INNOPOLIS UNIVERSITY

Same Top Fraction, Different Workload

Auditing Eligibility and Capacity

in Clinical Risk Evaluation

Vache Oganisyan (1) Dmitry Lvov (1, 2) Ilya Pershin (1, 3)
1 Innopolis University 2 ITMO University 3 Kazan Federal University
TAE (Trust-AI-Eval), NeurIPS 2026 d.lvov@innopolis.ru

The workload behind the same top 10%

Preterm birth, CDC 2023



+33.8% selected records for the same top 10%

All-record scoring is a retrospective upper bound, including no-care and unknown timing

Study design

2022

Fit and tune
80/20 train/validation

2023

Develop and evaluate audit
~3.48 million records

2024

Locked temporal replication
~3.51 million records

Preterm birth / NICU admission / Low birth weight

One frozen LightGBM per outcome. The same scores in all comparisons

U.S.-resident singleton live births. Nine maternal birth-certificate variables

Eligibility and selection budget

Capture = selected events / all population events

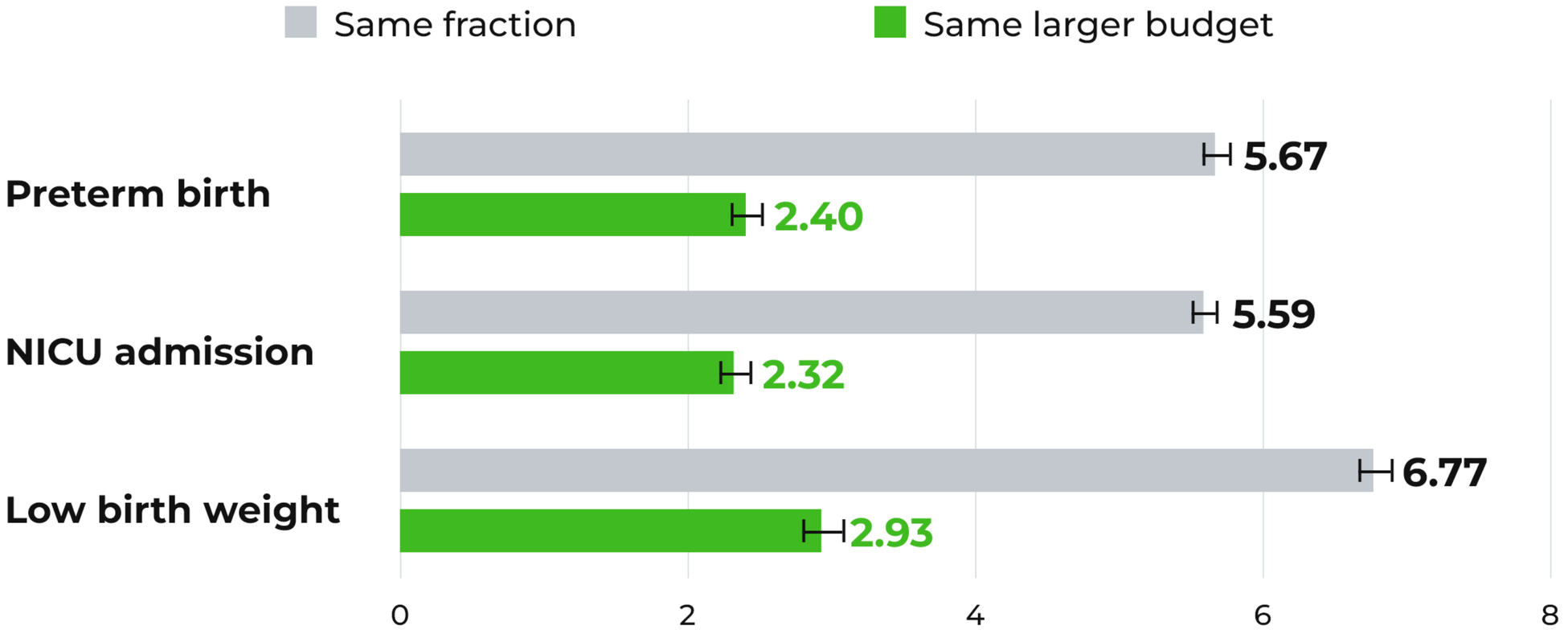
Candidates / selections	260,106	348,038
Early entry	**C00 14.63%**	**C01 17.89%**
All records	**C10 16.52%**	**C11 20.30%**

Same fraction: C11 - C00 = +5.67 pp

Same 348,038 selections: C11 - C01 = +2.40 pp

Preterm birth, CDC 2023. pp = percentage points. Differences use unrounded values

The equal-budget comparison



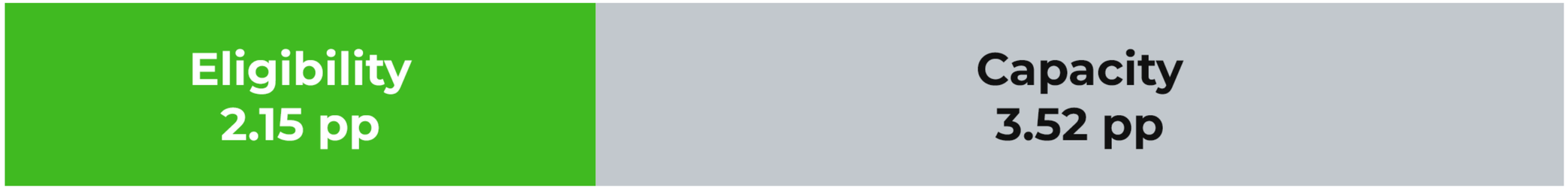
57-58% smaller gain at the larger fixed budget

CDC 2023, q = 10%. Gain in pp. 95% conditional bootstrap intervals

Separating eligibility from capacity

Preterm birth, CDC 2023, q = 10%

Symmetric Shapley allocation



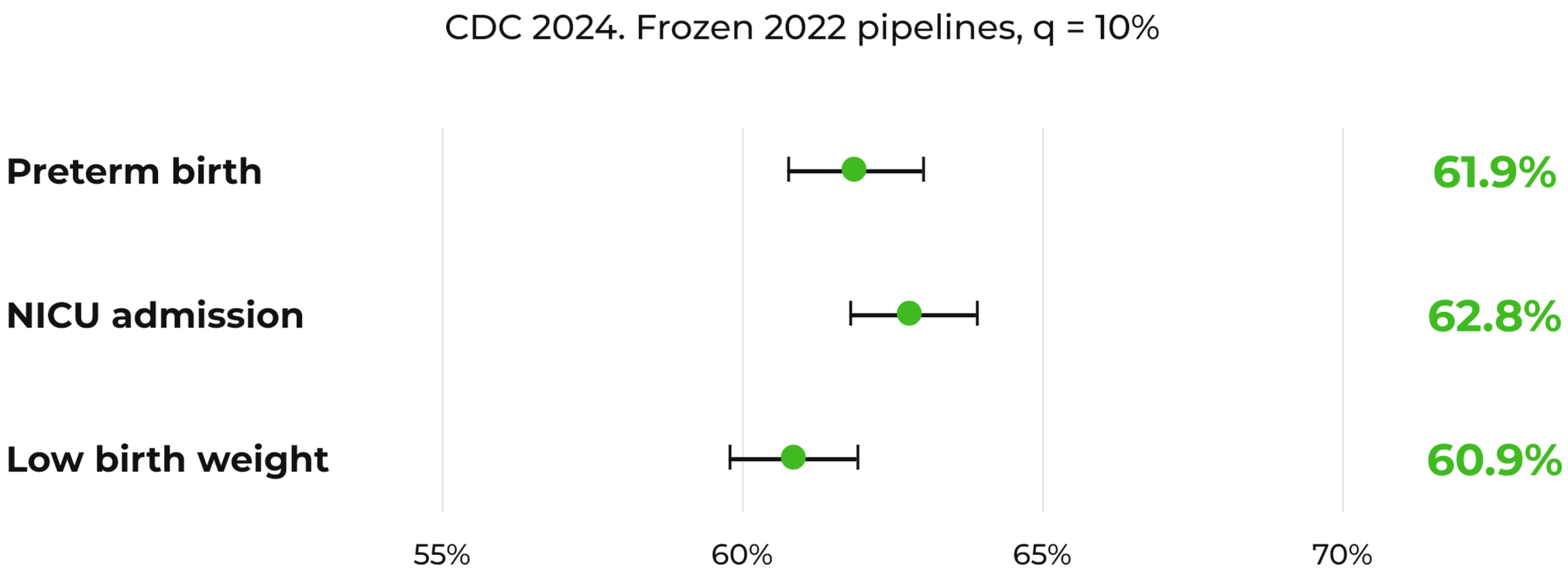
Total capture gain: 5.67 percentage points

Eligibility: (1.89 + 2.40) / 2 = 2.15 pp

Capacity accounts for 61.6-62.2% across the three outcomes

Symmetric Shapley allocation averages each factor's increment over both orders

Locked temporal replication



Capacity share remains 60.9-62.8% in the locked replication

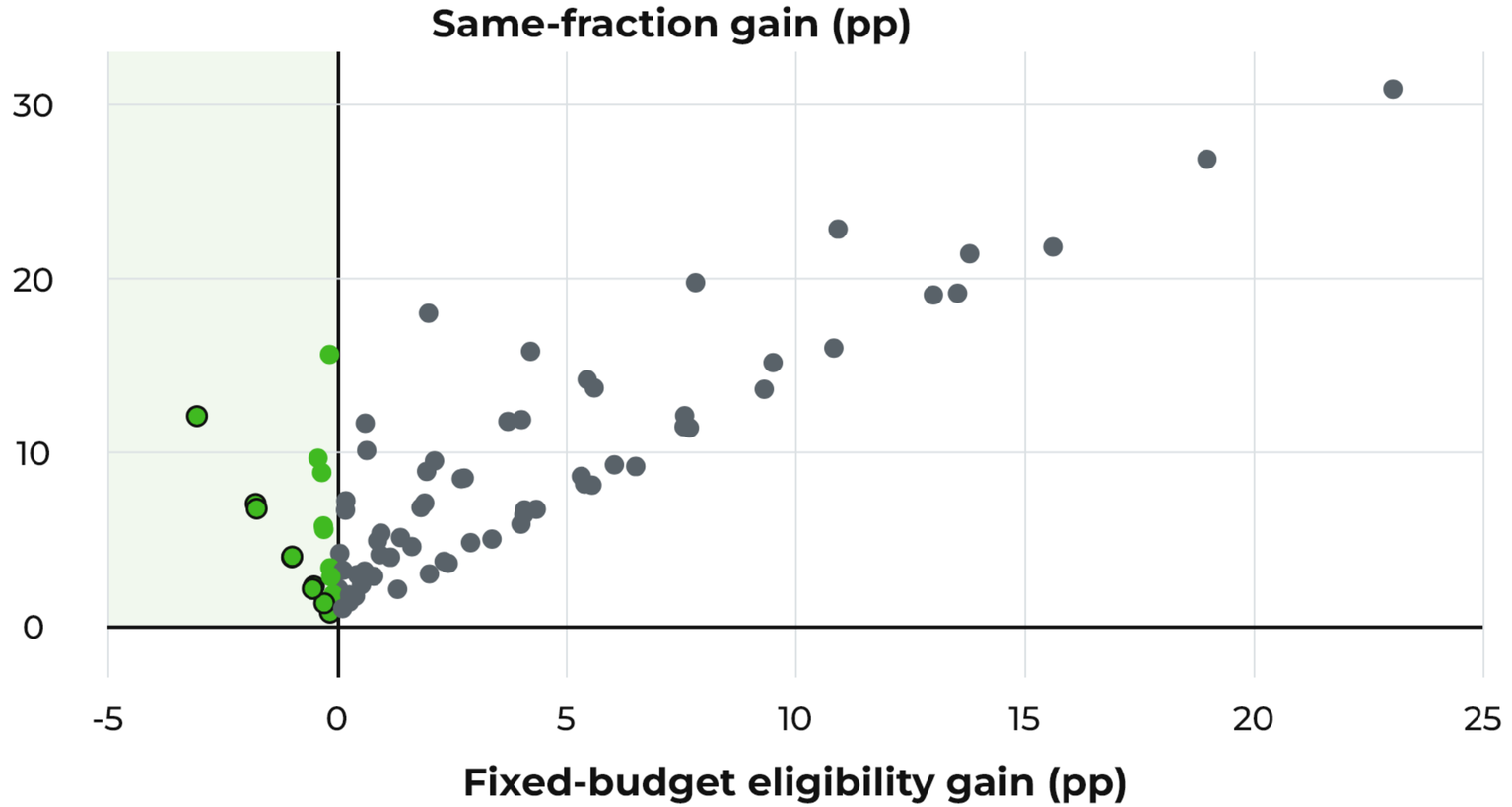
95% conditional bootstrap intervals. Scores and absolute budgets fixed

Controlled mechanism checks

18 / 81

sign reversals

Simulation: one point = one condition



The evaluation policy can reverse the conclusion

81 conditions, 100 repetitions each. Outlined: 9 central 95% replicate ranges below zero

Evaluation at a fixed review capacity

Compare the same absolute number of selections

For services with fixed review capacity

Target population / Eligibility rule
Selected count / Population-event denominator

Paper: openreview.net/forum?id=puYbvC47rw
Code: github.com/vache4ogan/prenatal-risk-scoreability

Paper

Code

d.lvov@innopolis.ru

Supported by the Ministry of Economic Development of the Russian Federation (agreement No. 139-10-2025-034 dd. 19.06.2025, IGK 000000C313925P4D0002).