Myriad difficult choices that researchers have to make to produce a ‘simple answer’

- How should the target cohort be defined?
- How should the outcome be defined?
- How should the time-at-risk be defined?
- How to account for patients with incomplete time-at-risk?
- Which statistical metrics should be reported?
- Which data should be used?

Myriad difficult choices that researchers have to make to produce a 'simple answer'

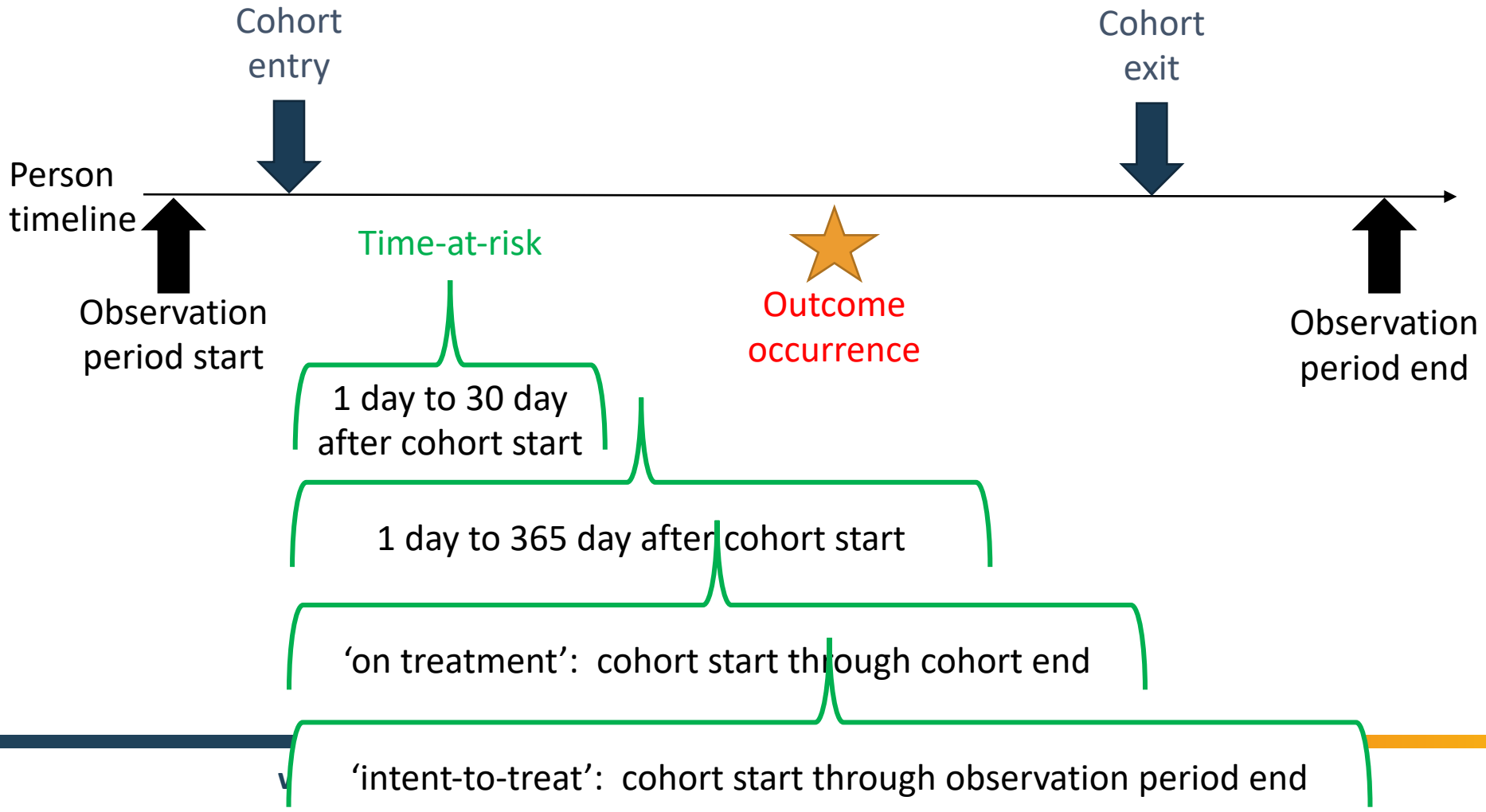
- **How should the target cohort be defined?**
  - For a cohort of 'new users of a drug', cohort entry can be defined as the date of first exposure
    - Should other inclusion criteria be imposed, such as requiring prior diagnosis of labeled indication? How do these criteria impact the generalizability of this estimate to the target population?
  - What minimum lookback period is required to ensure 'new user'?
    - Shorter period provides larger (and more generalizable) sample to yield more precise estimate
    - Longer period provides greater confidence that patient is truly 'newly exposed' and provides longer prior history to ensure outcome is incident occurrence

Myriad difficult choices that researchers have to make to produce a 'simple answer'

- **How should the outcome be defined?**
  - Alternative phenotype definitions often represent different sensitivity/specificity tradeoffs, though those operating characteristics are commonly unknown at the time of choosing the definition
  - 'First diagnosis' may be more sensitive but less specific than 'first diagnosis with hospitalization'
  - Outcome cohort can include 'first ever occurrence' vs. 'first occurrence post-exposure' vs. 'all occurrences'
  - Phenotype evaluation diagnostics required to quantify potential measurement error and calibrate incidence estimates

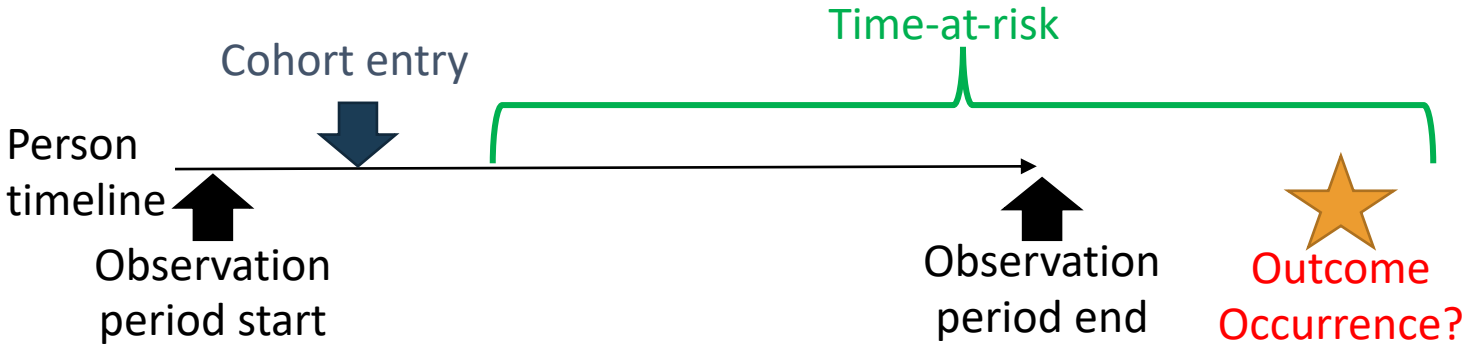
Myriad difficult choices that researchers have to make to produce a 'simple answer'

- **How should the time-at-risk be defined?**



Myriad difficult choices that researchers have to make to produce a 'simple answer'

- **How to account for patients with incomplete time-at-risk?**



  - Include persons with incomplete follow-up time
    - Assumes unobserved time did not have events
    - Lower bound of true incidence estimate
      = #observed_events / (#observed_events + #missed_events)
    - Worsens with increased censoring or more events in censored pts
  - Include only persons with full time-at-risk
    - Usually higher than true incidence estimate (if rate is uniform)
      ≈ #observed_events / (#observed_events – #missed_events)
    - Worsens with increased censoring (also smaller sample size)
    - Can flip if high rate of events in censored period

Myriad difficult choices that researchers have to make to produce a 'simple answer'

- **Which statistical metrics should be reported?**
  - Incidence proportion requires a defined time-at-risk
  - Incidence rate allows variable-length time-at-risk, but assumes constant hazard over time-at-risk
  - 95% confidence intervals commonly reported, but only represent sampling variability.
  - Within-source systematic error and between-source heterogeneity represent larger sources of uncertainty that are not adequately quantified in current practice
  - Characterizing the range of estimates across network analysis (e.g. minimum → maximum) may be more reflective of uncertainty than sampling statistics from any given data source

Myriad difficult choices that researchers have to make to produce a ‘simple answer’

- **Which data should be used?**
  - Incidence estimation requires a minimum longitudinal follow-up for the desired time-at-risk
  - Data should be represent patients that are contained within the target population of interest (but not necessarily be a random sample or fully representative of the target population)
  - A network analysis may provide heterogeneity across patients, health systems, geographies and represent different perspectives and health care process biases

Hierarchy of uncertainty

- Biology (genetics)
  - This is signal that you want to measure, not error
- Environment (i.e., its effect on biology)
  - Also signal that you want to measure
- Health care process bias
  - Measurement error
- Extract-transfer-load
  - ETL errors, and ETL interpretations
- Sampling error
  - Sampling error goes to zero with sample size
- Confounding
  - Different confounders in different populations

Problems with current practice

- For a majority of incidence questions of potential interest, there is no readily accessible evidence available
- When evidence is identified in the literature, it can be difficult to interpret:
  - Incidence metric – ambiguity in what's reported
  - Unspecified time-at-risk
  - Generalizability of target population
  - Diversity of phenotype definitions
  - Different evidence sources (RCT, systematic reviews, observational studies)
    - Systematic reviews synthesize results from different metrics/time-at-risk/phenotypes
  - Observational data have different sources of systematic error that are rarely quantified or corrected for

How could OHDSI help?

- Develop a standardized framework for incidence evidence generation and dissemination
- Fill the gaps where there is currently no available evidence
- Augment existing knowledge with new evidence systematically generated across the world's largest observational data network
  - Demonstrate reliability of current knowledge through replication
  - Reconcile discordant evidence observed in the literature through quantification of uncertainty
  - Apply causal effect estimates to overall incidence to assess attributable risk

“Things we know that we know”

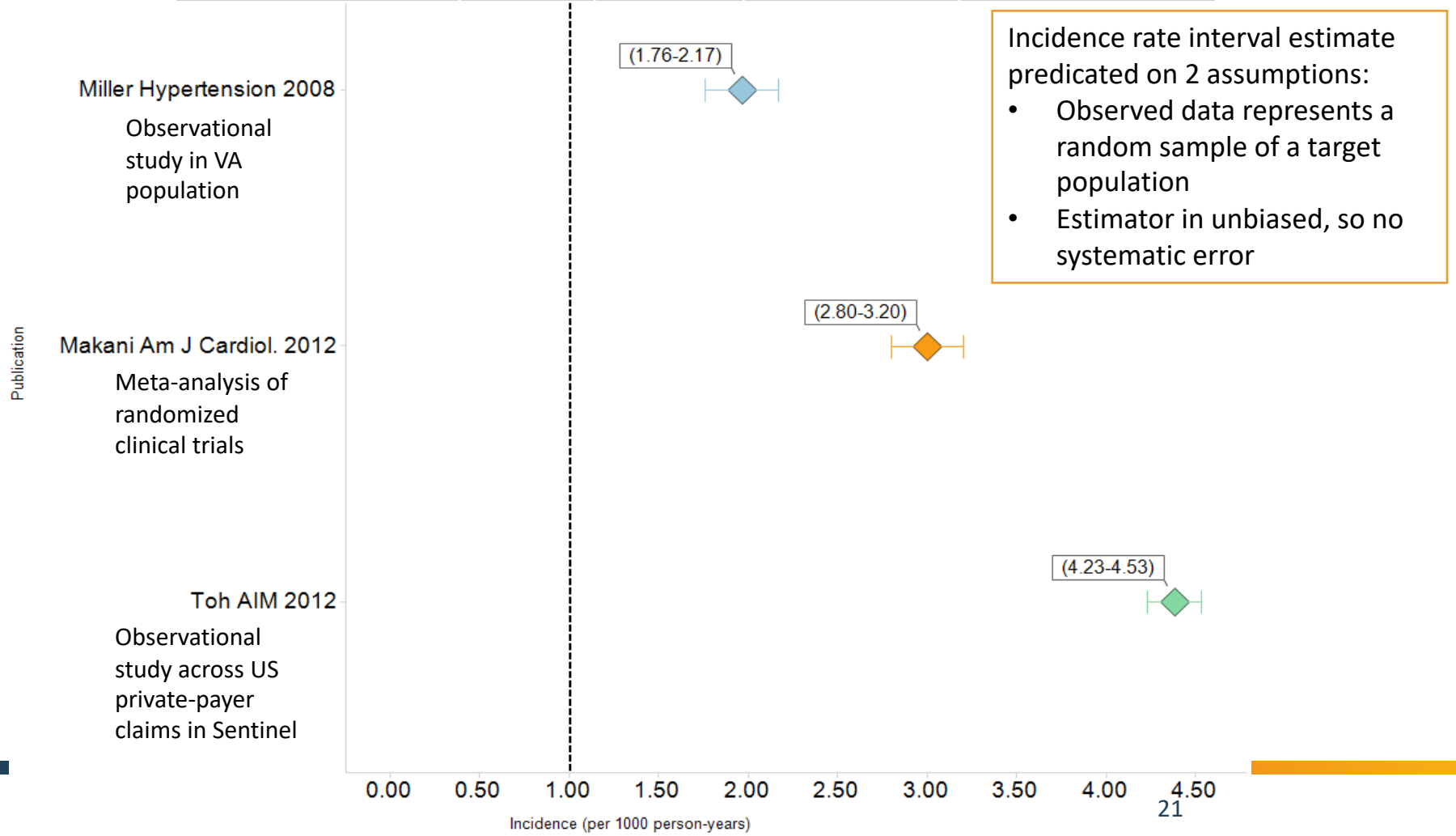
- What we think we know:
  - ACE inhibitors cause angioedema

- What we want to know:
  - Clinical characterization: Incidence of angioedema in patients exposed to ACE inhibitors
  - Population-level effect estimation:
    - Safety surveillance: Strength of association with ACE inhibitor vs. counterfactual
    - Comparative effectiveness: Strength of association with ACE inhibitor, relative to alternative treatments
    - Attributable risk
  - Patient-level prediction: Probability that a patient will experience event, given baseline characteristics

What's on the product label?

What's the published evidence?

Publication	Person-years	Events	Incidence (per 1000 person-years)	95% CI (Incidence rate per 1000 person-years)
Miller Hypertension 2008	179,088	352	1.97	(1.76-2.17)
Makani Am J Cardiol. 2012	185,067	394	3.00	(2.80-3.20)
Toh AIM 2012	753,105	3,301	4.38	(4.23-4.53)



Incidence rate interval estimate predicated on 2 assumptions:

- Observed data represents a random sample of a target population
- Estimator in unbiased, so no systematic error

How does it get distilled to clinicians?

UpToDate®

WHY UPTODATE? | PRODUCT | EDITORIAL | S

Topic Outline

INTRODUCTION
EPIDEMIOLOGY
CLINICAL FEATURES
- Affected areas
  - Face, mouth, and upper airway
  - Intestine
- Time course
- Severity
- Recurrence after stopping ACE inhibitor therapy

PATHOPHYSIOLOGY
- ACE inhibition
- Role of bradykinin in angioedema

RISK FACTORS
- Possible risk factors
- Predisposing genetic factors

DIAGNOSIS
- Evaluation of abdominal pain

DIFFERENTIAL DIAGNOSIS
TREATMENT
- Airway management
- Discontinue ACE inhibitor
- Other interventions
- Additional therapies for severe or persistent symptoms
  - Icatibant
  - Ecallantide
  - Fresh frozen plasma
  - Purified C1 inhibitor concentrate

ACE inhibit

Authors
Autumn Chandle
Aleena Banerji, M

INTRODUCTION

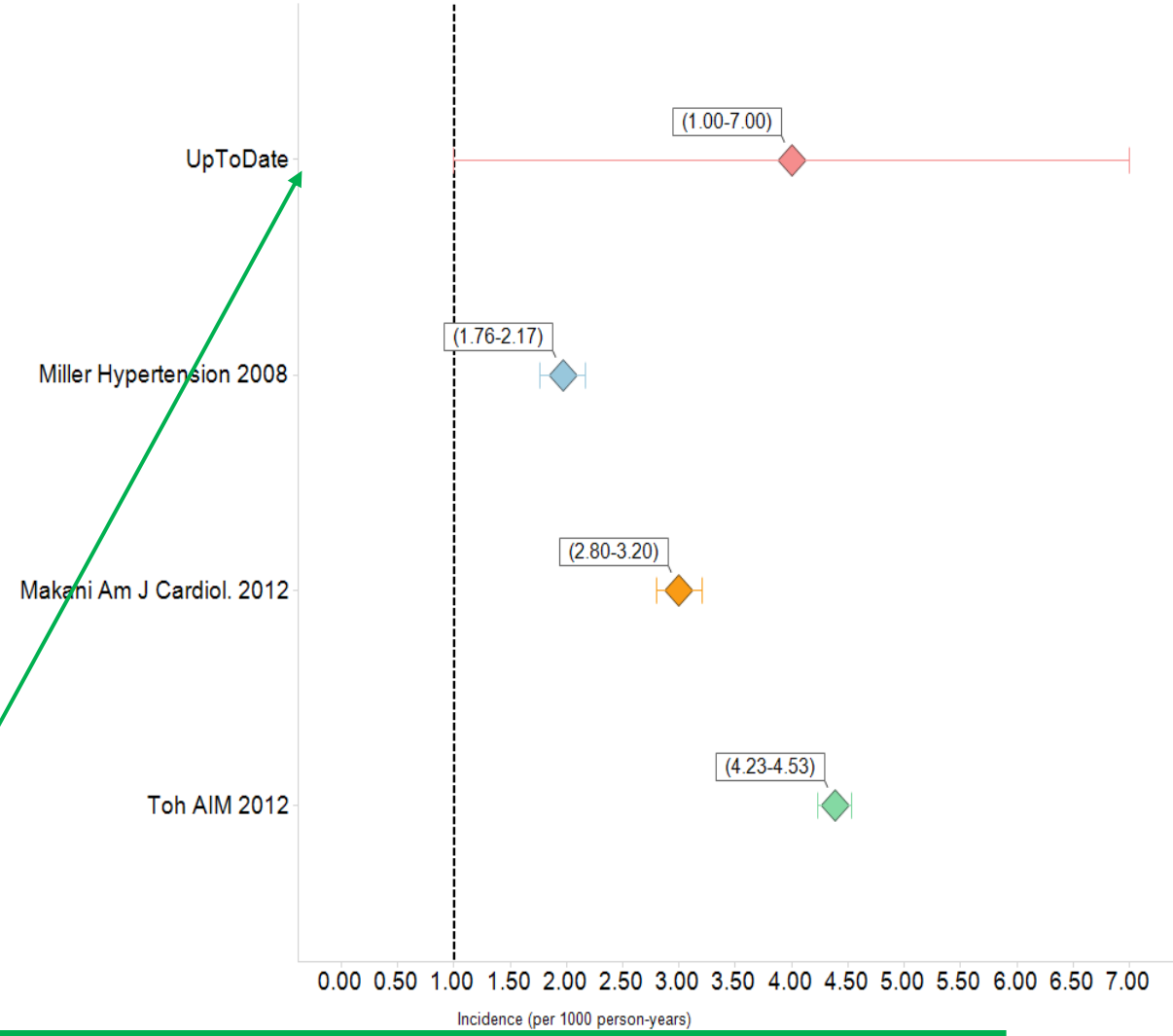
Angiotensin-conver
widely prescribed. F
pain due to intestina

This topic reviews t
induced angioedem
causes is found els
Pathogenesis and c

EPIDEMIOLOGY

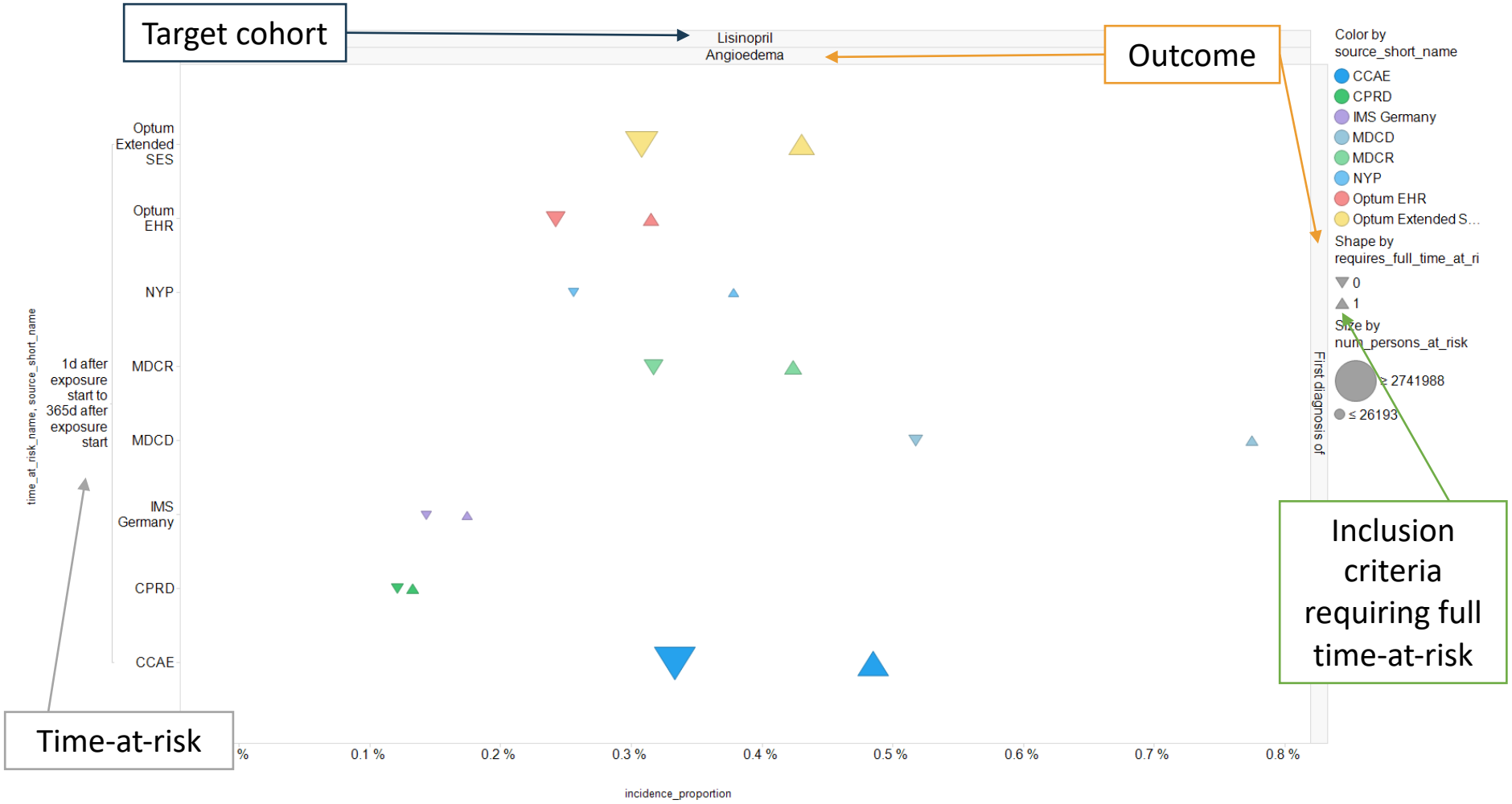
Angiotensin-conver
angioedema is up t

Although the risk to
cause of drug-induc
year [10-12]. Appro
and more than 40 m
myocardial infarctio



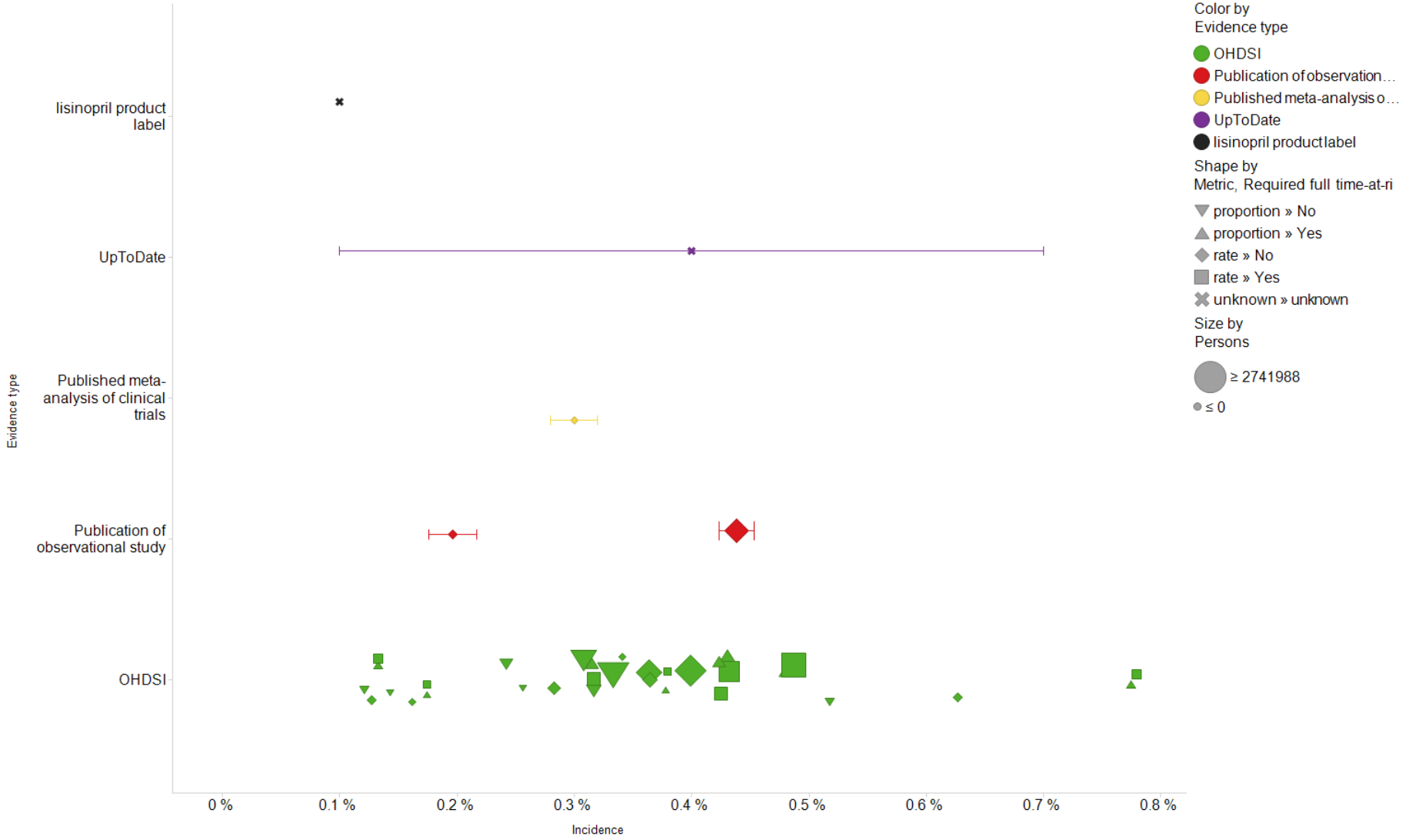
The overall incidence of angioedema related to ACE inhibitors has been estimated between 0.1 percent and 0.7 percent [1-5,14-16]. However, the lower end of this range may overlap with the background rate of angioedema in the general population. In the TRANSCEND trial of ACE inhibitor-intolerant individuals given an angiotensin II receptor blocker (ARB) or placebo, rates of angioedema were 0.07 and 0.1 percent in the ARB and placebo groups, respectively [17].

What if a standardized incidence estimation was consistently applied across the OHDSI network?



Range of incidence proportions from across 8 sources in the OHDSI data network: 0.1% - 0.8%

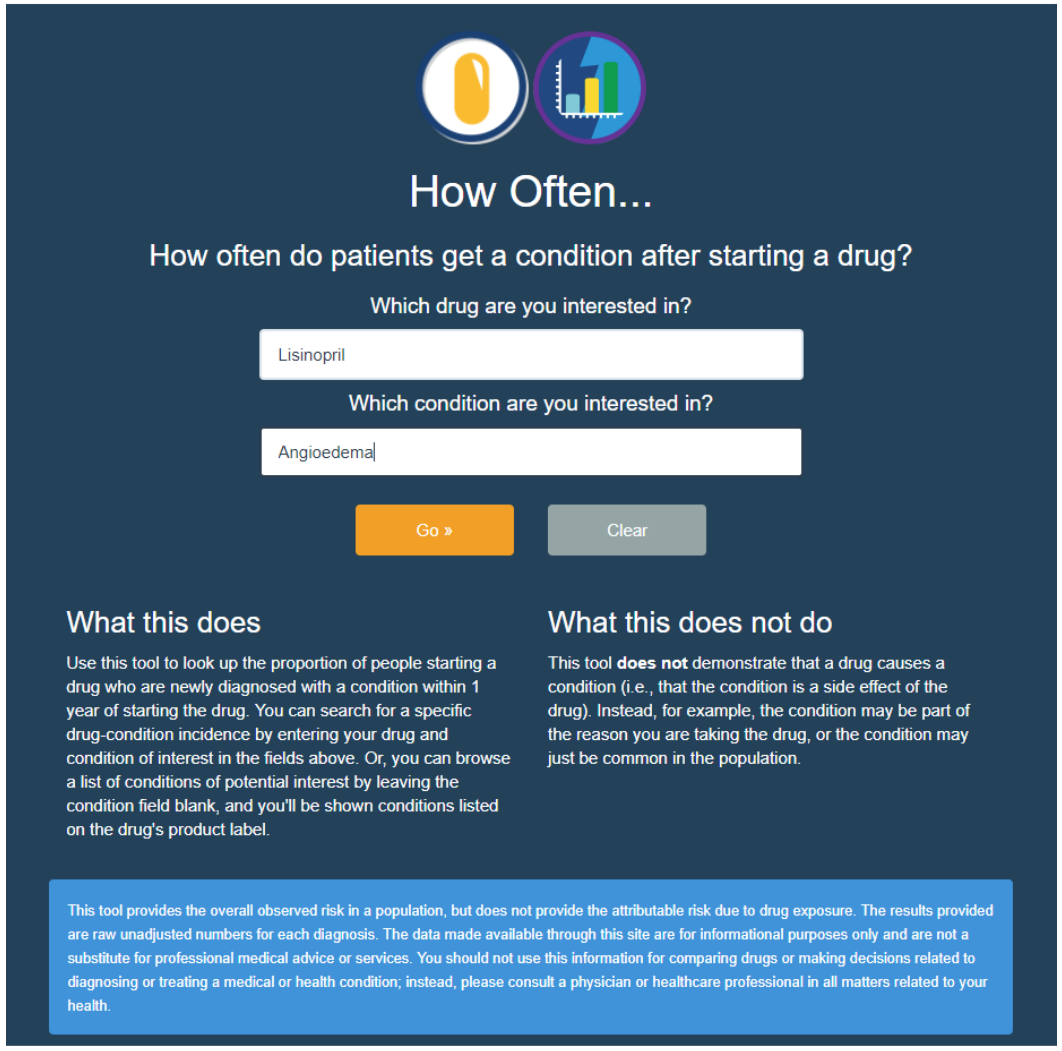
How does OHDSI evidence compare with prior evidence?



Caveats to All-by-All Incidence

- Why might rate be high
  - (Recall that indications reduced b/c first occurrence is after exposure)
  - High in the underlying population
  - Indication is a risk
  - Things associated with indication
  - Reversed timing (Drug -> Indication)
  - Or could be causal (attributable risk)
- But if rate is low and side effect is not serious, then side effect may not be important

Went live 2017



Experience

- Columbia-NYP emergency department
  - Used by some staff on patients with unexplained symptoms
  - No formal evaluation
- Was the trigger for Anna Ostropolets's Data Consult Service
  - Angioedema with penicillin
- Reliable for several examples but some exceptions (sexual disfunction)
- Did not maintain the proof of concept

We need you!

- We have shown proof-of-concept
- But this will only work if everyone contributes
- How can you help?

Original team

OHDSI collaborators

- Marc Suchard
- Martijn Schuemie
- David Madigan
- Jon Duke
- Patrick Ryan
- George Hripcsak

Columbia team

- Ray Chen
- Mark Velez
- Karthik Natarajan
- Jungmi Han
- Peng Jin

OHDSI Infrastructure

- Lee Evans

Global Symposium

Global Symposium

Oct. 20-22 • East Brunswick, NJ, USA

ohdsi.org/OHDSI2023

OHDSI 2023 Global Symposium

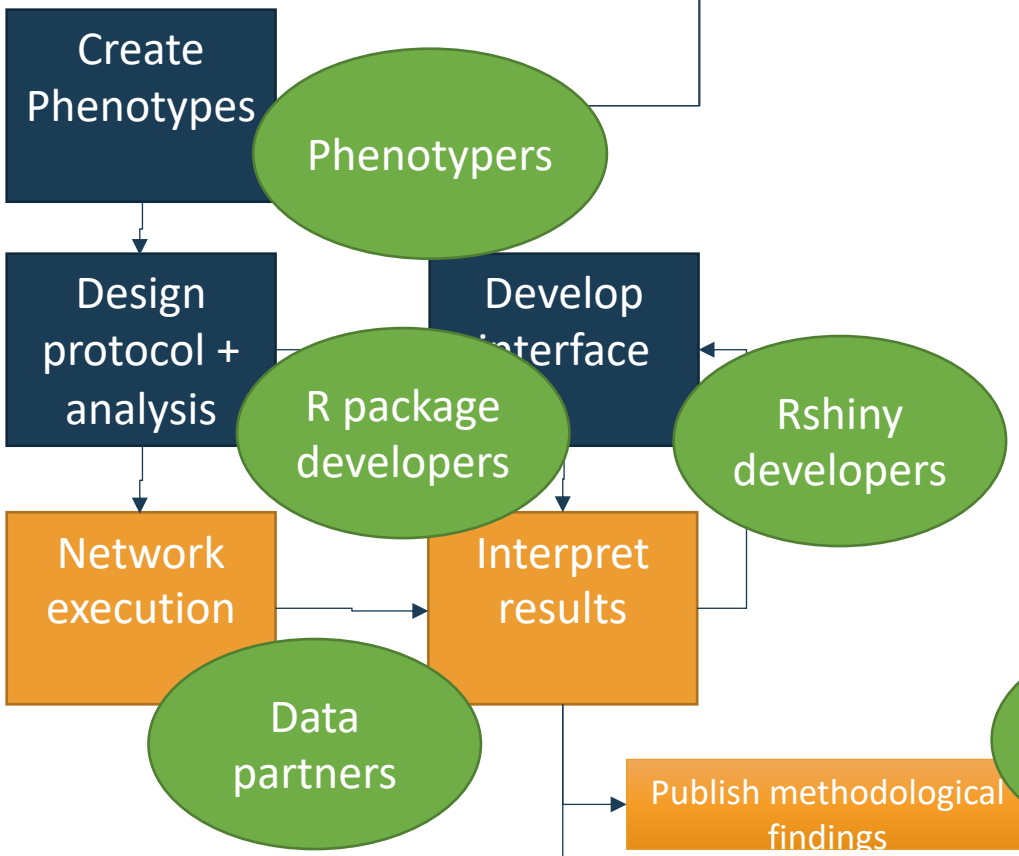
October 20-22 • East Brunswick, NJ, USA

** This agenda is tentative and subject to change*

	Friday, Oct 20	Saturday, Oct 21	Sunday, Oct 22
8:00am	Welcome to OHDSI2023!	Intro to OHDSI Tutorial & OHDSI workgroup activities	OHDSI collaborative workshop: HowOften
9:00am	State of the Community		
10:00am	Community networking		
11:00am	Plenary session		
12:00pm	Lunch	**Collaborator Showcase: posters & demos**	**Collaborator Showcase: posters & demos**
1:00pm	Panel: Network studies	OHDSI collaborative workshop: HowOften	HDSI workgroup activities
2:00pm	**Collaborator Showcase: posters & demos**		
3:00pm	**Collaborator Showcase: Lightning talks**		
4:00pm	**Collaborator Showcase: posters & demos**		
5:00pm	Closing talk	ree time ☺	Time to go home ☹
6:00pm	OHDSI Got Talent!		

HowOften community effort

We need you!



Target cohorts:

1. 'general population'
   1Jan2018, 1Jan2022, 1st visit 2018, 1st visit 2022….
2. 'important indications'
   Diabetes, Hypertension, Renal impairment, hepatic impairment, cardiac impairment, cancers (breast, lung, prostate), mental health (depression, bipolar, …hizophrenia), infectious …e (Covid, HIV), eye …blinding diseases), …men's health (pregnancy, endometriosis), ….
3. …drug classes'
   …ACE inhibitors, SSRI, …GF, ….

Outcome cohorts:

1. Adverse events of special interest (AESI)
   Guillain-Barre Syndrome, Thrombocytopenia, Ischemic stroke, Transverse myelitis, …
2. Designated medical events (DME)
   Stevens-Johnson Syndrome, pancreatitis, rhabdomyolysis acute kidney injury, …
3. Indication outcomes
   End-stage renal disease, acute myocardial infarction, hepatic failure, …
4. Side effects of drugs
   Headache, diarrhea, anaphylaxis, …

Stratification factors:
1. Database, 2. Age, 3. Sex, 4. Calendar year

HowOften next steps

- Pre-Symposium:
  - Draft protocol to allow data partners to get approval to participate
  - Develop and evaluate all phenotypes for targets and outcomes
    - All outcomes to be used in HowOften must be included in OHDSI Phenotype Library
  - Release analysis package that includes all phenotypes and analysis to instantiate cohorts and characterize incidence of all target-outcome pairs
- During Symposium:
  - Execute HowOften analysis package across OHDSI network
  - Deploy viewer to allow exploration of all results
  - Collaborate on appropriate use of evidence
    - Methodological questions: how to ensure results are reliable?
    - Development questions: how to improve user interface to disseminate results?
    - Clinical questions: what have we learned that can fill evidence gaps and improve decision-making?

When poll is active, respond at **PollEv.com/patrickryan800**

What phenotypes do you think should be included as target or outcome cohorts in HowOften?

Top

No responses received yet. They will appear here...

When poll is active, respond at **PollEv.com/patrickryan800**

What are novel ways that you think HowOften might be used by stakeholders?

Top

No responses received yet. They will appear here...