



# Closing: Collaborating on Evidence at Scale

Patrick Ryan, PhD  
Johnson & Johnson  
Columbia University



We thank the FDA for their generous  
support of the 2024 OHDSI symposium  
through the FDA SCIENTIFIC CONFERENCE  
GRANT PROGRAM (R13FD006972)



# Thank you OHDSI 2024 Symposium Sponsors!

**Johnson & Johnson**



**<epam>** | ODYSSEUS  
DATA SERVICES





# Thank you for OHDSI 2024 Scientific Review Committee

- Atif Adam
- Nsikak Akpakpan
- Thamir Alshammari
- Donna Armaignac
- Juan Banda
- Bill Baumgartner
- Clair Blacketer
- Sandeep Chandra Bollepalli
- Fan Bu
- Ed Burn
- Cindy Cai
- Marti Catala Sabate
- Adrien Coulet
- Deborah Duran
- Yudha Saputra
- Mengling Feng
- Davera Gabriel
- Narendra Garbyal
- Asieh Golozar
- Ben Hamlin
- Jill Hardin
- Kevin Haynes
- Jason Hsu
- Elisse Katzman
- Chungsoo Kim
- Kristin Kostka
- Christophe Lambert
- Christine Lu
- Rupa Makadia
- Benjamin Martin
- Evan Minty
- Paul Nagy
- Daniel Nugroho
- Edwin Ocasio
- Anna Ostropolets
- Melanie Philofsky
- Nicole Pratt
- Daniel Prieto Alhambra
- Hanieh Razzaghi
- Patrick Ryan
- Amir Sarayani
- Sarah Seager
- Azza Shoaibi
- Cynthia Sung
- Joel Swerdel
- Kim Van Naarden Braun
- Mui Van Zandt
- Rohit Vashisht
- Kiran Vupparaboina
- Andrew Williams
- Anna Yang
- Seng Chan You
- Linying Zhang
- Summera Zhou



# Congratulations, 2024 Titan Award nominees!

Alexander Davydov • **Andrew Kanter** • Anna Ostropolets • **Anthony Sena** • April Olympians Team • **Asieh Golozar** • Ben Martin • **Benjamin Viernes** • Christopher Mecoli • **Cindy Cai** • Clair Blacketer • **Cynthia Sung** • Daniel Morales • **Danielle Boyce** • DARWIN EU Development Team • **Elisse Katzman** • Evanette Burrows • **Eye Care and Vision Research Workgroup** • Frank DeFalco • **George Hripcsak** • Greg Klebanov • **Henrik John** • Hsin Yi Chen • **J Swetha Kiranmayi** • Jack Janetzki • **James Weaver** • Jared Houghtaling • **Jen Park** • Joel Swerdel • **John Gresh** • Jung Ho Kim • **Justin Manjourides** • Kyle Zollo-Venecek • **Liesbet Peeters** • Linying Zhang • **Louis Hendricks** • Maarten van Kessel • **Manlik Kwong** • Marc Suchard • **Marta Pineda-Moncusi** • Marti Catala Sabate • **Martijn Schuemie** • Martin Lavalley • **Maxim Moinat** • Michael Gurley • **Michael Matheny** • Michel Walravens • **Michelle Hribar** • Minnesota EHR Consortium Health Trends Across Communities Project Team • **Montse Camprubi** • Mengling 'Mornin' Feng • **Natthawut 'Max' Adulyanukosol** • OHDSI APAC ETL Team • **OHDSI Standardized Vocabularies Team** • Oleg Zhuk • **Parthiban Sulur** • Polina Talapova • **Qi Yang** • Renske Los • **Rich Boyce** • Robert Koski • **Robert Miller** • Roger Carlson • **Scott DuVall** • Thamir Alshammary • **Theresa Burkard** • Thomas Falconer • **Tom Seinen** • Vishnu Chandrabalan • **Vlad Korsik** • Will Kelly • **Zhen Lin**





## 2024 Titan Awards



Alexander Davydov



Vlad Korsik



Anna Ostroplets



Oleg Zhuk

[www.ohdsi.org](http://www.ohdsi.org)

Data Standards

#JoinTheJourney

## 2024 Titan Awards



Cindy Cai



Jung Ho Kim



Jack Janetzki

[www.ohdsi.org](http://www.ohdsi.org)

Clinical Applications

#JoinTheJourney

## 2024 Titan Awards



Linying Zhang



[www.ohdsi.org](http://www.ohdsi.org)

Methodological Research

#JoinTheJourney

## 2024 Titan Awards



Natthawut 'Max'  
Adulyanukosol



[www.ohdsi.org](http://www.ohdsi.org)

Community Collaboration

#JoinTheJourney

## 2024 Titan Awards



Clair Blacketer



[www.ohdsi.org](http://www.ohdsi.org)

Community Leadership

#JoinTheJourney

## 2024 Titan Awards



Martin Lavallee



[www.ohdsi.org](http://www.ohdsi.org)

Open-Source Development

#JoinTheJourney

## 2024 Titan Awards



Montse Camprubi



Elisse Katzman

[www.ohdsi.org](http://www.ohdsi.org)

Community Support

#JoinTheJourney



# Thank you to those who made today happen

- Elisse Katzman
  - Craig Sachson
  - Ann Marshak
  - Anita Barrett
  - Sofia Ellis-Chin
  - Charika Johnson
  - Patrick Ryan
  - OHDSI Steering Workgroup
  - Adit Anand
  - Tara Anand
  - James Baker
  - Cindy Chen
  - Fangyi Chen
  - Young Choi
  - Pooja Desai
  - Yilu Fang
  - Florent Pollett
  - Harry Reyes
  - Robert Barrett
  - Brooke Lawler
  - Hannah Lee
  - Star Liu
  - Ben Martin
  - Jen Park
-



Thank you to the anonymous contributor





# Collaborating at scale

2022:



2023:





Let's set a world record! (unofficially)





# The scale of LEGEND-T2DM

Open access Protocol

**BMJ Open** Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies

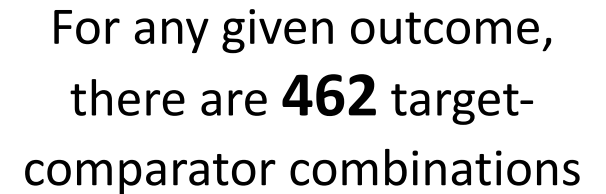
- 4 drug classes
- 5 DPP4i + 6 GLP1RA + 4 SGLT2 + 7 SU = 22 drug ingredients
- 32 outcomes

Rohan Khera  
Anna Ostrop  
Harlan M Krui

**Table 1** T2DM drug classes and individual agents within each class

| DPP4 inhibitors | GLP1 receptor antagonists | SGLT2 inhibitors | Sulfonylureas  |
|-----------------|---------------------------|------------------|----------------|
| Alogliptin      | Albiglutide               | Canagliflozin    | Chlorpropamide |
| Linagliptin     | Dulaglutide               | Dapagliflozin    | Glimepiride    |
| Saxagliptin     | Exenatide                 | Empagliflozin    | Glipizide      |
| Sitagliptin     | Liraglutide               | Ertugliflozin    | Gliquidone     |
| Vildagliptin    | Lixisenatide              |                  | Glyburide      |
|                 | Semaglutide               |                  | Tolazamide     |
|                 |                           |                  | Tolbutamide    |

DPP4, dipeptidyl peptidase-4; GLP1, glucagon-like peptide-1; SGLT2, sodium-co-transporter-2; T2DM, type 2 diabetes mellitus.

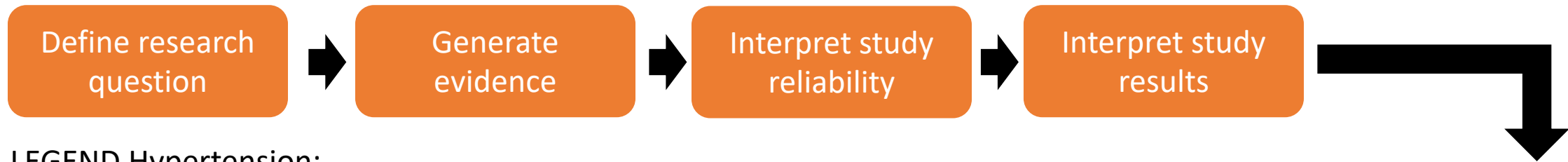


Across 32 outcomes,  
there are **14,784**  
target-comparator-  
outcome combinations

Every target-comparator-outcome is a valid clinical question that patients deserve reliable evidence for

# Pre-specification of a systematic approach

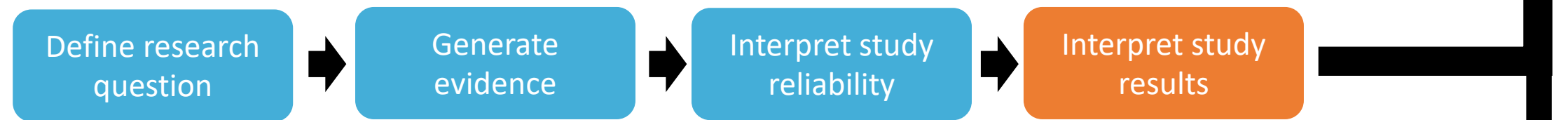
Traditional observational study:



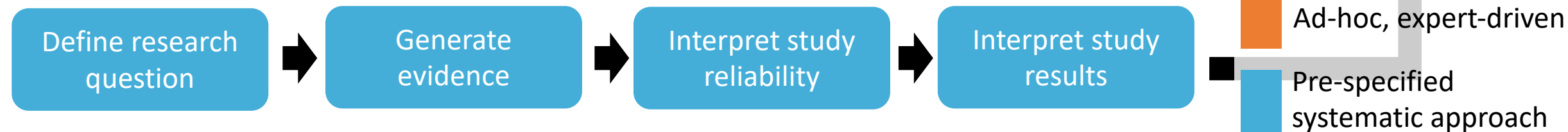
LEGEND Hypertension:



New recommendation:



Future:







# A rubric for interpreting studies

- Quality of study
  - Study design
    - Multiple designs to evaluate robustness (ex: comparative cohort and SCCS)
    - Multiple analyses within design (ex: PS matching vs stratification, on-treatment vs. ITT )
    - Objective diagnostics to test statistical assumptions, quantify residual error, and establish unblinding rules
  - Transparency
    - Pre-specified protocol
    - Publicly accessible analytic source code
    - Provenance of full resultset
  - Diversity of databases
    - Populations
    - Geographies
    - Data capture processes





# A rubric for interpreting studies

- Strength of evidence
  - Continuum:
    - No evidence: No databases pass diagnostics
    - Weak: One database pass diagnostics, lots of databases fail
    - Strong: lots of diverse databases pass and few fail diagnostic (including diagnostic for heterogeneity)
- Certainty in estimate
- Size of effect



# What you need to collaborate on evidence at scale

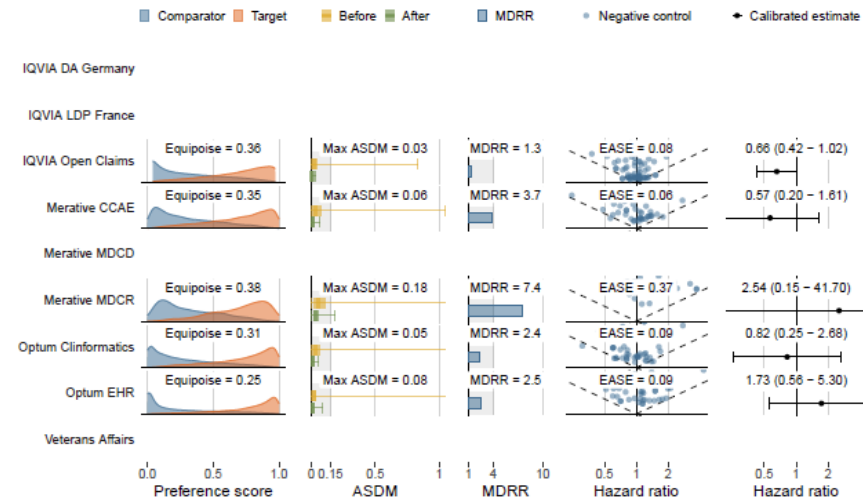
## LEGEND-T2DM Evidence Dissemination Summary

- Target (class): Semaglutide (GLP-1 Receptor Agonists)
- Comparator (class): Glimepiride (Sulfonylureas)
- Outcome: Acute pancreatitis

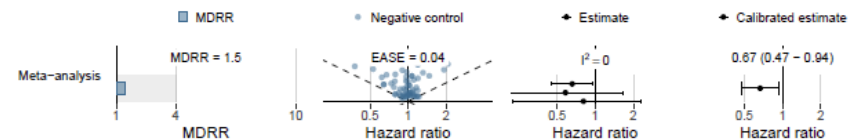
### How Often? (Incidence rates in the PS-matched target cohorts)

| Data source         | Persons exposed | Person-time (yrs) | Persons with outcome | IR (/1,000 PY) |
|---------------------|-----------------|-------------------|----------------------|----------------|
| IQVIA DA Germany    | -               | -                 | -                    | -              |
| IQVIA LDP France    | -               | -                 | -                    | -              |
| IQVIA Open Claims   | 99,708          | 52,939            | 60                   | 1.13           |
| Merative CCAE       | 20,240          | 9,388             | 14                   | 1.49           |
| Merative MDCD       | -               | -                 | -                    | -              |
| Merative MDCR       | 619             | 278               | <5                   | <17.97         |
| Optum Clinformatics | 7,607           | 3,811             | 8                    | 2.10           |
| Optum EHR           | 6,717           | 2,098             | 7                    | 3.34           |
| Veterans Affairs    | 1,258           | 883               | -                    | 0.00           |

### How Reliable Are the Effect Estimates? (Objective diagnostics)



### What have we learned from the OHDSI Network? (Meta-analysis diagnostics and estimate)





# Database diagnostics

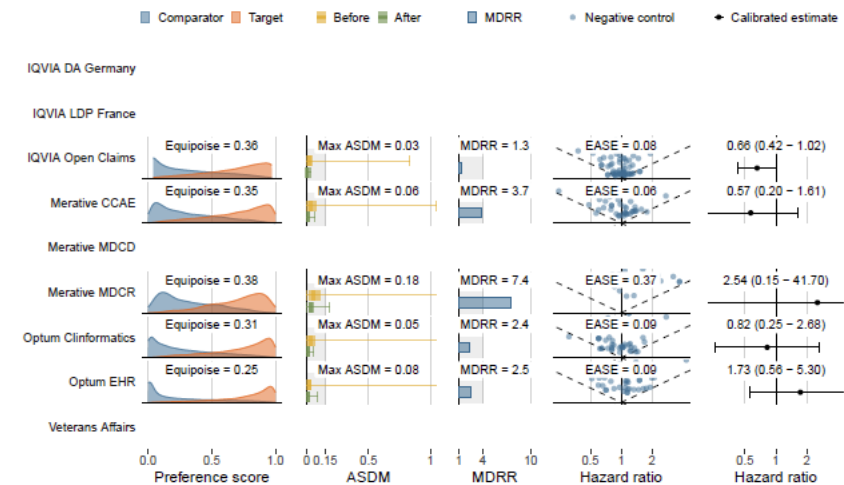
## LEGEND-T2DM Evidence Dissemination Summary

- Target (class): **Semaglutide** (GLP-1 Receptor Agonists)
- Comparator (class): **Glimepiride** (Sulfonylureas)
- Outcome: **Acute pancreatitis**

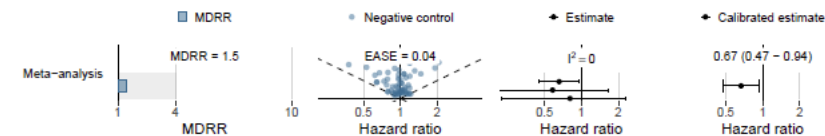
### How Often? (Incidence rates in the PS-matched target cohorts)

| Data source         | Persons exposed | Person-time (yrs) | Persons with outcome | IR (/1,000 PY) |
|---------------------|-----------------|-------------------|----------------------|----------------|
| IQVIA DA Germany    | -               | -                 | -                    | -              |
| IQVIA LDP France    | -               | -                 | -                    | -              |
| IQVIA Open Claims   | 99,708          | 52,939            | 60                   | 1.13           |
| Merative CCAE       | 20,240          | 9,388             | 14                   | 1.49           |
| Merative MDCD       | -               | -                 | -                    | -              |
| Merative MDCR       | 619             | 278               | <5                   | <17.97         |
| Optum Clinformatics | 7,607           | 3,811             | 8                    | 2.10           |
| Optum EHR           | 6,717           | 2,098             | 7                    | 3.34           |
| Veterans Affairs    | 1,258           | 883               | -                    | 0.00           |

### How Reliable Are the Effect Estimates? (Objective diagnostics)



### What have we learned from the OHDSI Network? (Meta-analysis diagnostics and estimate)



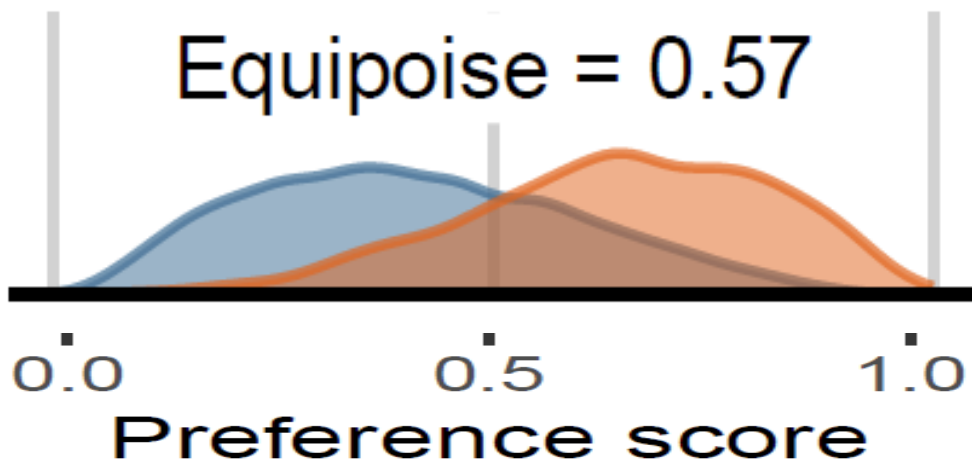


# Empirical Equipoise: Preference score

- Diagnostic: what proportion of the target population is close to treatment indifference?
- Threshold: PASS if Equipoise  $> 0.25$

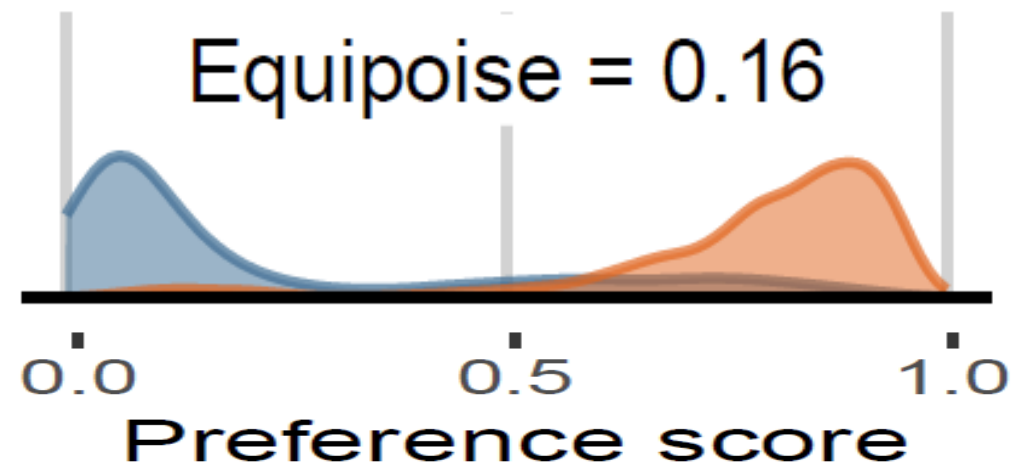
Pass! 😊

Equipoise = 0.57



Fail! 😞

Equipoise = 0.16

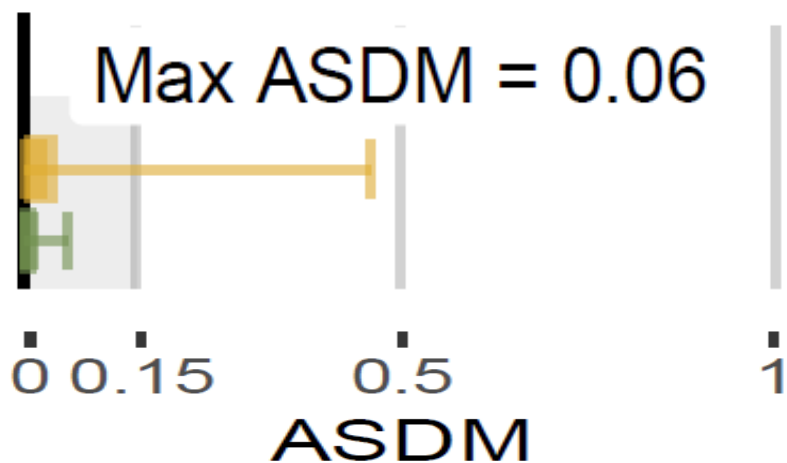




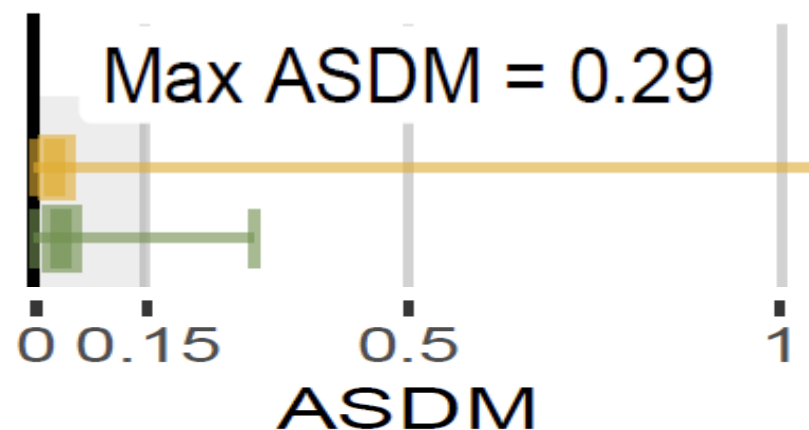
## Covariate balance: Standardized difference of means

- Diagnostic: are all observed baseline characteristics sufficiently similar between target and comparator cohorts?
- Threshold: PASS if Maximum Absolute Standardized Difference of Means after adjustment (Max ASDM)  $< 0.15$

Pass! 😊



Fail! 😞

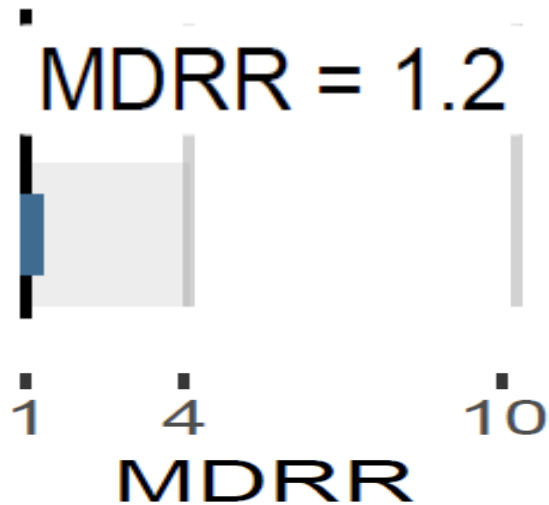




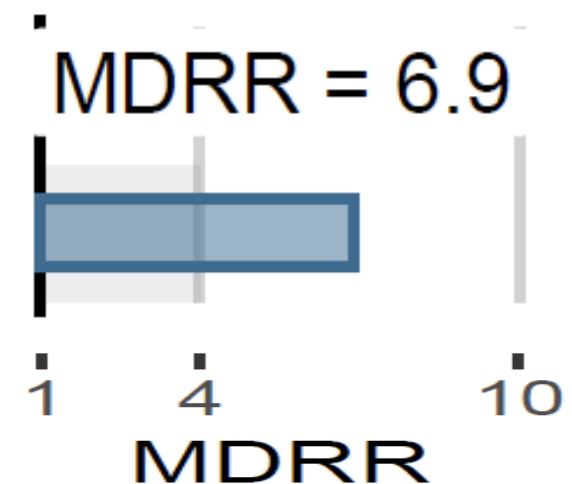
# Statistical Power

- Diagnostic: how much data is sufficient to provide useful information?
- Threshold: PASS if Minimum Detectable Relative Risk (MDRR)  $< 4$

Pass! 😊



Fail! 😞





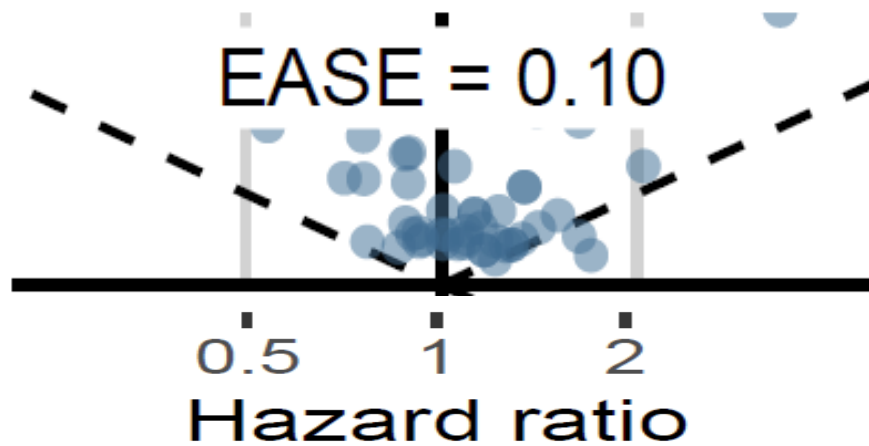


# Residual bias: Expected Absolute Systematic Error (EASE)

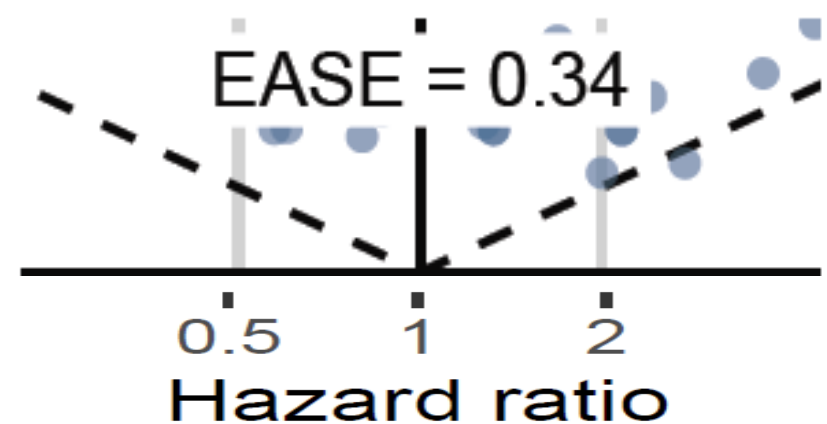
Diagnostic: is the residual bias observed from negative controls small enough to accept that calibrated effect estimates can be trusted as unbiased?

Threshold: PASS if Expected Absolute Systematic Error (EASE)  $< 0.25$

Pass! 😊



Fail! 😞





# Database diagnostics

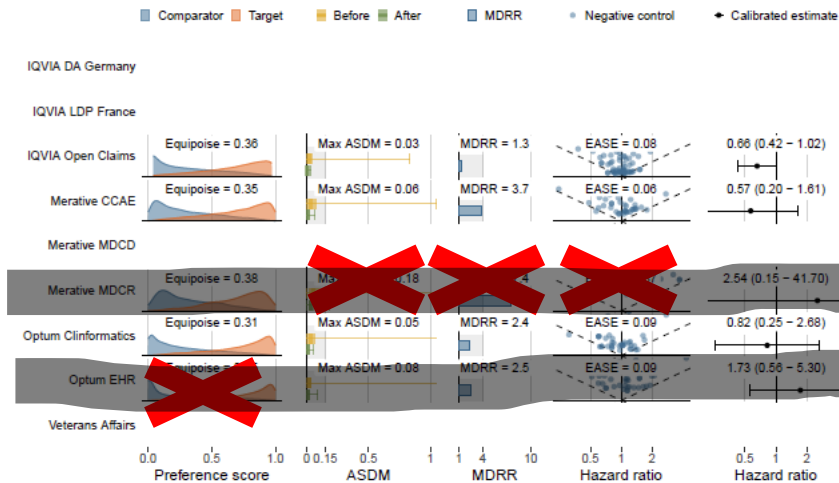
## LEGEND-T2DM Evidence Dissemination Summary

- Target (class): **Semaglutide (GLP-1 Receptor Agonists)**
- Comparator (class): **Glimepiride (Sulfonylureas)**
- Outcome: **Acute pancreatitis**

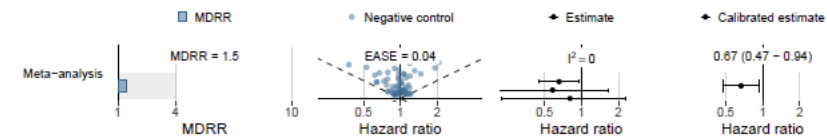
### How Often? (Incidence rates in the PS-matched target cohorts)

| Data source         | Persons exposed | Person-time (yrs) | Persons with outcome | IR (/1,000 PY) |
|---------------------|-----------------|-------------------|----------------------|----------------|
| IQVIA DA Germany    | -               | -                 | -                    | -              |
| IQVIA LDP France    | -               | -                 | -                    | -              |
| IQVIA Open Claims   | 99,708          | 52,939            | 60                   | 1.13           |
| Merative CCAE       | 20,240          | 9,388             | 14                   | 1.49           |
| Merative MDCD       | -               | -                 | -                    | -              |
| Merative MDCR       | 619             | 278               | <5                   | <17.97         |
| Optum Clinformatics | 7,607           | 3,811             | 8                    | 2.10           |
| Optum EHR           | 6,717           | 2,098             | 7                    | 3.34           |
| Veterans Affairs    | 1,258           | 883               | -                    | 0.00           |

### How Reliable Are the Effect Estimates? (Objective diagnostics)



### What have we learned from the OHDSI Network? (Meta-analysis diagnostics and estimate)





# What if you failed a database diagnostic?

- If you didn't evaluate your diagnostic, you may have inappropriately reported unreliable evidence
  - How many papers have you read that give you an estimate but don't convince you that they have reliable evidence?
- Reporting you failed a diagnostics is informative to understand if issue is design or data or both
- The value of a diverse data network, it is possible that some may fail and others will pass forward



# Meta-analysis diagnostics

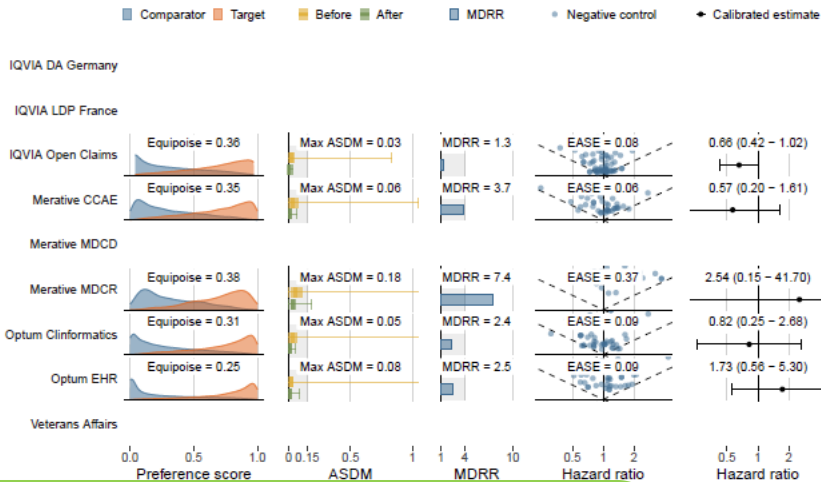
## LEGEND-T2DM Evidence Dissemination Summary

- Target (class): **Semaglutide (GLP-1 Receptor Agonists)**
- Comparator (class): **Glimepiride (Sulfonylureas)**
- Outcome: **Acute pancreatitis**

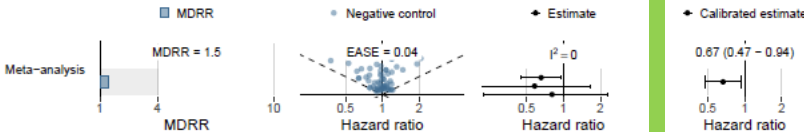
### How Often? (Incidence rates in the PS-matched target cohorts)

| Data source         | Persons exposed | Person-time (yrs) | Persons with outcome | IR (/1,000 PY) |
|---------------------|-----------------|-------------------|----------------------|----------------|
| IQVIA DA Germany    | -               | -                 | -                    | -              |
| IQVIA LDP France    | -               | -                 | -                    | -              |
| IQVIA Open Claims   | 99,708          | 52,939            | 60                   | 1.13           |
| Merative CCAE       | 20,240          | 9,388             | 14                   | 1.49           |
| Merative MDCD       | -               | -                 | -                    | -              |
| Merative MDCR       | 619             | 278               | <5                   | <17.97         |
| Optum Clinformatics | 7,607           | 3,811             | 8                    | 2.10           |
| Optum EHR           | 6,717           | 2,098             | 7                    | 3.34           |
| Veterans Affairs    | 1,258           | 883               | -                    | 0.00           |

### How Reliable Are the Effect Estimates? (Objective diagnostics)



### What have we learned from the OHDSI Network? (Meta-analysis diagnostics and estimate)

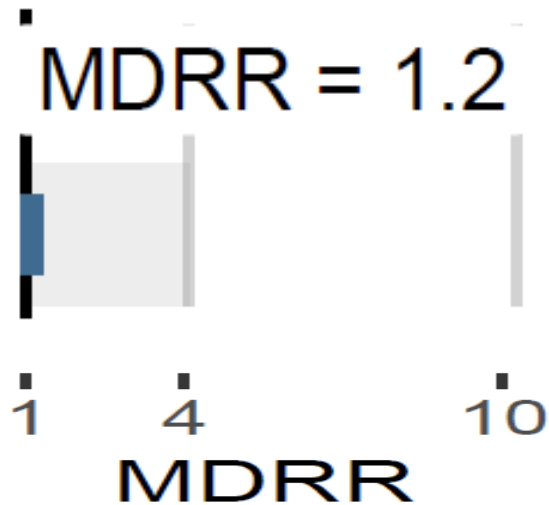




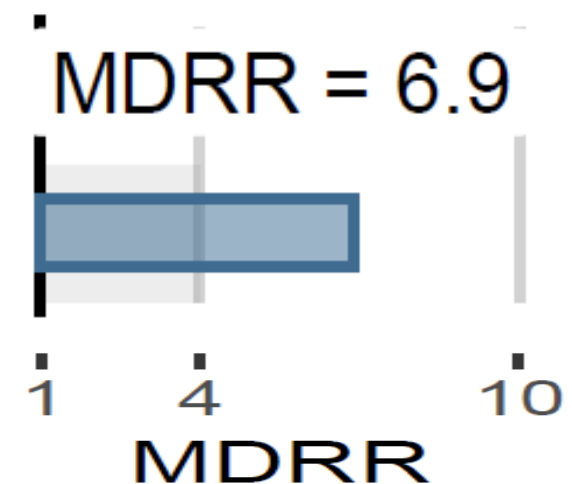
# Statistical Power

- Diagnostic: how much data is sufficient to provide useful information?
- Threshold: PASS if Minimum Detectable Relative Risk (MDRR)  $< 4$

Pass! 😊



Fail! 😞





## Residual bias:

### Expected Absolute Systematic Error (EASE)

Diagnostic: is the residual bias observed from negative controls small enough to accept that calibrated effect estimates can be trusted as unbiased?

Threshold: PASS if Expected Absolute Systematic Error (EASE)  $< 0.25$

Pass! 😊

EASE = 0.10

0.5 1 2  
Hazard ratio

Fail! 😞

EASE = 0.34

0.5 1 2  
Hazard ratio

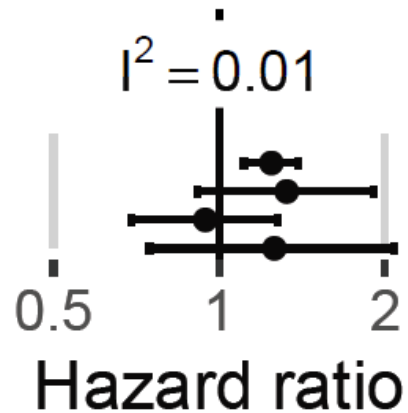
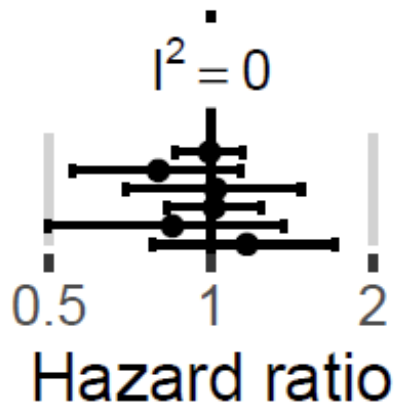


# Heterogeneity

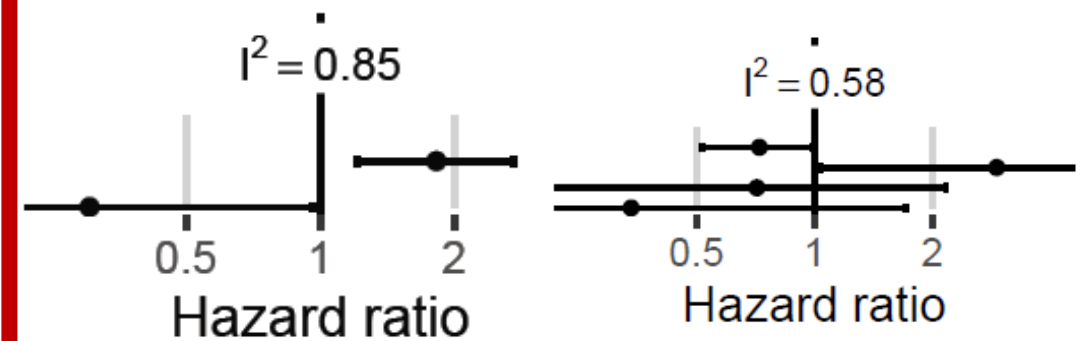
Diagnostic: Do the databases provide consistent estimates of the magnitude of the effect?

Threshold: PASS if  $I^2 < 0.40$

Pass! 😊



Fail! 😞





# Meta-analysis diagnostics

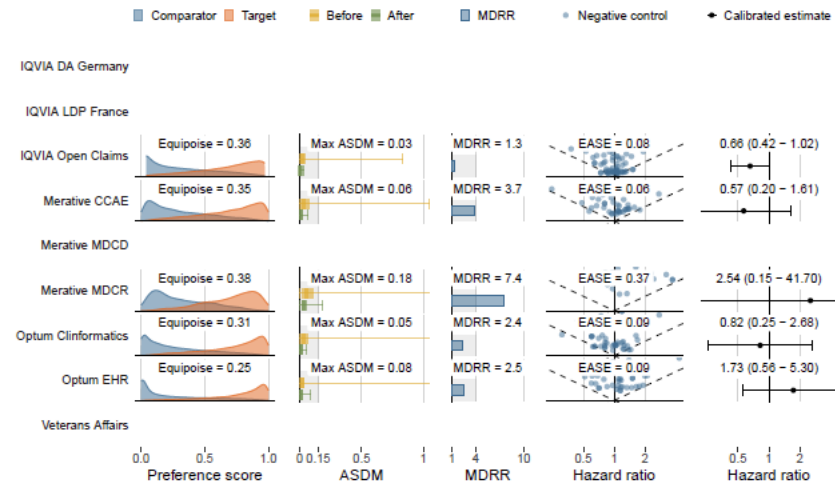
LEGEND-T2DM Evidence Dissemination Summary

- Target (class): Semaglutide (GLP-1 Receptor Agonists)
- Comparator (class): Glimepiride (Sulfonylureas)
- Outcome: Acute pancreatitis

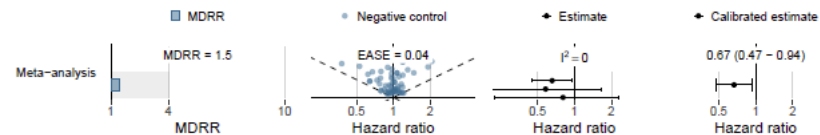
## How Often? (Incidence rates in the PS-matched target cohorts)

| Data source         | Persons exposed | Person-time (yrs) | Persons with outcome | IR (/1,000 PY) |
|---------------------|-----------------|-------------------|----------------------|----------------|
| IQVIA DA Germany    | -               | -                 | -                    | -              |
| IQVIA LDP France    | -               | -                 | -                    | -              |
| IQVIA Open Claims   | 99,708          | 52,939            | 60                   | 1.13           |
| Merative CCAE       | 20,240          | 9,388             | 14                   | 1.49           |
| Merative MDCD       | -               | -                 | -                    | -              |
| Merative MDCR       | 619             | 278               | <5                   | <17.97         |
| Optum Clinformatics | 7,607           | 3,811             | 8                    | 2.10           |
| Optum EHR           | 6,717           | 2,098             | 7                    | 3.34           |
| Veterans Affairs    | 1,258           | 883               | -                    | 0.00           |

## How Reliable Are the Effect Estimates? (Objective diagnostics)



## What have we learned from the OHDSI Network? (Meta-analysis diagnostics and estimate)



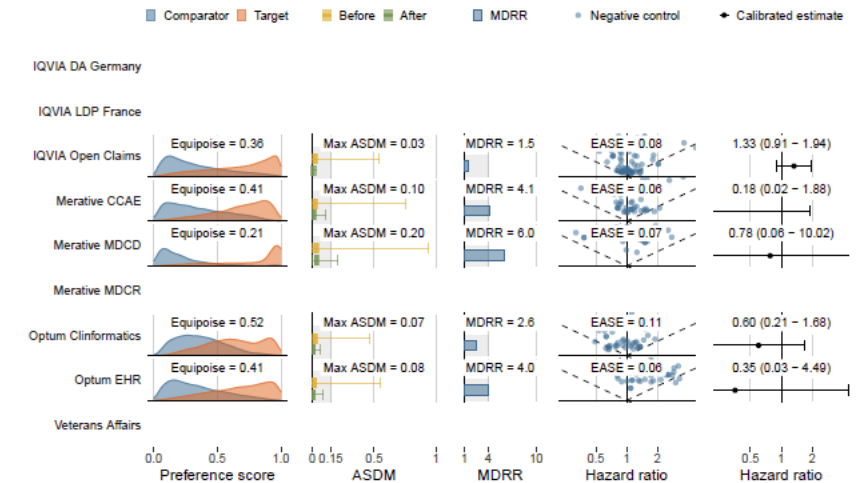
LEGEND-T2DM Evidence Dissemination Summary

- Target (class): Linagliptin (DPP-4 inhibitors)
- Comparator (class): Dulaglutide (GLP-1 Receptor Agonists)
- Outcome: Acute pancreatitis

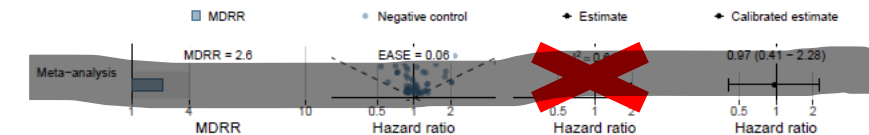
## How Often? (Incidence rates in the PS-matched target cohorts)

| Data source         | Persons exposed | Person-time (yrs) | Persons with outcome | IR (/1,000 PY) |
|---------------------|-----------------|-------------------|----------------------|----------------|
| IQVIA DA Germany    | -               | -                 | -                    | -              |
| IQVIA LDP France    | -               | -                 | -                    | -              |
| IQVIA Open Claims   | 59,428          | 44,537            | 86                   | 1.93           |
| Merative CCAE       | 3,107           | 2,206             | <5                   | <2.27          |
| Merative MDCD       | 629             | 397               | <5                   | <12.59         |
| Merative MDCR       | 408             | 330               | <5                   | <15.13         |
| Optum Clinformatics | 6,063           | 4,449             | 10                   | 2.25           |
| Optum EHR           | 4,309           | 897               | <5                   | <5.58          |
| Veterans Affairs    | -               | -                 | -                    | -              |

## How Reliable Are the Effect Estimates? (Objective diagnostics)



## What have we learned from the OHDSI Network? (Meta-analysis diagnostics and estimate)





# Strength of evidence

- How many databases do you have that pass diagnostics?
  - None
  - At least 1
  - $>1$
  - $>2$
  - $>3$
  - $>4$
- Do you have results from  $\geq 1$  claims and  $\geq 1$  EHR?
- Do you have results from  $\geq 1$  in US and  $\geq 1$  in Europe?



# Meta-analysis estimate

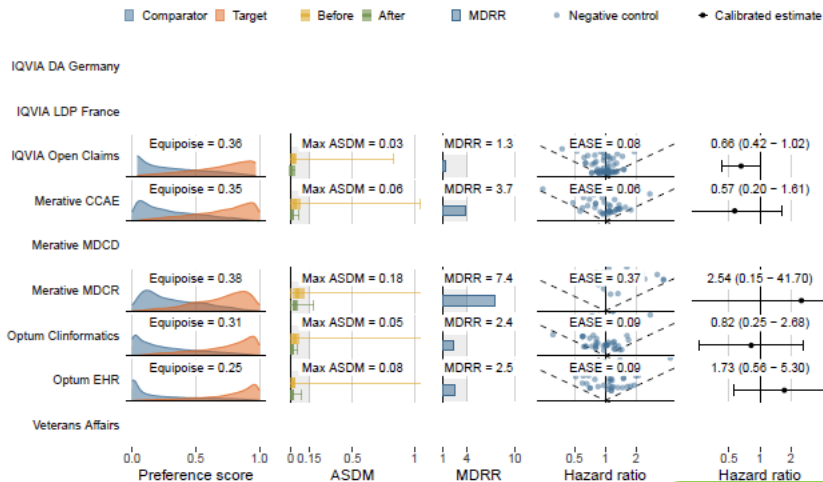
## LEGEND-T2DM Evidence Dissemination Summary

- Target (class): **Semaglutide (GLP-1 Receptor Agonists)**
- Comparator (class): **Glimepiride (Sulfonylureas)**
- Outcome: **Acute pancreatitis**

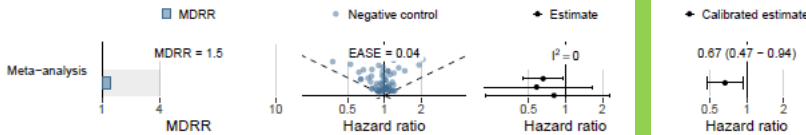
### How Often? (Incidence rates in the PS-matched target cohorts)

| Data source         | Persons exposed | Person-time (yrs) | Persons with outcome | IR (/1,000 PY) |
|---------------------|-----------------|-------------------|----------------------|----------------|
| IQVIA DA Germany    | -               | -                 | -                    | -              |
| IQVIA LDP France    | -               | -                 | -                    | -              |
| IQVIA Open Claims   | 99,708          | 52,939            | 60                   | 1.13           |
| Merative CCAE       | 20,240          | 9,388             | 14                   | 1.49           |
| Merative MDCD       | -               | -                 | -                    | -              |
| Merative MDCR       | 619             | 278               | <5                   | <17.97         |
| Optum Clinformatics | 7,607           | 3,811             | 8                    | 2.10           |
| Optum EHR           | 6,717           | 2,098             | 7                    | 3.34           |
| Veterans Affairs    | 1,258           | 883               | -                    | 0.00           |

### How Reliable Are the Effect Estimates? (Objective diagnostics)



### What have we learned from the OHDSI Network? (Meta-analysis diagnostics and estimate)





# Certainty of estimate

- Rule of thumb: Divide upper bound by the lower bound, if its  $< 1.5$ , that's pretty certain. If its  $> 3$ , that's pretty uncertain.
- Rule of pinky: If your confidence interval is wider than your pinky finger, than you are pretty uncertain (subject to measurement error)



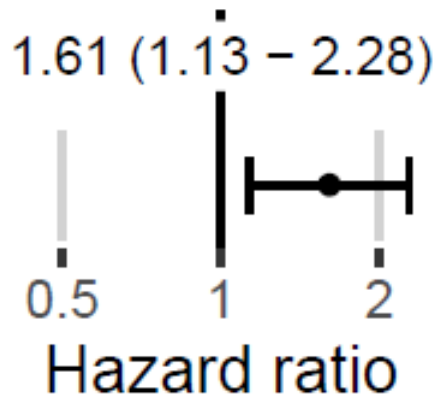
All estimates are equally good,  
there's nothing magic about  
 $p < 0.05$

---

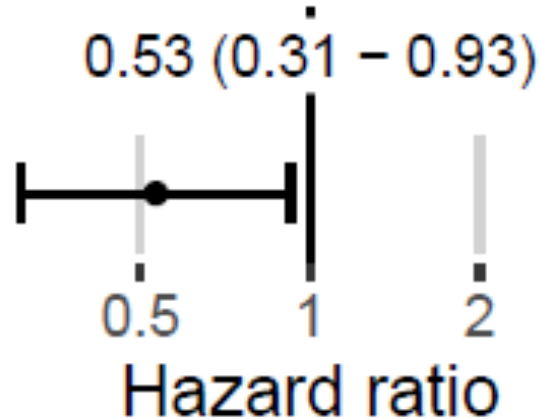


# You could be holding reliable evidence

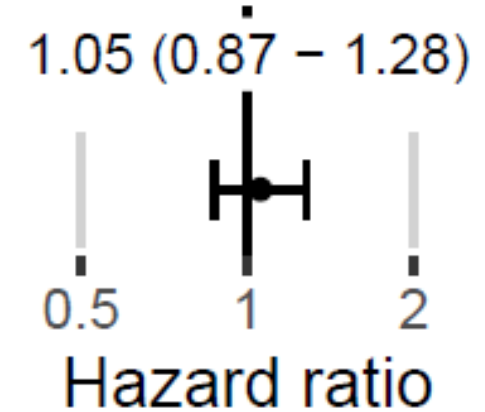
• Calibrated estimate



• Calibrated estimate



• Calibrated estimate



Ask not what the p-value can do for you.....instead ask yourself:

- What is the quality of the study?
- What is the strength of evidence?
- What is the certainty of estimate?



# Magnitude of effect

- All estimates are equally good, there's nothing magic about  $p < 0.05$ 
  - Example: high strength, high certainty,  $RR=1$  (0.98 – 1.02) is extremely informative because high likelihood of no difference
  - Publication bias is a problem, let's not contribute to that, we all need to disseminate whether you have a positive effect or not
- Need characterization plus estimation results for full context
  - Incidence rate: provides absolute (background + attributable) risk
  - Relative risk: provides causal estimate for counterfactual or comparisons
- We need to translate estimate and its uncertainty into recommendations for action based on clinical context and utility





# What if you failed your meta-analysis diagnostics?

- Most interesting scenario: You've observed heterogeneity, now what?
  - Reflection: is there true population differences or differential bias?
  - Action: Report the database estimates and learn why those estimates are heterogeneous
  - This scenario is both why we need a diverse data network and a great opportunity for more methodological research!

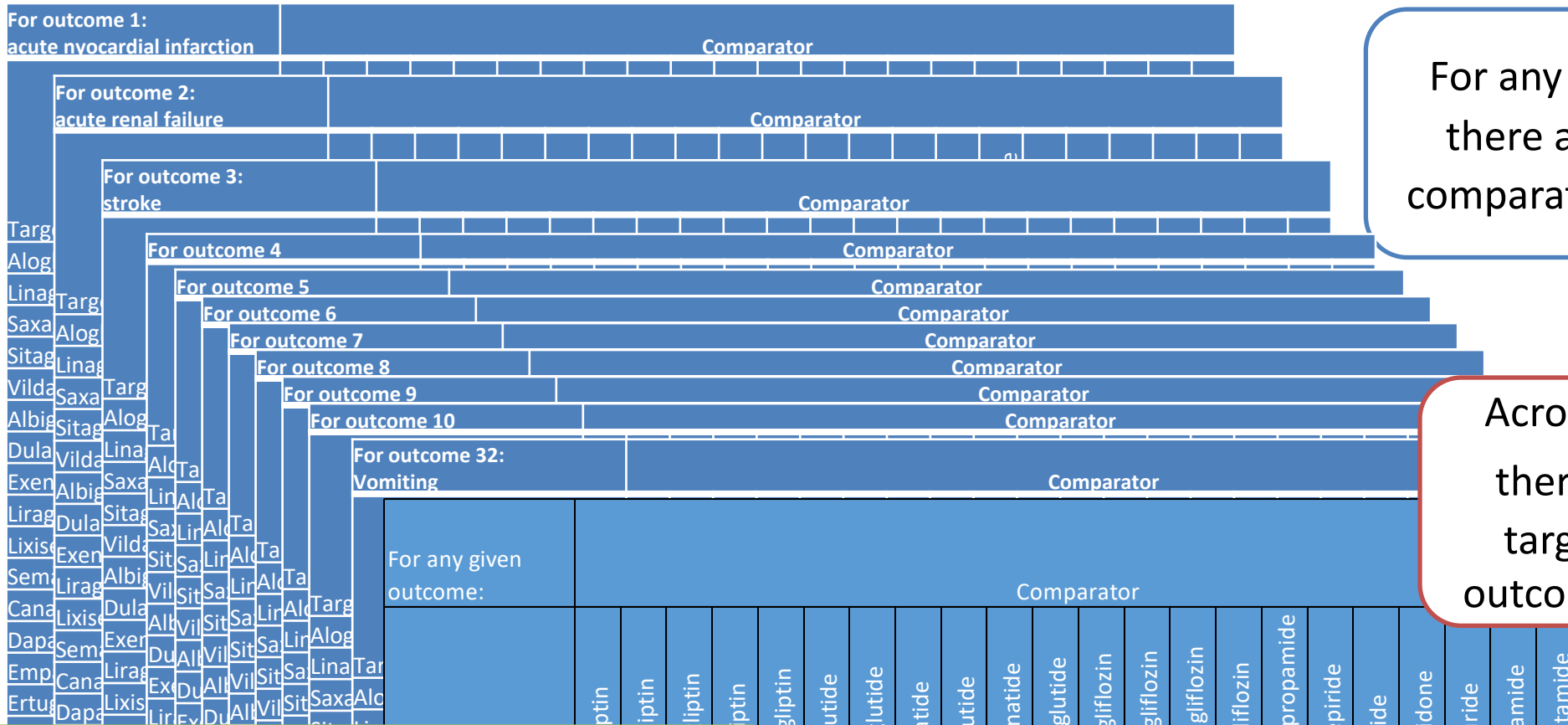


# What if you failed all database diagnostics?

- Informative: this design with this network is incapable of answering the question
- LEGEND is the biggest network study ever undertaken and still can't answer many questions



# The scale of LEGEND-T2DM



For any given outcome, there are **462** target-comparator combinations

Across 32 outcomes, there are **14,784** target-comparator-outcome combinations

Every target-comparator-outcome is a valid clinical question that patients deserve reliable evidence for...

...it is our responsibility to communicate the reliable evidence we have to everyone who can benefit, and to also communicate when we can't produce reliable evidence



## Next steps

- Most of you have in your hand a piece of reliable evidence that could be published and make a difference
  - All of you have permission to disseminate that evidence, and explicit encouragement from the LEGEND leadership and OHDSI community to do so (just coordinate with the LEGEND team and stick to our OHDSI authorship guidelines)
  - Or even better, all of you made friends today and a dream team to move forward on the opportunity together to impact the health of millions of patients
-





# Thank you for joining the journey!

## Now its Group Photo time out in the Atrium!



## And then come back to the Ballroom for Game Night!