Synthesizing Evidence for Rare Events: a Novel Zero-Inflated Bivariate Model to Integrate Studies with Double-Zero Outcomes

Lu Li, Ph.D. candidate at the University of Pennsylvania
Advisor: Dr. Yong Chen
Joint work with Drs. Lifeng Lin, Haitao Chu, Yong Chen
2023 OHDSI Symposium
Real-world case study

Explores the potential use of probiotics as a treatment for Clostridium difficile-associated diarrhea (CDAD) caused by antibiotic use

Whether probiotics cause any side effects when used to prevent CDAD
### Analysis 1.24. Comparison 1 Probiotics versus control, Outcome 24 Adverse Events: complete case.

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Experimental</th>
<th>Control</th>
<th>Risk Ratio</th>
<th>Weight</th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>n/N</td>
<td>n/N</td>
<td>M-H, Random, 95% CI</td>
<td>M-H, Random, 95% CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allen 2013</td>
<td>294/1470</td>
<td>284/1471</td>
<td>1.04[0.9,1.2]</td>
<td>13.09%</td>
<td></td>
</tr>
<tr>
<td>Arvola 1999</td>
<td>0/61</td>
<td>0/58</td>
<td>Not estimable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Double-zero-event study (DZS)**
Background

Relative Risk

\[ \frac{0.57}{0.53} \]

Undefined

\[ \frac{0.527}{0.473} \]

Undefined

\[ \frac{1.344}{1.144} \]

Non-event

Event

How should we deal with the studies with double zeros?

DARWIN EU*

* Data Analysis and Real-world Interrogation Network
Should we drop them?

**ORIGINAL ARTICLE**

Exclusion of studies with no events in both arms in meta-analysis impacted the conclusions

Chang Xu, Ling Li, Lifeng Lin, Haitao Chu, Lehana Thabane, Kang Zou, Xin Sun

*Chung Medicine Program, Technical University, Shanghai University, China*

**Stat Methods Med Res.** Author manuscript; available in PMC 2013 Dec 1.

Published in final edited form as:

Published online 2010 Dec 21. doi: 10.1177/0962280210393712

Bivariate Random Effects Models for Meta-Analysis of Comparative Studies with Binary Outcomes: Methods for the Absolute Risk Difference and Relative Risk

Haitao Chu, Lei Nie, Yong Chen, Yi Huang, and Wei Sun
Existing approach to incorporate DZS:

**Bivariate Generalized Linear Mixed Model (BGLMM)**

- A bivariate random effects model that *jointly* analyzes the risks in treatment and control groups

<table>
<thead>
<tr>
<th>Group (for the <em>i</em>-th study)</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of events</td>
<td>(Y_{i1})</td>
<td>(Y_{i0})</td>
</tr>
<tr>
<td>Sample size</td>
<td>(N_{i1})</td>
<td>(N_{i0})</td>
</tr>
<tr>
<td>Event risk</td>
<td>(P_{i1})</td>
<td>(P_{i0})</td>
</tr>
<tr>
<td>Fixed effects</td>
<td>(\mu_1)</td>
<td>(\mu_0)</td>
</tr>
<tr>
<td>Random effects</td>
<td>(v_{i1})</td>
<td>(v_{i0})</td>
</tr>
</tbody>
</table>

\[ Y_{ik} \sim \text{Binomial} \left( N_{ik}, P_{ik} \right); \quad g(P_{ik}) = \mu_k + v_{ik} \]

\[
\mathbb{P}(Y_{i0} = y_{i0}, Y_{i1} = y_{i1}) = \prod_{k=0}^{1} (P_{ik})^{y_{ik}} (1 - P_{ik})^{N_{ik} - y_{ik}}
\]

**Limitation:**

BGLMM treats all DZS similarly to the other studies.

Should we treat DZS similarly to the other studies?

Short answer: **NO**

**Rationale:** Sample size of the studies are informative

**Conclusion:** We **should not** treat all the double zero studies the same as the other studies

**Assuming an event rate of 1%:**

Could happen by chance.  Very unlikely.

0.99^{110} \approx \frac{33}{100}

0.99^{1,000} \approx \frac{4}{100,000}
To differentiate DZS: Zero-Inflated Models

- Zero-inflated models separate observed zeros into two distinct categories.

\[ Y_{ik} \sim \begin{cases} \text{Binomial } (N_{ik}, P_{ik}), & \text{with probability } 1 - \pi \\ 0, & \text{with probability } \pi \end{cases} \]

1 – \( \pi \): at-risk population

“at-risk” or “chance” zeros correspond to a latent group of individuals who are at risk for an event but have a recorded count of zeros.

\( \pi \): low-risk population

“structural” zeroes represent individuals who are not susceptible to a specific event, thereby having no chance of a positive count.

Proposed method

- Zero-Inflated Bivariate Generalized Linear Mixed Model (ZIBGLMM)

**Advantages:**

1. No studies are dropped from the analysis.

$1 - \pi$: at-risk population

$\pi$: low-risk population

Recap BGLMM:

$$P(Y_{i0} = y_{i0}, Y_{i1} = y_{i1}) = \prod_{k=0}^{1} (P_{ik})^{y_{ik}} (1 - P_{ik})^{N_{ik} - y_{ik}}$$

$$g(P_{ik}) = \mu_k + \nu_{ik}$$
Revisit the case study

- 10 of 32 studies are double-zero-event studies (DZS), with sample size ranging from 18 to 144.
- Concluded that probiotics reduce the risk of AE by 17%:
  - RR 0.83 (95% CI 0.71 to 0.97)
- Using our proposed method (ZIBGLMM):
  - RR 0.70 (95% CI 0.55 to 0.88)
- Conclusion:
  - Including the DZS could potentially result in estimates that differ by a large degree (>0.1).
  - Using ZIBGLMM offers a more comprehensive analysis of the available data.
Summary

- Zeros in double-zero-event studies (DZS) may arise due to heterogeneity in the population.
- ZIBGLMM offers a more comprehensive analysis of the available data.

For **OHDSI**, ZIBGLMM is useful especially for **larger network studies** and for studies involving **rare events**.
Correspondence to:

Yong Chen, Ph.D.
Email: ychen123@upenn.edu

Lu Li
Email: luli1@sas.upenn.edu

Acknowledgments:
Jiayi Tong,
Qiong Wu, Ph.D.,
Dazheng Zhang,
Bingyu Zhang,
Jiajie Chen, Ph.D.,
Tianyu Zhang,
and other lab members for your help with the project.

👉 Poster: # 506
👉 Fri 10/20 4:15 – 5:00 pm

View our code at:
Overview of ASSURE
OHDSI Symposium 2023

Kevin Haynes
Justin Bohn
Jenna Reps
Gowtham Rao
Mitch Conover
Standardizing regulatory-grade real-world evidence generation

Standardized design
- Indication
- Target
- Outcome
- Comparator
- Time-at-risk

Standardized analytics
- Open community data standards
- Methodological research
- Open-source development

JNJ/Epi standardized data network (OMOP CDM)
- US private claims x3
- US EHR x2
- US Hospital
- US Medicaid
- US Medicare suppl
- Germany EHR
- France EHR
- Australia EHR
- Japan claims

Standardized evidence
- Characterization
  - Disease natural history
  - Treatment utilization
  - Outcome Incidence
  - Time-to-event
- Population-level estimation
  - Comparative cohort
  - Self-controlled case series
- Patient-level prediction

Innovation
Transforming RWE generation from bespoke studies taking months to a systematic process taking days, while enabling transparent reproducibility and ensuring scientific best practices in causal inference and machine learning

Use cases
Current focus:
- Safety signal detection and evaluation
- Enhanced surveillance
Future opportunities:
- Comparative effectiveness
- Disease interception

Results delivered in 2023
- 23 Requests
- Impact on regulatory decision making
Where does ASSURE fit into the life of a safety signal?

- Early awareness of signals enables preparation and validation of input specifications
- Standardization enables evidence generation within a short timeline

**Standardized inputs**  **Standardized analytics**  **Standardized databases**  **Standardized results**
ASSURE Analyses: Inputs and Outputs

- Analysis designer
  - ATLAS/CIRCE Cohort Design interface
  - Analysis parameter selection interface

- Phenotype library
  - Clinical descriptions
  - Cohort definitions
  - Diagnostics
  - Measurement error estimates

- Specifications
  - Cohorts
  - Parameters
  - Databases

- Cohort Generation
  - Indication I
  - Treatment T
  - Subgroup S
  - Outcome O

- Execution engine
  - Characterization
    - Treatment pathways (T in I)
    - Temporal characterization (I, T, O in S)
    - Incidence (O in T/I, in S)
    - Prevalence (O, T in I, in S)
    - Treatment utilization (T, in S)
  - Population-Level Estimation
    - Absolute effects (O in T1)
    - Comparative effects (O in T1vT2)
    - Designs: CM, SCCS, SCC
    - Use case: Identification, Estimation
  - Patient-level Prediction
    - Baseline risk
    - Attributable risk

- Data analysis infrastructure 1
  - CDM1
  - CDM2
  - CDM3

- Infrastructure j
  - CDM k

- Results viewer
  - Interactive exploratory interface
    - Characterization
    - Estimation
      - Identification
      - Estimation
      - Prediction
  - Self-contained static standardized reporting
    - Specs doc
    - Study report
    - Supplemental materials

- • 164 Janssen products
  • 935 alternate treatments
  • 39 treatment indications
  • 45 outcome events
1. Treatment/Comparator/Indication/Outcome
   • Comparator Selection Tool
2. Phenotype Development
   • Disease Advisory Board
3. Analytic Design and Implementation
   • Negative Control Selection
   • Time at Risk Selection
4. Result Interpretation
   • Shiny Dashboard
5. Documentation and Communication
   • Standardized Output

A Day in the Life of the ASSURE Team
tcis <- list(
  list(
    targetId = 13771,
    comparatorId = 13774,
    indicationId = NULL,
    genderConceptIds = c(8507, 8532), # use valid minAge = 18, # Age 18+. Can be NULL
    maxAge = NULL, # No max age. Can be NULL
    excludedCovariateConceptIds = c(1154029, 1103640)
  )
)

sccsTi <- list(
  list(
    targetId = 13771,
    indicationId = NULL, # NO INDICATION REQUIRED
    genderConceptIds = c(8507, 8532), # use valid minAge = 18, # Age 18+. Can be NULL
    maxAge = NULL # No max age. Can be NULL
  )
)

outcomes <- tibble(
  cohortId = c(12308),
  cleanwindow = c(90)
)

negativeConcepSetId <- 5749

timeAtRisks <- tibble(
  label = c("On-treatment"),
  riskWindowStart = c(1),
  startAnchor = c("cohort start"),
  riskWindowEnd = c(0),
  endAnchor = c("cohort end"),
)

# Try to avoid intent-to-treat TARs for SCCS:
sccsTimeAtRisks <- tibble(
  label = c("On-treatment"),
  riskWindowStart = c(1),
  startAnchor = c("cohort start"),
  riskWindowEnd = c(0),
  endAnchor = c("cohort end"),
)

# Try to use fixed-time TARs for PLP:
plpTimeAtRisks <- tibble(
  riskWindowStart = c(1),
  startAnchor = c("cohort start"),
  riskWindowEnd = c(365),
  endAnchor = c("cohort start"),
)

studyStartDate <- "" # YYYYMMDD
studyEndDate <- "" # YYYYMMDD
Patient’s outcomes after endoscopic retrograde cholangiopancreatography (ERCP) using reprocessed duodenoscope accessories: a descriptive study using real-world data

Jessica Mayumi Maruyama
Eduardo Sleiman Beljavskis
Laila Colações
Lisandry Aquino
Renata Martins
Sarah Rodrigues
Suellen dos Santos
Julio Cesar Barbour Oliveira
1. Background

Concerns related to duodenoscope-related infections due to material reprocessing.

ERCP: Significant impact on management and prognosis of biliary and pancreatic diseases.

Study objective using an OMOP CDM harmonized dataset from Brazil:
- To compare the % of readmissions post-ERCP between Single-use (SUG) and Non-single-use (NSUG) institutions.

Source: https://www.sages.org/
2. Methods

Data source: Hospital and Ambulatory Information System from Brazilian Administrative Database, mapped to OMOP CDM v 5.4. A deterministic linkage algorithm was developed to connect hospitals with outpatient records using the key information of zip code, date of birth, and gender.

Study period: January 2020 to January 2023

- Patients with no history of cancer
- ERCP procedure, excluding due to sepsis, acute pancreatitis, or cholangitis
- Readmission within 30 day
- Causes for readmission: sepsis, acute pancreatitis, or cholangitis
3. Methods

Identification of ERCP procedures:

Specific SUS coding system, named Table of the Procedure, Medication, Orthotics, Prosthetics, and Special Materials Management System of the SUS (SIGTAP)

Statistical analysis: Atlas

Identification of SUG and NSUG hospitals:

- 3 Single-use institutions
- 15 Non-single use institutions
4. Results

Table 1. Descriptive information of total and readmitted patients in SUG and NSUG groups

<table>
<thead>
<tr>
<th></th>
<th>SUG</th>
<th>NSUG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Readmitted patients</td>
</tr>
<tr>
<td>N</td>
<td>669</td>
<td>20</td>
</tr>
<tr>
<td>Male (%)</td>
<td>30.9</td>
<td>50.0</td>
</tr>
<tr>
<td>Mean age (SD)</td>
<td>55.0 (19.0)</td>
<td>55.0 (17.9)</td>
</tr>
</tbody>
</table>

Note. SUG – single-use group; NSUG – non-single-use group; SD – standard deviation;
Readmitted patients included patients who were hospitalized within 30 days after a patient’s ERCP due to sepsis, acute pancreatitis, or cholangitis.

In comparison to the readmitted patients from SUG, the readmitted patients from NSUG had a higher proportion of female individuals and patients with a lower mean age.
5. Results

**Non-Single-Use (NSUG)**
Readmission: 4.8% (43)
- No readmission: 844
- Readmission within 30 days: 43

**Single-Use (SUG)**
Readmission: 2.9% (20)
- No readmission: 649
- Readmission within 30 days: 20

**Difference between NSUG Group and SUG Group:**
The NSUG group had a percentage of readmissions approximately 65% higher compared to the SUG group.
6. Conclusion and next steps

Real-world data from Brazilian administrative dataset

Higher % of readmissions in NSUG institutions compared to SUG institutions

Next step: estimation study adjusting for confounders and unbalanced data

Inform clinical decision-making and optimal ERCP management practices
Does COVID-19 Increase Racial/Ethnic Differences in Prevalence of PASC/Long COVID in Children and Adolescents?
— Findings from Difference-in-Differences Analyses using an EHR-Based Cohort from the RECOVER Program

Bingyu Zhang
PhD student, University of Pennsylvania
Advisor: Dr. Yong Chen
2023 OHDSI Symposium, October 20
What is PASC?

For some people, symptoms last weeks or months after the acute infection has passed. For other people, new symptoms may appear after the acute infection has passed whether they had symptoms during the acute infection or not. Together, these and other health effects of the virus are called post-acute sequelae of SARS-CoV-2 infection, or PASC.


https://recovercovid.org/long-covid
https://covid19community.nih.gov/what-you-need-to-know-about-long-covid
RECOVER: Researching COVID to Enhance Recovery

- The National Institutes of Health (NIH) created the RECOVER Initiative to learn about the long-term effects of COVID
- The goal of RECOVER is to rapidly improve our understanding of and ability to predict, treat, and prevent PASC

- **PI for pediatric RECOVER:**
  - Christopher Forrest (Children’s Hospital of Philadelphia)
- **PI for adult RECOVER:**
  - Rainu Kaushal (Weill Cornell)
- **Biostatistics Core Director:**
  - Yong Chen
  - for PCORnet Pediatric RECOVER

[https://recovercovid.org](https://recovercovid.org)
Selected Publications on PASC within RECOVER

Identifying who has long COVID in the USA: a machine learning approach using N3C data

Emily R Pfaff 1, Andrew T Girvin 2, Tellen D Bennett 3, Abhishek Bhatia 4, Ian M Brooks 5, Rachel R Deer 6, Jonathan P Dekermanian 7, Sarah Elizabeth Jolley 8, Michael G Kahn 9, Kristin Kostka 10, Julie A McMurry 11, Richard Moffitt 12, Anita Walden 13, Christopher G Chute 13, Melissa A Haendel 11, N3C Consortium

Clinical Features and Burden of Postacute Sequelae of SARS-CoV-2 Infection in Children and Adolescents

Suchitra Rao 1, Grace M Lee 2, Hanieh Razeghi 3, Vitaly Lorman 3, Asuncion Mejias 4, Nathan M Pajor 5, Deepika Thacker 6, Ryan Webb 3, Kimberley Dickinson 3, I. Charles Bailey 3, Ravi Jhaveri 7, Dimitri A Christakis 8, 9, Tellen D Bennett 1, Yong Chen 10, Christopher B Forrest 3

Cardiac Complications After SARS-CoV-2 Infection and mRNA COVID-19 Vaccination – PCORnet, United States, January 2021–January 2022

Jason P Block, Tegan K Boehmer, Christopher B Forrest, Thomas W Carton, Grace M Lee, Umed A Ajani, Dimitri A Christakis, Lindsay G Cowell, Christine Draper, Nidhi Chiddayal, Aaron M Harris, Michael D Kappelman, Jean Y Ko, Kenneth H Mayer, Kshmara Nagavedu, Matthew E Oster, Anuradha Paranjape, Jon Puro, Matthew D Ritchey, David K Shay, Deepika Thacker, Adi V Gundlapalli

Data-driven analysis to understand long COVID using electronic health records from the RECOVER initiative

Chengxi Zang 1, Yongkang Zhang 1, Jie Xu 2, Jiang Bian 2, Dmitry Morozyk 1, Edward J Schenck 3, Drhuv Khullar 1, Anna S Nordvig 4, Elizabeth A Shenkman 2, Russell L Rothman 5, Jason P Block 6, Kristin Lyman 7, Mark G Weiner 1, Thomas W Carton 7, Fei Wang 8, Rainu Kaushal 1

Long COVID risk and pre–COVID vaccination in an EHR–based cohort study from the RECOVER program

M Daniel Brannock 1, Robert F Chew 2, Alexander J Preiss 2, Emily C Hadley 2, Signe Redfield 3, Julie A McMurry 4, Peter J Leese 5, Andrew T Girvin 6, Miles Crosskey 7, Andrea G Zhou 8, Richard A Moffitt 9, 10, Michele Jonsson Funk 5, Emily R Pfaff 5, Melissa A Haendel 4, Christopher G Chute 11, N3C; RECOVER Consortia

Data-driven identification of post–acute SARS-CoV-2 infection subphenotypes

Hao Zhang 1, Chengxi Zang 1, Zhenxing Xu 1, Yongkang Zhang 1, Jie Xu 2, Jiang Bian 2, Dmitry Morozyk 1, Drhuv Khullar 1, Yiye Zhang 1, Anna S Nordvig 4, Edward J Schenck 4, Elizabeth A Shenkman 2, Russell L Rothman 5, Jason P Block 6, Kristin Lyman 7, Mark G Weiner 1, Thomas W Carton 7, Fei Wang 8, Rainu Kaushal 1
Racial/ethnic Differences in PASC Prevalence

Millions of people have had COVID-19 — and in many ways, people of color have been hit hardest.

Studies show that some groups and communities are more likely to go to the hospital for health issues related to COVID-19. This is because people don't have equal access to health care and information about COVID. And some people live or work in places where they are more likely to catch COVID-19.

AMERICAN INDIAN OR ALASKA NATIVE
2.5 times more likely to go to the hospital

BLACK OR AFRICAN AMERICAN
2.1 times more likely to go to the hospital

HISPANIC OR LATINO
1.9 times more likely to go to the hospital

ASIAN
0.7

Published online 2023 Feb 16. doi: 10.1007/s11606-022-07997-1

Racial/Ethnic Disparities in Post-acute Sequelae of SARS-CoV-2 Infection in New York: an EHR-Based Cohort Study from the RECOVER Program

Dhruv Khullar, MD, MPP,1,2 Yongkang Zhang, PhD,1 Chengxi Zang, PhD,1 Zhenxing Xu, PhD,1 Fei Wang, PhD,1 Mark G. Weiner, MD,1 Thomas W. Carton, PhD,3 Russell L. Rothman, MD, MPP,4 Jason P. Block, MD, MPH,5 and Rainu Kaushal, MD, MPH1

https://stacks.cdc.gov/view/cdc/105453
Clinical Question

Does there exist **racial/ethnic differences** in potential PASC symptoms and conditions among children and adolescents after **SARS-CoV-2 infection**?
Typical Solutions

• regression model adjusted for confounders

Typical Solutions

Standard regression model

Step 1: large-scale propensity score (LSPS) stratification/matching/weighting

• Fit LSPS model: Race/ethnicity ~ confounders

• Stratify or match or weight on propensity scores

Step 2: Outcome regression model

• Regression model, with propensity score adjusted

LEGEND pipeline

How many differences are attributable to COVID infection?

JOURNAL ARTICLE

Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using hypertension as a case study

Martijn J Schuemie, Patrick B Ryan, Nicole Pratt, RuiJun Chen, Seng Chan You, Harlan M Krumholz, David Madigan, George Hripcsak, Marc A Suchard

How Many Differences are Attributable to COVID Infection?

- Difference-in-differences approach
Proposed Solution

**Standard regression model**
- regression model adjusted for confounders

**LEGEND pipeline**
- **Step 1**: large-scale propensity score (LSPS) stratification/matching/weighting
  - Fit LSPS model: Race/ethnicity ~ confounders
  - Stratify or match or weight on propensity scores
- **Step 2**: Outcome regression model
  - Regression model, with propensity score adjusted

**Proposed method**
- **Step 1**: large-scale propensity score (LSPS) stratification/matching/weighting
  - Fit LSPS model: Race/ethnicity ~ confounders
  - Stratify or match or weight on propensity scores
- **Step 2**: Outcome regression model
  - Difference-in-differences analyses to control pre-COVID racial/ethnic differences
  - Regression model, with propensity score adjusted
Study Cohort

Inclusion Criteria

- Documented SARS-CoV-2 infection
- Age < 21 years
- Had at least one visit during the baseline period
- Had at least one visit during the follow-up period

225,723 patients across 13 institutions in the US
- **Stratify by COVID-19 acute phase severity**

- **Three minority groups compare to Non-Hispanic White group respectively**
Conclusion

Racial/ethnic differences related to COVID-19 vary across:

- racial/ethnic groups
- severity of acute COVID-19
- particular PASC symptom and condition

- Stratify by COVID-19 acute phase
- Three minor groups compare to Non-Hispanic White group respectively

<table>
<thead>
<tr>
<th>Respiratory</th>
<th>Non-severe</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory signs and symptoms</td>
<td>At least one groups differential risk difference &gt; 0</td>
<td>At least one groups differential risk difference &lt; 0</td>
</tr>
<tr>
<td>Skin symptoms</td>
<td>Difference of risk difference</td>
<td>Difference of risk difference</td>
</tr>
</tbody>
</table>

- AAPI vs. NHW
- Hispanic vs. NHW
- NHB vs. NHW
Takeaways

- Help understand racial/ethnic differences in PASC after SARS-CoV-2 infection among children and adolescents

- Cover a broad spectrum of the US pediatric population
  
  **LEGEND principle 1:** Generate evidence at a large scale  
  **LEGEND principle 9:** Generate evidence across a network of multiple databases

- Handle measured confounders using propensity score matching

- Control pre-COVID racial/ethnic differences using difference-in-differences analyses
  
  **LEGEND principle 5:** Generate evidence using best practices to minimize bias

- Future work
  
  - Explore methods to adjust for systematic bias
Acknowledgment

Research Team

- Dazheng Zhang^, Bingyu Zhang^, Qiong Wu, PhD, Ting Zhou, MD, Jiayi Tong, Yiwen Lu, Jiajie Chen, PhD, Deena J. Chisolm, PhD, Ravi Jhaveri, MD, Rachel C Kenney, PhD, Russell L Rothman, MD, MPP, Suchitra Rao, MD, David A. Williams, MD, Mady Hornig, MA, MD, Jeffrey S. Morris, PhD*, Christopher B. Forrest, MD, PhD*, and Yong Chen, PhD*

^ co-first author
* senior author
Eye Care and Vision Research Workgroup
Our Journey

GETTING STARTED

INITIAL STEPS

MILESTONES

NEXT STEPS
Getting Started

- OHDSI Eye Care and Vision Research Workgroup was started in spring 2022
  - Members of American Academy of Ophthalmology (AAO) Data Standards Workgroup identified need for ophthalmic data elements in the OMOP common data model
  - Ophthalmic concepts in source terminologies had not been updated consistently in over a decade

- Goals
  - Create access to large diverse datasets of ophthalmic and systemic data
  - Enable research in vision and systemic health
Initial Steps

- Created subgroups for tasks & subspecialties
  - Tasks: Concept mapping, visual acuity concept mapping, visual impairment phenotype, image integration, ETL scripts
  - Subspecialties: Glaucoma, retina, pediatrics/strabismus, uveitis
- Recruited colleagues to participate
- National Eye Institute (NEI) at National Institutes of Health (NIH) hired DATA Scholar to manage the project
Milestones

- **Membership**
  - 122 total, ~40 active
  - 13 trainees, 10 AI-READI (Bridge2AI) interns
  - Ophthalmologists, optometrists, informaticists, vision scientists

- **Meetings**
  - 17 Teams workgroup meetings
  - 3 in person meetings
  - ~42 subgroup meetings
  - Countless ad-hoc meetings
Milestones

- Collaborations
  - 9 OHDSI workgroups
  - 10 external groups including:
    - American Academy of Ophthalmology (AAO)
    - Association for Research in Vision and Ophthalmology (ARVO)
    - National Eye Institute
    - NIH Bridge2AI
    - NIH All of Us
    - SNOMED International and LOINC
Milestones

- **Data Concepts**
  - >3700 ophthalmic data elements analyzed & mapped
  - 11 retina condition codes submitted to SNOMED International
  - 224 visual acuity concepts submitted to LOINC
  - Glaucoma concepts currently in discussion with SNOMED International
Epic EHR Concept Matches

Milestones

- Phenotypes
  - 3 visual impairment
  - 6 uveitis*
  - 3 new anti-VEGF users*
  - 1 blinding disease*
  - 5 diabetic retinopathy

*Submitted to How Often
Milestones

- Publications
  - 9 papers, 4 EyeWiki pages
  - 5 more in progress

- Presentations
  - 18 talks, 5 posters

- Support
  - 1 NEI/NIH Data Scholar
  - 2 Grant submissions
Milestones

- **SOS Challenge 2023**
  - Led by Cindy X. Cai MD MS from Johns Hopkins University
  - Comparison of 3 anti-VEGF agents for risk of kidney injury when injected intravitreally
  - Results: no increased risk for kidney injury in any pairwise comparisons
  - Manuscript is in process
Next Steps

- Pilot at test sites
  - Image integration
  - Concept mappings (prioritized set)
- More eye care and vision research community outreach and education
- More network studies
- More funding support
First Year Challenges

Schedules
- Clinic schedules
- Time zones

Concept Modifiers
- Measurements often have multiple modifiers
- Pre-coordination results in thousands of concepts

Resources
- All volunteer effort
- DATA Scholar position is only 2 years

Diversity
- Members are from academic medical centers
- Need more diverse partners
Summary

- Eye Care and Vision Research Workgroup had a productive year
- Working towards goal of including ophthalmic data and imaging in the OMOP common data model
- Still much more work to do—come join us!

Workgroup meeting is Sunday, Oct. 22 at 1 – 5 pm.
Thank you!

- **OHDSI Community**
  - Clair Blacketer, Paul Nagy, Elisse Katzman, Nathan Hall, Patrick Ryan, Craig Sachson, Anna Ostropolets
  - SOS Challenge collaborators

- **Eye Care and Vision Research Workgroup**
  - Co-leads: Kerry Goetz, Sally Baxter
  - Subgroup leads: Cindy Cai, Gayathri Srinivasan, Brian Stagg, Kavi Thakoor, Brian Toy
  - All of our wonderful members!

- **National Eye Institute and National Institutes of Health**