Public



Task Aligned Association Preservation and Downstream Utility of Synthetic Clinical Trial Adverse Event Data

When can synthetic AE data support a real prediction task?

Meishuo Ouyang

Public

Contents

1.When can synthetic AE data support a real prediction task?

2.The controlled disruption experiment

3.What the two trials showed

4.Implications for clinical trial Teams

Public

1.The Utility Gap in Synthetic Trial Data

A synthetic dataset may reproduce the frequency of each variable while weakening the relationships that a model needs.

Marginal similarity

The observed frequencies of individual fields look similar.

Task utility

A model trained on synthetic records performs on separate real events.

The study tested whether these two properties can come apart.

Public

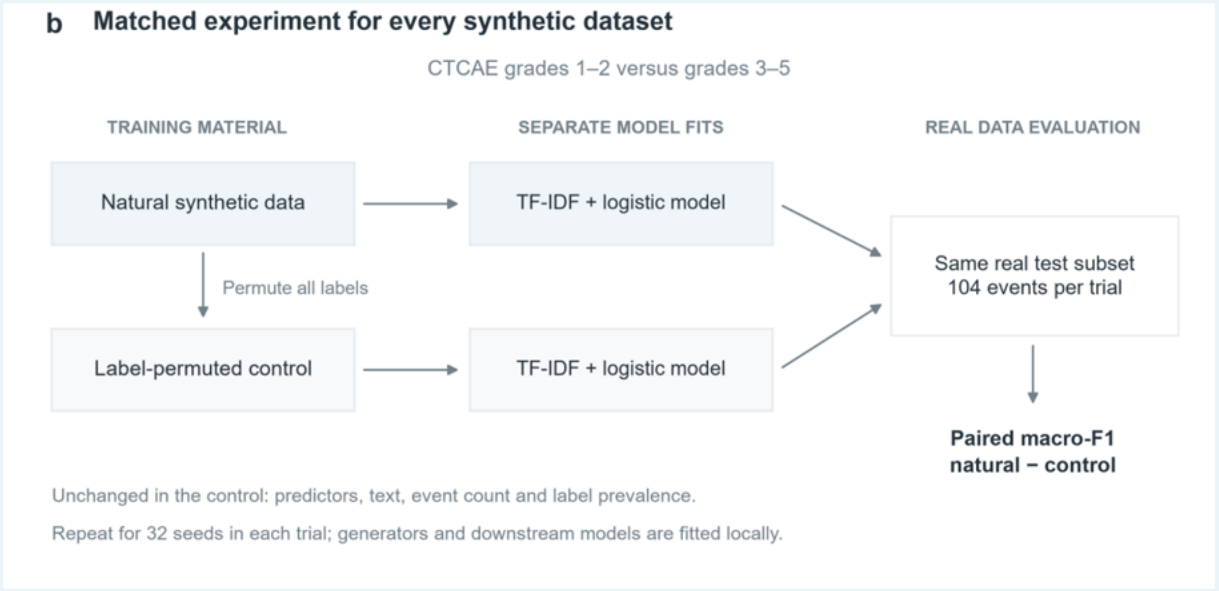
Two Questions in This Study

- Does disrupting predictor–outcome alignment reduce downstream performance?
- Can an association-preservation score identify the better natural synthetic datasets?

Public

Study Overview





Public

What Changed in the Control

Preserved
Predictor values
Rendered model text
Number of AE events
CTCAE class prevalence

Disrupted
Alignment between each record's predictors and its AE-severity label

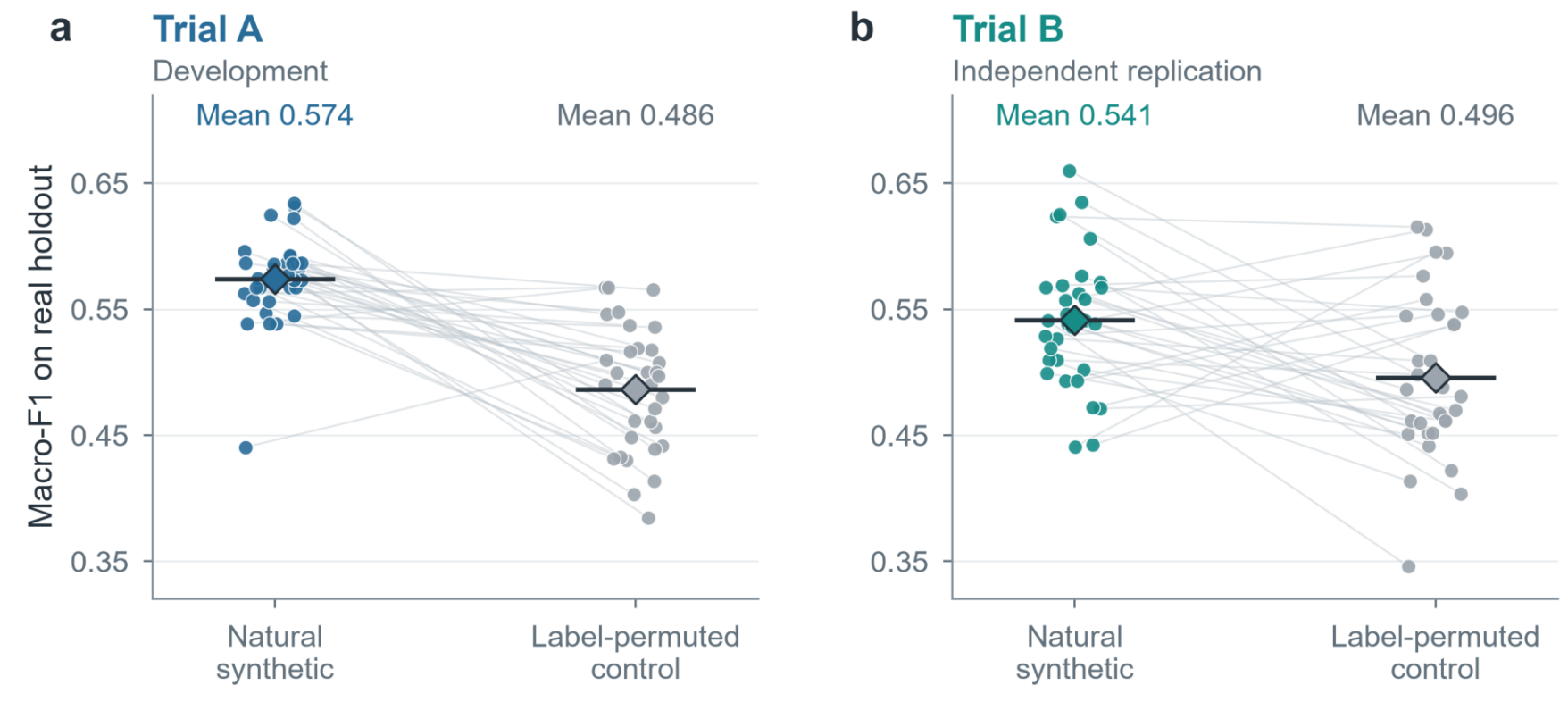
Train on Synthetic, Test on Real

Synthetic training material
A separate TF-IDF/logistic model was fitted for each natural dataset and its matched control.

Real evaluation set
Both models were evaluated on the same 104 real AE events in the relevant trial.

Public

Study Results: Real-Data Performance



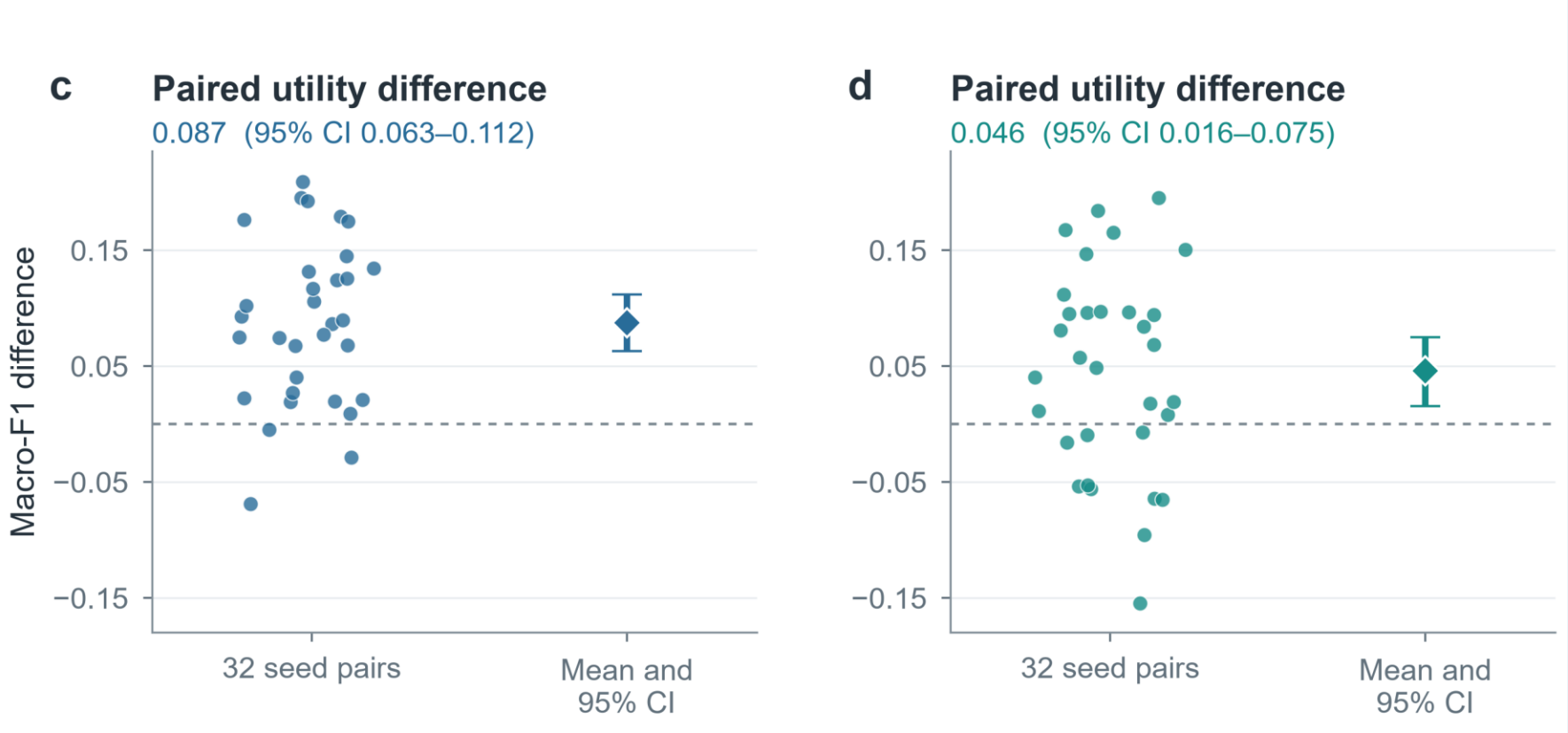
Trial A
0.574 vs 0.486

Trial B
0.541 vs 0.496

Each line connects a natural synthetic dataset with its matched control.

Public

Study Results: Paired Utility Difference



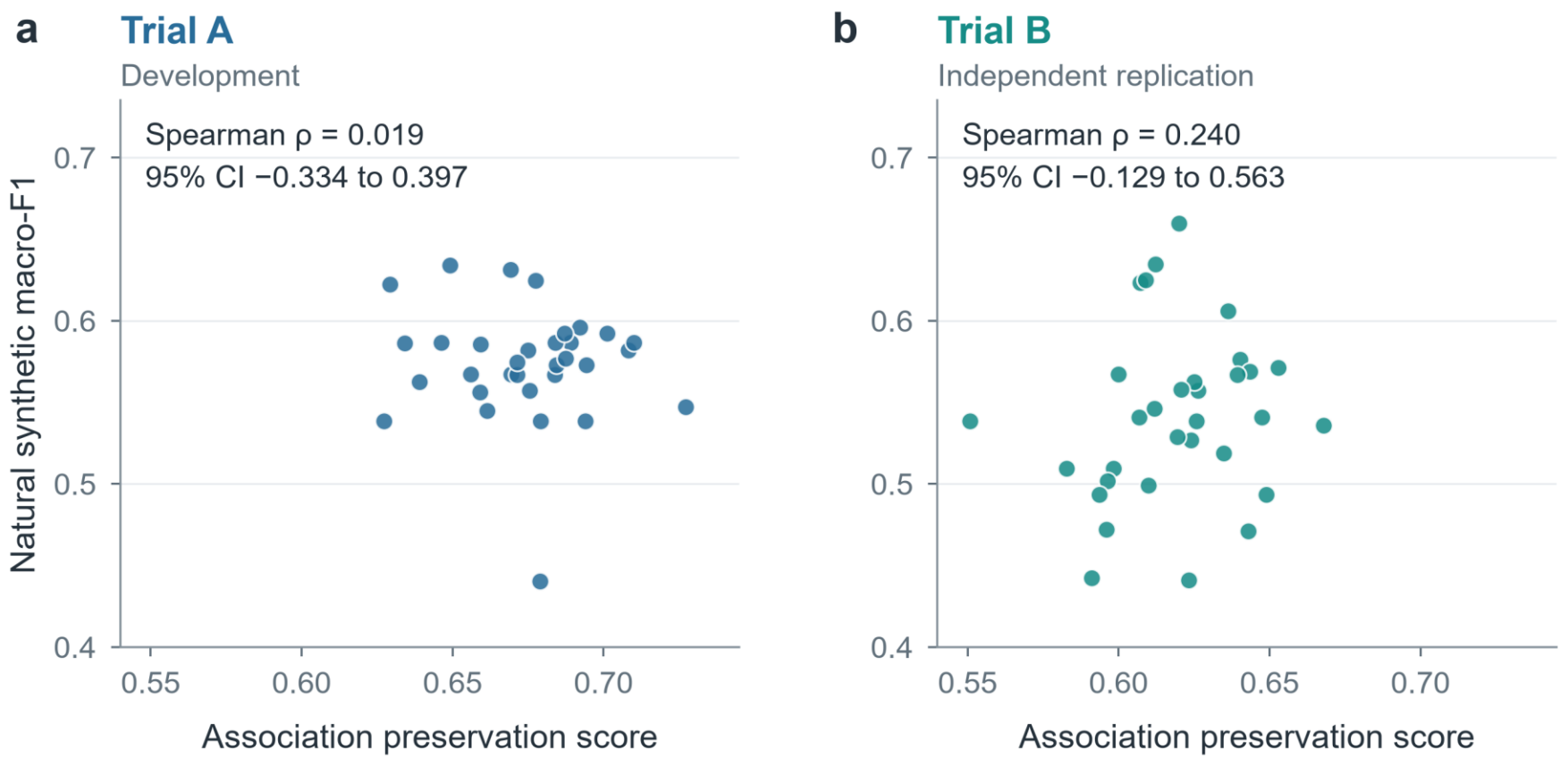
Trial A
+0.087
95% CI 0.063–0.112

Trial B
+0.046
95% CI 0.016–0.075

The contrast was directionally replicated in the second trial.

Public

Study Results:Association Score and Natural-Data Utility



Trial A
ρ = 0.019

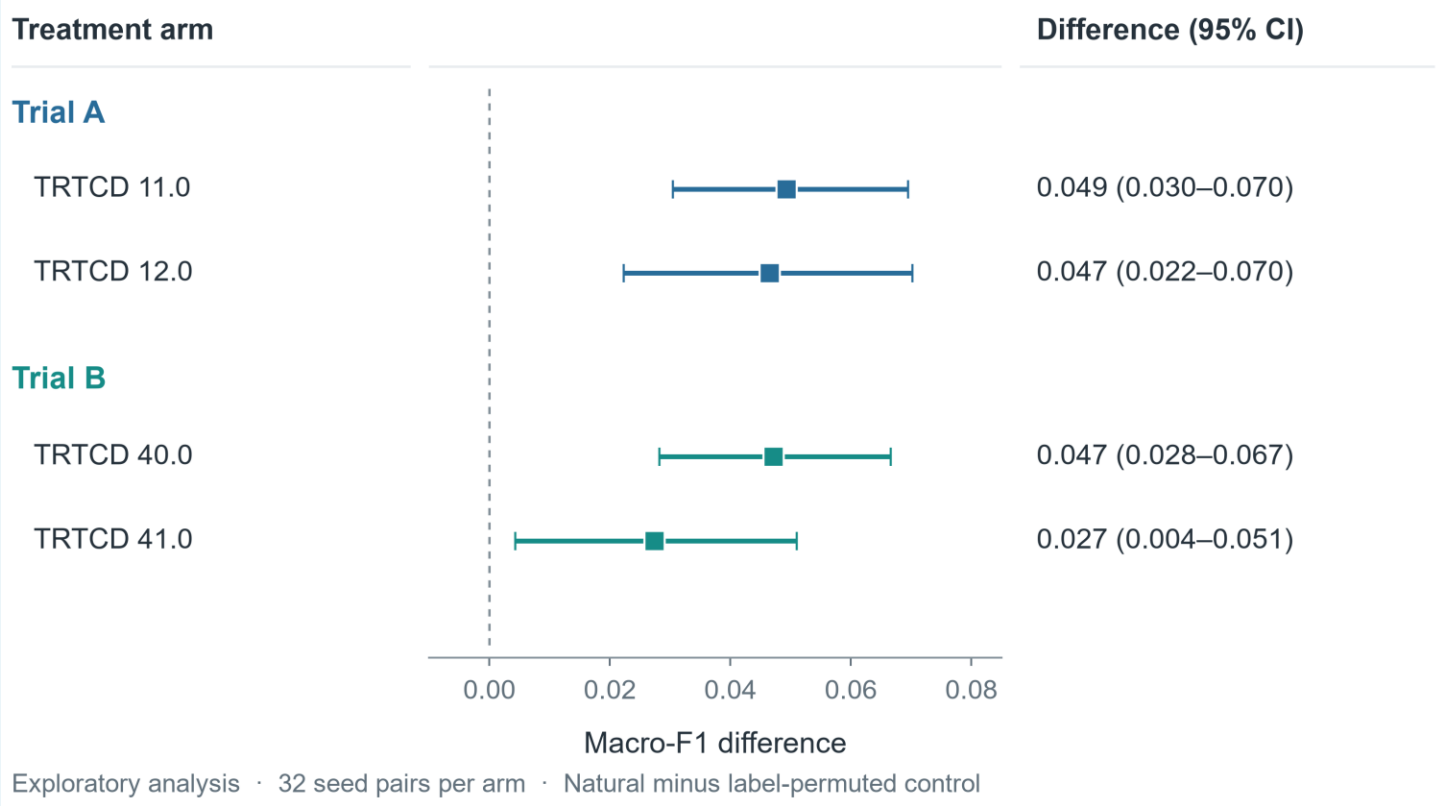
Trial B
ρ = 0.240

Both 95% intervals include zero.

A score that detects complete disruption did not reliably identify the best natural candidates.

Public

Study Results:Exploratory Treatment–Arm Contrasts



All four arm-level differences were positive.

These analyses do not compare treatment effects.

Within-arm consistency is supporting evidence, not proof of an arm-specific effect.

Public

Conclusion and Future

Disrupting predictor–outcome alignment reduced downstream AE-severity classification performance in both trials.
The association score did not reliably rank natural synthetic datasets. Direct, task-specific testing on real holdout data remains necessary when predictive utility matters.

Next:
evaluate other generators, endpoints and trials, with privacy and subgroup fairness assessed independently.

Public

Implications for Clinical Trial Teams

Data management	Preserve field definitions, provenance and event timing.
Medical and safety	Check whether clinically relevant AE relationships remain plausible.
Statistics and AI	Specify the task, control leakage and test on independent real data.
Governance	Assess disclosure risk and fairness separately from utility.

Public

Introduction of Presenter

Meishuo Ouyang, Ph.D

- Clinical data scientist and strategy leader specializing in clinical development, biostatistics, and artificial intelligence.
- Vice President of Strategy & Market at Q-Square Business Intelligence, with previous postdoctoral research experience at Duke University School of Medicine.
- Research interests include AI-enabled clinical development, synthetic clinical data, statistical validation, and privacy-preserving analytics.
- Experience spanning randomized clinical trials, real-world data, health economics, causal inference, and machine learning.
- Current work focuses on integrating synthetic data, AI agents, and cryptographic verification into reliable and scalable clinical research workflows.