Externally Controlled Trials

A Review of Design and Borrowing Through a Causal Lens

On behalf of the ASA Biopharmaceutical Section RWE Scientific Working Group — HCT Subgroup

Shu Yang · Professor of Statistics, North Carolina State University

ASA Biopharmaceutical Section Webinar · September 11, 2026 · 11:00 AM EST

Review: Zhu et al. (2026), arXiv:2605.03282

ROADMAP

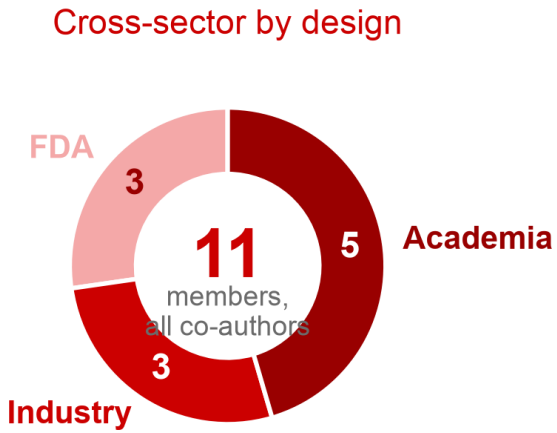
This 15-minute overview

1. **The HCT subgroup**
   Motivation, members, and activities of the RWE SWG's subgroup on externally/hybrid controlled trials

2. **The review paper**
   What we wrote, for whom, and the three contributions

3. **Today's webinar**
   How the next 90 minutes are organized

4. **Review overview**
   Designs, causal roadmap, bias objects, and method selection



Why the RWE Working Group formed an HCT subgroup

The scientific problem

RCTs are the gold standard but can be infeasible, unethical, or too slow — rare disease, oncology, pediatrics, post-approval

External controls (historical trials, registries, natural history, RWD) are increasingly proposed as comparators or augmentation

Hybrid controlled trials — RCTs augmented with external controls — are where methodological activity is concentrated, yet regulatory guidance is thinnest

The community gap

Methods literature fragmented by disease area and by Bayesian vs frequentist paradigm

Sponsors and reviewers lack a shared basis for comparing, selecting, and evaluating borrowing methods

Regulatory momentum (FDA external-control guidance, EMA/MHRA, ICH M14, Bayesian draft guidance) raises the stakes for getting this right

A multi-stakeholder ASA subgroup is a natural home for this synthesis

Subgroup charge: bring industry, regulatory, and academic statisticians together on ECT/HCT design, analysis, and evaluation.

Subgroup members

Regulatory (U.S. FDA)

Pallavi Mishra-Kalyani — FDA

Hana Lee — FDA

Lei Nie — FDA

Industry

Rima Izem — Novartis, Statistical Methodology

Herbert Pang — Genentech, PD Data Science & Analytics

Birol Emir — Pfizer R&D, Data Sciences & Analytics

Academia

Mark van der Laan — UC Berkeley (causal inference / TMLE)

Ying Yuan — MD Anderson (Bayesian designs)

Peng Yang — MD Anderson

Shu Yang — NC State (data integration; speaker)

Ke Zhu — NC State (lead author)

All eleven are co-authors of the subgroup's review paper — a deliberately cross-sector team.

A year + of subgroup activities



Regular cross-sector meetings on ECT/HCT design, analysis, and evaluation challenges

Separated the two design settings (SAT+EC vs HCT); identified covariate shift vs outcome drift as the organizing bias objects

Mapped Bayesian and frequentist methods, software, and regulatory guidance onto one framework

Co-authored Zhu et al. (2026) — the subgroup's main written deliverable

This webinar: overview + multi-stakeholder panel + open Q&A for the broader Section

One-year arc: from open questions to a shared framework to public dissemination.

REVIEW PAPER

The subgroup's review paper

"Externally Controlled Trials: A Review of Design and Borrowing Through a Causal Lens"

Zhu, Izem, Yang, Yuan, Pang, van der Laan, Nie, Emir, Mishra-Kalyani, Lee, Yang (2026)

arXiv:2605.03282 · under review

Decision framework

Not a methods catalog — written for industry, regulators, and methodologists

Two designs

SAT+EC and HCT on the same causal spine

Bias-aware taxonomy

Bayesian and frequentist borrowing compared by covariate shift vs outcome drift

Practice-ready

Method-selection table, 19 R packages, evidence-package guidance



Externally Controlled Trials: A Review of Design and Borrowing Through a Causal Lens

Ke Zhu[1,2], Rima Izem[3], Peng Yang[4], Ying Yuan[4], Herbert Pang[5,2], Mark van der Laan[6], Lei Nie[7], Birol Emir[8], Pallavi Mishra-Kalyani[7], Hana Lee[7], and Shu Yang*[1]

[1]Department of Statistics, North Carolina State University, Raleigh, NC, USA
[2]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, USA
[3]Statistical Methodology, Novartis Pharma AG, Basel, Switzerland
[4]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
[5]PD Data Science & Analytics, Genentech, South San Francisco, CA, USA
[6]Division of Biostatistics, University of California, Berkeley, CA, USA
[7]Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA†
[8]Pfizer Research & Development, Chief Medical Office, Data Sciences and Analytics, New York, NY, USA

Abstract

Externally controlled trials (ECTs) are increasingly used when randomized controls are infeasible, unethical, or insufficient, including applications in rare diseases, oncology, pediatrics, and post-approval effectiveness research. Although methodological work has expanded rapidly across causal inference, Bayesian dynamic borrowing, and hybrid trial designs, the literature remains fragmented. We adopt a six-step scientific roadmap to organize modern ECT methodology in two primary settings: (i) single-arm trials that evaluate efficacy through comparison with external controls, and (ii) hybrid controlled trials that augment the internal control arm with external controls drawn from real-world data or historical studies. The roadmap clarifies causal estimands, identifiability assumptions, and how statistical parameters arise from identification, and shows how modeling and borrowing strategies trade off efficiency and robustness, especially under covariate shift and outcome drift. Within this framework, we synthesize and evaluate recent Bayesian and frequentist developments, compare their strengths, limitations, operating characteristics, and available software, and emphasize the role of sensitivity analysis. By re-framing ECT methodology through a causal lens, this work establishes a coherent foundation for integrating external data into regulatory and clinical decision-making and highlights core challenges and opportunities for future research.

Keywords: Bayesian dynamic borrowing; causal inference; hybrid trial designs; real-world evidence; sensitivity analysis.

Manuscript first page

WEBINAR

How today's 90 minutes are organized

Overview · 15 min | Moderated panel · 60 min | Q&A · 15 min

11:00 11:15 12:15 12:30

Overview talk

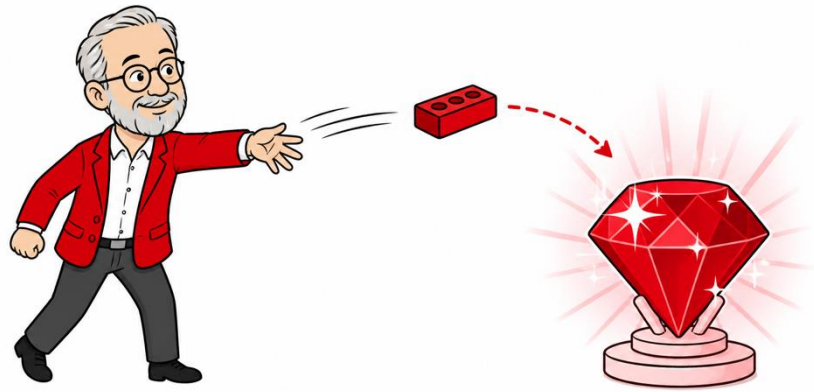
Subgroup context and the main messages of the review (this talk)

Moderated panel

Regulatory, industry, and academic discussants

Audience Q&A

Questions monitored, answered, or routed by Ke Zhu and Peng Yang



Throw a brick to attract a gem

This overview is the brick; the panel is the gem — and it continues the subgroup's multi-stakeholder format.

Two ECT designs — the key split is randomization

A. Single-arm trial + external control (SAT+EC)

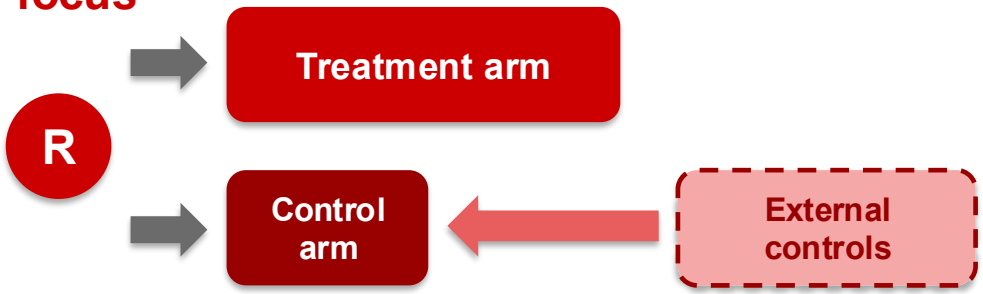
Single arm: all patients treated ⟷ **External controls (trials, registries, RWD)**

Non-randomized comparison — EC identifies the effect

Untestable no-unmeasured-confounding assumptions; confounding is the main threat

Enables evidence when randomization is infeasible or unethical (rare disease, oncology)

B. Hybrid controlled trial (HCT) — subgroup focus



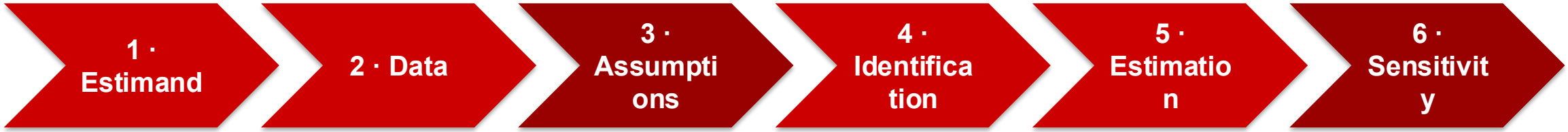
ECs augment the randomized control arm for efficiency

RCT analysis remains the validity benchmark

Internal controls help diagnose external-data bias; guidance thinner here — the gap the review targets

Same causal roadmap applies to both; what changes is the role of the external data.

The six-step causal roadmap



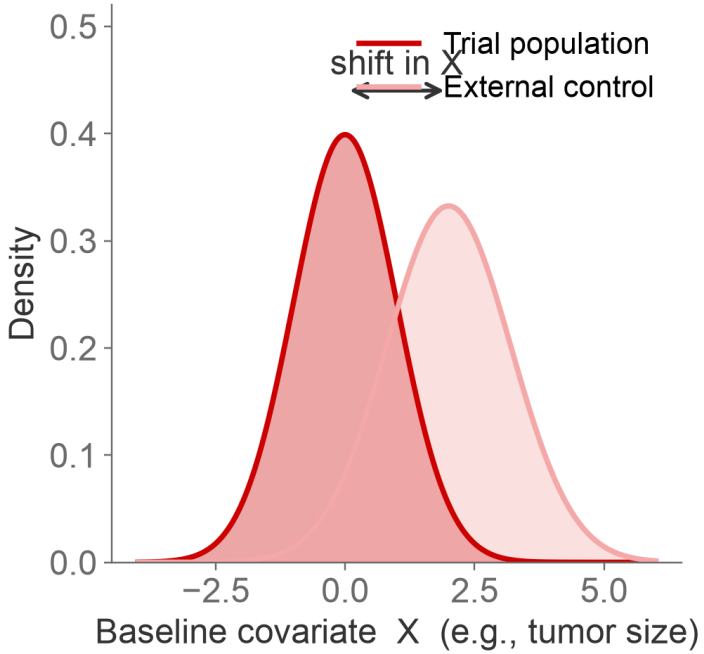
1 · Estimand	2 · Data	3 · Assumptions	4 · Identification	5 · Estimation	6 · Sensitivity
Scientific question: population, treatments, outcome, intercurrent events (ICH E9(R1))	Fit-for-purpose subject-level data; eligibility-aligned EC cohort	What must hold for the estimand to be recoverable — often untestable	Estimand written as a functional of the observed-data law	Borrowing strategy, inference, operating characteristics, software	Probe unverifiable assumptions; tipping points, E-values

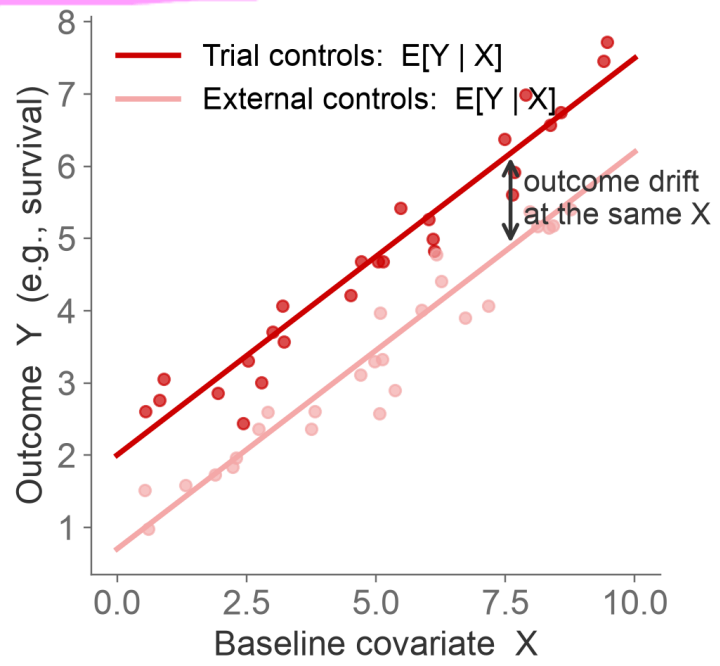
Not just organization — a decision tool

The roadmap decides which assumptions are design-critical, which can be probed with internal controls, and which are sensitivity-only.

Two bias objects drive method choice

A. Covariate shift — visible in the data B. Outcome drift — persists after adjusting for X





Covariate shift

Visible baseline differences in X

Fixed by alignment, weighting, outcome models, DR methods

Outcome drift

Residual gap in Y at the same X

Unmeasured confounding, concurrency, measurement, care patterns

Selection principle: fixing only covariate shift buys efficiency when drift is absent — and bias when it is not. Choose for the bias you cannot rule out.

HCT borrowing methods on one map

I — **RCT-only (No borrowing)**

No EC assumptions — the benchmark (covariate-adjusted, AIPW/TMLE)

II — **Borrow assuming no drift (Full borrowing)**

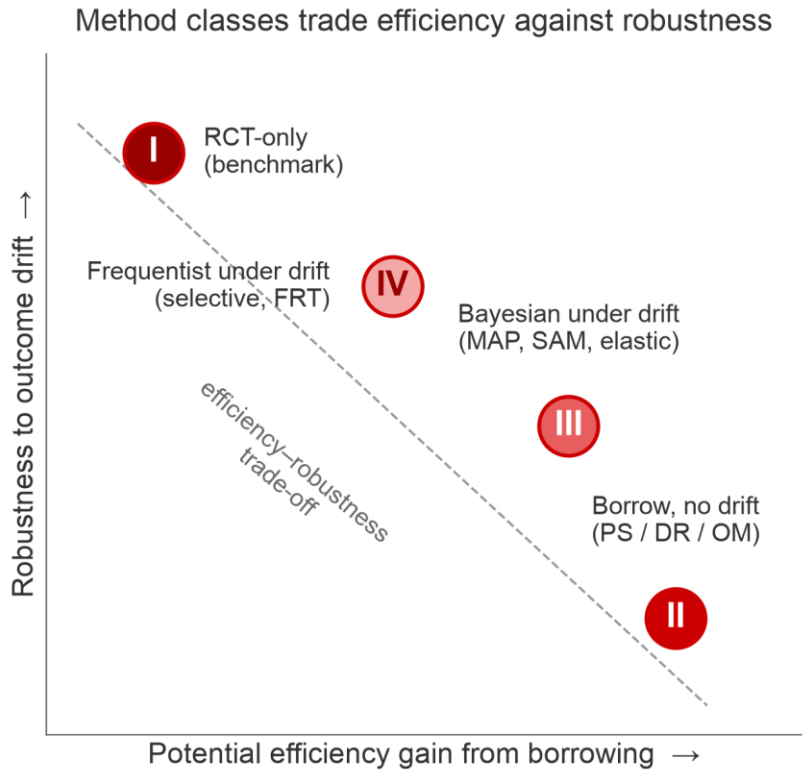
PS matching/weighting, outcome models, doubly robust — largest gain, highest bias risk

III — **Bayesian under drift (Dynamic borrowing)**

Power / commensurate / elastic priors, (r)MAP, SAM, hierarchical models, selective priors

IV — **Frequentist under drift (Selective borrowing)**

Prognostic adjustment, test-then-pool, selective borrowing, bias models, randomization inference



Grouped by assumptions addressed — not Bayesian vs frequentist first. Full comparison in Table 5 of the review.

From methods to practice

— SAT+EC: eligibility alignment, fit-for-purpose data, and sensitivity analysis are the load-bearing beams — sensitivity is not optional

— HCT: keep the RCT-only benchmark; treat borrowing as earned efficiency, not free sample size

— Pre-specify the borrowing rule, success criterion, and sensitivity plan in the SAP

— Show operating characteristics by simulation across drift scenarios — not only under exchangeability

— Quantitative sensitivity: tipping points e.g. — how much bias would overturn the claim?

— Software: 19 R packages catalogued in the review (RBesT, SAMprior, psborrow2, intFRT, SelectiveIntegrative, …)

Controversies addressed & where the field goes

Three recurring controversies (our take)

Type I error under borrowing? Controllable with prespecified, calibrated procedures

Power gains with error control? Yes, in realistic semiparametric settings

Dynamic borrowing illusory in small samples?

Detection is limited, but large conflicts can still be blunted

Open directions

Master protocols and platform/basket borrowing

Synthetic / AI-generated controls

Bayes–frequentist synthesis and calibration

Design-stage borrowing and sample-size planning

Decision-grade uncertainty communication for regulators

CLOSING

Takeaways — and over to the panel

— The RWE SWG HCT subgroup treated externally controlled trials as a strategic, cross-sector topic — the review is its written deliverable

— One six-step causal roadmap applies to both ECT designs; randomization changes the external data's role

— Select methods by covariate shift vs outcome drift, then by your power-bias tradeoff

— Fit-for-purpose external data come first; adaptive borrowing is not a substitute for compatibility

— Panel next: causal and Bayesian foundations, industry implementation, and regulatory discussions

Mark van der Laan

Jiann-Ping Hsu / Karl E. Peace Professor of Biostatistics and Statistics · University of California, Berkeley

Biography

Mark J. van der Laan, Ph.D., is the Jiann-Ping Hsu/Karl E. Peace Professor of Biostatistics and Statistics at the University of California, Berkeley. He developed targeted maximum likelihood estimation (TMLE) and super learning, and his research includes causal inference, semiparametric models, and censored data. He received the COPSS Presidents' Award in 2005 and is a founding editor of the Journal of Causal Inference and the International Journal of Biostatistics. He is speaking today on the causal foundations of externally controlled trials.

Questions for Mark van der Laan

— The roadmap places identification before estimation. In your view, what do practitioners most often get wrong when they skip from data to estimator without stating the identifying assumptions?

— Doubly robust estimation and TMLE are central to the frequentist borrowing toolkit. What are the practical advantages of targeted learning in small-sample hybrid trials, and what finite-sample cautions matter — nuisance estimation, undersmoothing, cross-fitting?

— Adaptive TMLE has been proposed for deciding how much to borrow. How do you see the frontier between data-adaptive borrowing and honest post-selection inference developing?

PANEL · DISCUSSANT 2 OF 7 · ACADEMIA · BAYESIAN

Ying Yuan

Bettyann Asche Murray Distinguished Professor and Chair · Department of Biostatistics · MD Anderson Cancer Center

Biography

Ying Yuan is the Bettyann Asche Murray Distinguished Professor and Chair of the Department of Biostatistics at the University of Texas MD Anderson Cancer Center. Dr. Yuan is internationally renowned for his pioneering research in innovative Bayesian adaptive designs. The designs and software developed by Dr. Yuan's lab (www.trialdesign.org) have been widely adopted by medical research institutes and pharmaceutical companies. Among these, the BOIN design, developed by Dr. Yuan's team, is a groundbreaking oncology dose-finding method recognized by the FDA as a fit-for-purpose drug development tool. The BOP2 design is widely used for Bayesian efficacy and toxicity monitoring. Dr. Yuan serves on the FDA Advisory Committee for Genetic Metabolic Diseases.

Questions for Ying Yuan

— A wide range of Bayesian borrowing methods are available, for example, power, commensurate, elastic, robust MAP, SAM priors. When selecting among these approaches, what are the two or three questions that should drive the choice?

— There is often confusion regarding whether information borrowing approaches should be considered “static” or “dynamic.” For example, should a heavy-tailed mixture prior be viewed as a static robust prior or a dynamic borrowing method? Can you unpack that distinction and its practical consequences for trial design?

PANEL · DISCUSSANT 3 OF 7 · INDUSTRY

Rima Izem

Statistical Methodology Director, Advanced Quantitative Sciences · Real-World Evidence Science Lead · Novartis Pharma AG

Biography

Rima Izem, Ph.D., is Director in Biostatistics and Real-World Evidence Science Lead in Advanced Quantitative Sciences, Medical Affairs & Access, at Novartis in Basel. She has nearly 20 years of experience in pharmaceutical research, including ten years at the U.S. FDA and earlier academic appointments. Her work has supported benefit–risk evaluation across the product lifecycle. Recent interests include causal inference and hybrid designs that combine clinical-trial and real-world data, especially in rare events and rare diseases.

Questions for Rima Izem

— In practice, what factors impact the decision to conduct (or not) a hybrid design, and the type of design and statistical methods used?

— How do you evaluate and communicate the added value and the risk of bias (with external data to the trial) for each design and method?

PANEL · DISCUSSANT 4 OF 7 · INDUSTRY

Herbert Pang

PD Data Science & Analytics Hybrid Trials and ECs initiative lead · Genentech

Biography

Herb obtained his PhD in Biostatistics from Yale, and BA in Mathematics and Computer Science from Oxford. He is currently the PD Data Science & Analytics Hybrid Trials and ECs initiative lead at Genentech. He is a member of the Data Innovation group within Real World and Clinical Data Strategy. He is currently serving as the contact (lead) PI on an FDA BAA award on Hybrid RCTs and has previously served as the contact (lead) PI on an FDA U01 grant on Hybrid RCTs. Herb has authored over 160 published articles in statistics, bioinformatics, biomedical and clinical journals; and has contributed to the development of the ‘psborrow’ and ‘rdborrow’ R packages.

PANEL · MODERATED QUESTIONS · INDUSTRY

Questions for Herbert Pang

— The review catalogues 19 R packages for hybrid designs. From an industry implementation standpoint, how do you evaluate and select software for a regulated analysis — validation burden, dependencies, maintenance, documentation?

— Operationally, what does it take to stand up an external-control analysis pipeline — data access, curation, eligibility alignment, reproducibility — and where do timelines usually slip?

— Are there operational or computational considerations (for example, simulation studies for operating characteristics, or randomization inference at scale) that methodologists tend to underestimate?

PANEL · DISCUSSANT 5 OF 7 · INDUSTRY

Birol Emir

RWE Statistics Group Head · Data Sciences and Analytics · Chief Medical Office · Pfizer Research & Development

Biography

Birol Emir, Ph.D., is a statistician with experience across academia, industry, and professional service. He currently leads Real-World Evidence Statistics at Pfizer and has been serving as an Adjunct Professor at Columbia University. He is an ASA Fellow and has been actively involved in ASA through section leadership, organizing invited JSM sessions, and supporting Connecticut chapter activities. His primary focuses have been on hybrid designs (external control, pragmatic trials, data linkage), causal inference methods, AI/ML, and data analysis. He co-authored Interface Between Regulation and Statistics in Drug Development (Alemayehu, Emir and Gaffney, 2021) and co-edited a volume on health economics and outcomes research.

Questions for Birol Emir

— How can organizations align cross-functional teams to routinely evaluate, design, and communicate the strategic trade-offs and residual risks of external control arm strategies to senior leadership?

PANEL · DISCUSSANT 6 OF 7 · REGULATORY

Pallavi Mishra-Kalyani

Deputy Director, Division of Biometrics V · Office of Biostatistics, CDER · U.S. FDA

Biography

Pallavi Mishra-Kalyani, Ph.D., is Deputy Director of the Division of Biometrics V in the Office of Biostatistics, Center for Drug Evaluation and Research (CDER), U.S. FDA. Since joining FDA in 2015, she has worked on statistical issues in the use of real-world data and real-world evidence for regulatory purposes. Her research interests include observational methods, causal inference, and non-randomized trial design. She received her Ph.D. in Biostatistics from Emory University and a Master's degree in Epidemiology from the Harvard T.H. Chan School of Public Health.

Questions for Pallavi Mishra-Kalyani

— From a review perspective, what distinguishes a strong externally controlled trial submission from a weak one? Which elements of the evidence package — pre-specification, eligibility alignment, sensitivity analyses — most often decide the outcome?

— The talk argues that sensitivity analysis should be treated as a core inferential component rather than an appendix. How should sponsors present tipping-point or bias-bound analyses so they are actually usable in regulatory decision-making?

— Where do sponsors most commonly go wrong when proposing an external control in a pre-submission meeting, and what would you like to see discussed earlier in development?

PANEL · DISCUSSANT 7 OF 7 · REGULATORY

Hana Lee

Senior Statistical Reviewer · Office of Biostatistics, CDER · U.S. FDA

Biography

Hana Lee, Ph.D., is a Senior Statistical Reviewer in the Office of Biostatistics, CDER, U.S. FDA. She leads FDA-funded work supporting the agency's real-world evidence program, including causal-inference frameworks for non-randomized studies, machine-learning methods, and sensitivity analysis for RWE studies. She is a co-lead of the ASA Biopharmaceutical Section Real-World Evidence Scientific Working Group, the group that organized today's webinar.

Questions for Hana Lee

— “Fit-for-purpose” external data is foundational to everything discussed today. In practice, how do you assess whether a registry, natural history study, or EHR-derived cohort is fit for purpose as an external control?

— The paper emphasizes that adaptive borrowing methods should not substitute for data compatibility. How should sponsors demonstrate compatibility — concurrency, outcome definitions, prognostic covariate capture, missingness — before any statistical borrowing is considered?

— What real-world evidence lessons from recent externally controlled submissions would you highlight for statisticians planning their first ECT?

Q&A SESSION · 12:15–12:30

Audience Q&A

15 minutes

Open floor — type questions in the webinar chat

How we will take questions

Type questions in the webinar chat as they come up.
Ke Zhu and Peng Yang monitor the queue.
They will answer directly, or route a question to the panel.
Discussants stay on through 12:30.

Ke Zhu

Postdoc fellow
North Carolina State University &
Duke University
Q&A monitor

Peng Yang

Postdoc fellow
MD Anderson Cancer Center
Q&A monitor

Thank you

Zhu, Izem, Yang, Yuan, Pang, van der Laan, Nie, Emir, Mishra-Kalyani, Lee, Yang (2026)

arXiv:2605.03282 · https://arxiv.org/abs/2605.03282

ASA Biopharmaceutical Section · RWE Scientific Working Group · HCT Subgroup

Thank you to the panel, the Q&A monitors, and the audience