Designing for Impact

Introducing Translational Design Studies

NIH Collaboratory Grand Rounds

Christopher J. Lindsell · Lloyd Michener · Boris B. Volkov · Reuben Adatorwovor · Justin B. Moore

14 August 2026

For the Journal of Clinical and Translational Science

Based on: Volkov, Adatorwovor, Michener, Lindsell & Moore. Journal of Clinical and Translational Science.

Disclosures

Authors are members of the CTSA consortium and serve on the editorial board of the Journal of Clinical and Translational Science

Translational failure often begins
before the trial,
before implementation,
and before evaluation.

It becomes visible much later.

What we call failure is just where failure becomes visible

Not where it originates.

THE LABEL WE USE	THE DESIGN DECISION IT TRACES BACK TO
Recruitment failure	Eligibility built from registry codes, not from who is reachable and willing
Endpoint failure	Outcome chosen for regulatory precedent; never shown to move in this population
Implementation failure	Delivery model designed for a staffing pattern the sites do not have
Stakeholder resistance	Priorities assumed from literature rather than elicited from partners
Poor generalizability	Site selection optimized for enrollment speed, not for the inference target
Unreliable data capture	EHR field assumed to be complete because it exists in the data dictionary

The protocol converts assumptions into commitments

Every specified parameter is a claim. Most are never labelled as one.

WHAT THE PROTOCOL STATES
Sample size: 1,240
Primary endpoint: 90-day readmission
Eligibility: adults with HF, EF < 40%
Delivery: nurse-led, 4 sessions
Data source: EHR extract, monthly
Analysis: ITT, logistic regression

rests on

WHAT IT ASSUMES — UNTESTED
…an event rate we have not observed here
…that readmission is attributable and meaningful to patients
…that EF is recorded, current, and comparable across sites
…that sites have nursing capacity for a 4-visit model
…that the extract is complete at the time of analysis
…that adherence is high enough for ITT to be informative

The work happens. The evidence does not accumulate.

PUBLISHED, CITED, CUMULATIVE



DESIGN-PHASE INQUIRY — THE MISSING LAYER

Qualitative inquiry with patients and clinicians · Workflow mapping · Feasibility and operational fit · Candidate outcome validation · Readiness assessment · Co-design and refinement · Simulation and data-fitness work

Fragmented · relegated to supplements · forced into ill-fitting categories · rarely transferable · never cited as design evidence

Design is too important to remain largely unevidenced.

If a decision determines whether downstream work can succeed,

the basis for that decision belongs in the scientific record.

Translational Design Study

> Systematic inquiry conducted before, or alongside, planned translational work, with the explicit goal of generating generalizable empirical evidence that shapes what should be done, how it should be done, and with and for whom.

BEFORE OR ALONGSIDE

Not after the fact. Not a *post hoc* reframing of a study that did not work.

GENERALIZABLE

The finding must be usable by a team other than the one that produced it.

SHAPES THE DECISION

The output is a changed, justified, or improved design choice.

The category is defined by intent, not by method

NOT A METHOD CLASS

Qualitative, quantitative, and mixed methods all qualify.

Simulation qualifies. Interview studies qualify. Secondary data analysis qualifies. Measurement validation qualifies.

The method follows the design question — not the reverse.

IT IS AN INTENT CLASS

The study exists to reduce a named design uncertainty.

Test for authors and reviewers:
name the decision that changes, or is defensibly held, because of this evidence. If you cannot name it, it is not a Translational Design Study.

What it is not

Adjacent categories, separated by what the work is accountable for.

	PRIMARY QUESTION	EVIDENCE PRESENTED	TRANSFERABLE?
Protocol	What do we plan to do?	None — a plan	No
Pilot study	Can we run the procedures here?	Local operational metrics	Rarely
Feasibility study	Is this study viable in this setting?	Local recruitment, retention, fit	Sometimes
Implementation study	Does the strategy work in practice?	Adoption, fidelity, outcomes	Yes — post-launch
Translational Design Study	Which design choice is right, and on what evidence?	Empirical test of a named design assumption	**Yes — by requirement**

Four vantage points, one question set

1. What design uncertainty does this vantage point see clearly?
2. What assumption commonly goes untested?
3. What evidence can be generated before the main investment?
4. Which design decision becomes better?
5. What downstream failure is avoided?

Community engagement
Lloyd Michener

Evaluation
Boris Volkov

Biostatistics & informatics
Reuben Adatorwovor

Implementation science
Justin Moore

Same spine. Different uncertainties. The point is not four disciplines — it is one design problem seen four ways.

The people affected know things the protocol does not

Not a value statement. An epistemic one.

WHAT PARTNERS CAN OBSERVE DIRECTLY	HOW THE PROTOCOL USUALLY ESTIMATES IT
Who is reachable, and at what cost of time, travel, and trust	*Recruitment feasibility is estimated from registry counts*
Whether the outcome we measure is the outcome that matters	*Endpoint relevance is inferred from prior trials*
What participation actually costs a household	*Burden is judged by visit count, not by lived cost*
Which language, framing, and messenger are credible here	*Materials are adapted for reading level, not for credibility*

When community evidence changes the design

ASSUMPTION IN THE DRAFT DESIGN	DESIGN-PHASE EVIDENCE	CONSEQUENT DESIGN CHANGE
Monthly in-person visits are acceptable	Structured burden elicitation across 4 partner sites; wage-loss and childcare quantified	Two visits moved to remote; retention assumption revised downward and budgeted
Hospital-free days is a meaningful outcome	Ranking exercise with patients and caregivers; symptom burden ranked above utilization	Co-primary added; utilization retained as secondary with stated rationale
Clinic staff will refer eligible patients	Observation and interviews: referral competes with three other active protocols	Referral pathway redesigned around existing workflow; site load capped

Illustrative composites. The reportable unit is the third column — the changed decision, with its evidentiary basis.

Engagement as evidence, not ornament

Frameworks tell you what good engagement looks like. They do not tell you what works here.

WHAT WE HAVE

CDC/NIH Principles of Community Engagement

PCORI Engagement Rubric

NIH Roadmap for Engaging the Public;

NExTRAC ENGAGE report

Principles and expectations — necessary, and not self-executing.

WHAT IS STILL AN OPEN EMPIRICAL QUESTION

- Which governance structures actually shift decision authority
- Whether compensation models change who is able to participate
- How to measure participation quality rather than participation count
- What institutional coordination reduces cumulative burden on the same communities

A plausible pathway is still a claim



EACH ARROW CARRIES AN UNSTATED ASSUMPTION

…that these resources arrive, and persist beyond the grant	…that activities are delivered as specified, at dose	…that outputs are the right proximal signal of change	…that this mechanism operates in this context	…that effects hold once the study team leaves

The logic model is not a diagram of what we believe. It is a set of testable propositions about mechanism, context, and measurement.

Theory of change as a testable design object

DESIGN UNCERTAINTY	PRE-LAUNCH EVIDENCE	DESIGN IMPLICATION
Does the proposed mechanism operate here?	Rapid-cycle tests of the mechanism in two contrasting sites; measured on the proximal construct	Retain, respecify, or abandon the mechanism before the definitive trial
Is the proximal measure sensitive to change on the study's timescale?	Repeated measurement in a small cohort over the intended follow-up window	Change the measure, the timing, or the effect size the trial is powered to detect
Do partners and funders share a success criterion?	Structured elicitation of what each party counts as success, and at what threshold	Specify the criterion in advance, or state openly that it is contested

The design decision is the unit of learning

ITERATION READ AS WEAKNESS

"The protocol changed three times."

"They ran a pilot and then modified the intervention."

"The logic model was revised after the first year."

ITERATION READ AS EVIDENCE

"Three design assumptions were tested; two did not hold, and here is what changed."

"The revision is the finding — documented, methodologically defensible, and reusable."

Prove to improve: the contribution is better design, not a premature verdict on whether an intervention works.

Data availability is not data fitness

The field the query returns is not the field the design assumed.

AVAILABLE	COMPLETE	ACCURATE	TIMELY	COMPARABLE
The variable exists in the warehouse and the query returns rows	It is populated for the population you will actually enroll	It reflects the clinical state, not the billing artifact	It is present when the decision or the analysis needs it	It means the same thing across sites, systems, and time

THE COMMON FAILURE

Feasibility is established at the first column. The design is then built as though all five held.

The data can fail before the trial starts

DESIGN ELEMENT	THE ASSUMPTION	PRE-LAUNCH TEST
Endpoint	*The outcome is ascertainable, and varies enough to be informative*	Retrospective ascertainment audit; observed event rate and variance in this population
Eligibility	*Coded criteria identify the intended population*	Chart review on a sample; positive predictive value of the computable phenotype
Missingness	*Missing values are ignorable*	Characterize the missingness mechanism; sensitivity analysis under plausible alternatives
Heterogeneity	*One effect estimate is meaningful across sites*	Baseline variance decomposition; ICC estimation to inform clustering and power
Estimand	*Everyone means the same thing by 'the effect'*	Specify population, intervention, comparator, outcome, and summary measure explicitly

Simulation is evidence for design

Not a power calculation appendix. A study of how the design behaves under conditions you can name.

VARY WHAT YOU ARE UNCERTAIN ABOUT

- Plausible effect sizes and their priors
- Between-site variability and ICC
- Missing-data mechanisms and rates
- Accrual and dropout trajectories
- Operational constraints: staffing, timing, capacity

REPORT WHAT THE DESIGN CANNOT SURVIVE

- The range of conditions under which the trial is still informative
- The parameter at which power collapses — and how close reality is to it
- Whether an adaptive rule improves decisions or only adds complexity
- The decision threshold that is robust across plausible worlds

A simulation study that changes the endpoint, the sample size, or the decision rule is design evidence — and it is transferable to anyone facing that structure.

Can these data support this decision?

A decision screen for designs with predictive or risk-stratified components.

1. **Is the model discriminating and calibrated in this population?**
   Discrimination and calibration, not accuracy alone
2. **Was validation internal, and is external validation feasible?**
   Cross-validation or bootstrap; external where possible
3. **Are the predictors available at the moment the rule must fire?**
   Point-of-care timing, not retrospective completeness
4. **Does the rule change what anyone actually does?**
   Net benefit at the operating threshold, not AUC

IF ANY ANSWER IS NO

The finding is the design implication: simplify the rule, change the predictors, revise the threshold, or drop the stratification and state why.

Report against TRIPOD.

A negative answer, reported well, is a contribution.

Implementation failure often starts in design

THE SETTING THE DESIGN ASSUMED

- Dedicated study staff
- Protected time for delivery
- Stable teams and low turnover
- One competing priority
- Uniform workflow across sites
- Fidelity as the default

meets



THE SYSTEM THAT RECEIVES IT

- Clinical staff absorbing the work
- Delivery competing with throughput
- Turnover mid-study, and retraining costs
- Three other active initiatives
- Workflow that differs by site and by shift
- Adaptation as the default

The gap is not an implementation problem discovered later. It is a design decision made earlier, without evidence.

Context is not noise. It is part of the design space.

Fit

Does the intervention match the problem as the site experiences it?

Readiness

Is there capacity and commitment now, or only stated willingness?

Burden

What does delivery displace, and who absorbs the cost?

Adaptability

Which components are core, and which are negotiable?

Sustainability

What persists when study resources end?

THE DESIGN-PHASE VERSION OF THESE QUESTIONS

Not "did fidelity hold?" — but "which components can this system deliver, and which will it necessarily adapt?"

Asked before launch, answered empirically, and reported so the next system can use the answer.

Study the fit before testing the promise

THE DESIGN WINDOW

Cheap to change. Rarely studied. Rarely published.

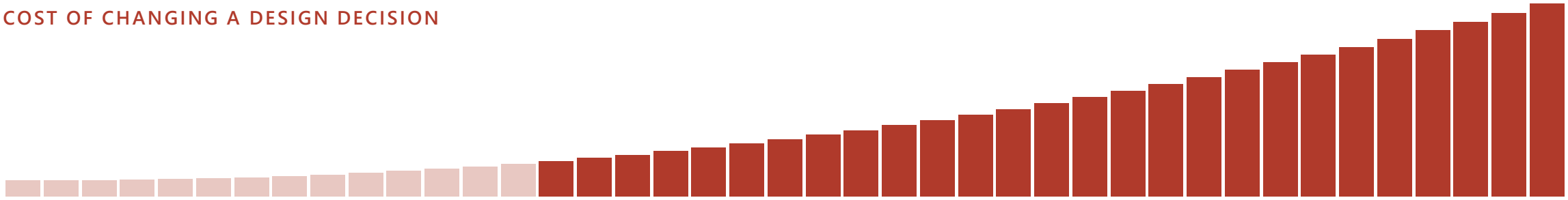
Intervention developed

Definitive trial

Implementation study

Scale-up

COST OF CHANGING A DESIGN DECISION



Questions asked in the design window are answerable with weeks of work. The same questions asked after launch cost a trial.

Four vantage points. One class of evidence.

COMMUNITY	EVALUATION	BIOSTATISTICS & INFORMATICS	IMPLEMENTATION
Relevance, burden, trust, outcome meaning	Mechanism, context, measurement, success criteria	Data fitness, estimand, endpoint, decision rule	Fit, readiness, adaptability, sustainability

A named design assumption, tested with reproducible methods, with a stated design implication that transfers

The four lenses do not each need their own article type. They need the same one.

What qualifies

Design-phase work that produces knowledge someone else can use.

- Engagement studies that change intervention or trial design
- Endpoint and outcome-measure validation in the target population
- Readiness and capacity assessments that alter site selection or scope
- Workflow and contextual analyses affecting delivery or implementation
- Simulations informing sample size, endpoints, adaptive rules, or thresholds
- Missingness and data-quality analyses that change the design
- Predictive-model validation for risk-stratified or tailored designs
- Theory-of-change or evaluation-logic studies that test design assumptions

In every case: a clear design question, an explicit assumption, reproducible methods, and results others can apply, adapt, and build on.

What does not qualify

Often valuable work. Not design evidence.

A protocol without empirical findings	*Describes a plan; presents no evidence*
A generic planning narrative	*No design question, no test, no result*
A routine pilot with no transferable insight	*Answers 'can we run it here' and stops*
A local feasibility report	*Real evidence, no generalizable knowledge claim*
A failed efficacy study reframed after the fact	*The design was not at risk when data were collected*
Stakeholder consultation without systematic inquiry	*Engagement occurred; no methods, no findings*
Operational project management work	*Important; not a scientific contribution*

The exclusions are not judgments of quality. They are judgments about what the work is accountable for.

What reviewers should evaluate

Judge the design question and the rigor — not proximity to conventional endpoints.

Importance	Is this design decision consequential for downstream translation?
Explicitness	Is the assumption or uncertainty named, not implied?
Method	Are the methods appropriate to the design question and reproducible?
Credibility	Does the evidence actually support the claim, with uncertainty quantified?
Implication	Is the design consequence stated clearly and specifically?
Transferability	Can a team elsewhere use this, and are the moderating conditions stated?
Accumulation	Does it add to a cumulative body of design knowledge, or stop at the project?

The last two are the ones we most often forget to ask — and they are what make this a literature rather than a collection.

These studies already exist

The challenge is not whether Translational Design Studies exist. They already exist — in your files, in your supplements, in your grant applications.

The challenge is whether we recognize them, evaluate them rigorously, publish them systematically, and learn from them collectively.

WHAT WE ARE ASKING FOR

- Submit the design work you have already done
- Review under the design rubric, not the trial rubric
- Argue with the boundaries — publicly
- Bring the cases where this framing fails

What important design decisions are currently being made in your science without evidence?

Discussion