FROM THE RESEARCH BENCH

What is external validity? A result needs a transfer boundary

External validity asks whether a result transfers across a named population, treatment, outcome, and context. Name the target before generalizing.

Sinan Isoglu, MBA (Quantic)

Commercial growth leader, lecturer and doctoral researcher

PUBLISHED
6 September 2026

READING TIME
12 min

WORDS
2,613

https://isoglu.com/insights/journal/what-is-external-validity/

CONTENTS

MANAGEMENT SUMMARY

External validity asks whether a result transfers to a specified target population, treatment, outcome, and context. It is not a universal property of a study. This article separates internal credibility from external transfer, distinguishes effect-generalization from sign-generalization, and uses a synthetic transfer matrix to show how a changed population, treatment, outcome, or context changes the claim. Egami and Hartman provide the primary framework; Campbell supplies historical planned-change and replication framing. The matrix, plain-language formulas, and review card are author conventions. They do not compute a transport estimate, validate a current company result, or replace target data.

Keywords: External validity · Generalizability · Transportability · Causal inference · Research design · Transfer boundary

A randomized experiment can be well designed and still answer a narrower question than a commercial team wants to ask. It may estimate an effect for eligible trial users in one country, under one onboarding treatment, measured with one event over 14 days. The moment a manager asks whether the same result holds for enterprise accounts, a sales-led implementation, or 90-day renewal, the question has changed.

That changed question is the starting point for external validity. It asks whether a result transfers to a specified target. It does not stamp a study as universally relevant.

The customer-model article keeps next purchase, partial defection, and profitability separate as predictive outcomes. The case-study design article keeps a case boundary and its transfer conditions visible. The measurement-invariance article asks whether a construct is measured comparably across English and German groups. This page owns the question between a result and a named target: what would have to remain comparable before the result could travel?

What does external validity mean?

Internal validity and external validity answer different questions. Internal validity asks whether the comparison or causal contrast is credible within the study design. External validity asks whether the same conclusion is credible for a target defined outside, or differently from, that source design.

Egami and Hartman formalize the distinction by defining the target explicitly. A researcher has to name the target population, target treatment, target outcome, and target context before asking whether the causal conclusion transfers. Their point is stronger than "state the limitations": no experiment is universally externally valid because every transfer question has a target.

The practical definition is therefore relational:

External validity is the credibility of a stated source result for a stated target, conditional on the dimensions and assumptions that connect them.

The source result and the target claim are not the same object. A study can be internally credible and still have an unresolved transfer boundary. Conversely, a broad target does not repair a weak source comparison. The two validity questions must be reviewed in sequence.

Which four dimensions change the transfer question?

Use four fields before using the word "generalizes." They correspond to the X-, T-, Y-, and C-validity dimensions in Egami and Hartman's framework:

Table 1 Which four dimensions change the transfer question?

Dimension	Object that can change	Diagnostic question	Typical hidden substitution
X-validity	Population or unit	Who was in the source, and who is in the target?	Trial users become enterprise accounts
T-validity	Treatment or implementation	Is the target intervention the same treatment in the relevant sense?	An email sequence becomes a sales-led program
Y-validity	Outcome or measurement	Is the target outcome the same construct, event, and horizon?	First report exported becomes 90-day renewal
C-validity	Context or setting	Could geography, institution, channel, language, regulation, or time change the mechanism?	One market's operating conditions become another's

Table from this essay. Sources and interpretation are given in the article.

The dimensions can change together. A German enterprise implementation may change population, treatment, context, and the practical meaning of the outcome at once. That is not a reason to give up. It is a reason to stop calling the transfer "the same result" until the changes and their consequences are visible.

Measurement deserves its own line inside the outcome field. A target can use the same label while recording a different event, threshold, or time horizon. The measurement-invariance question is not identical to external validity, but a changed measurement can block the outcome comparison on which transfer depends.

What is the difference between generalizability, transportability, and replication?

The labels vary by discipline, so do not let the label replace the target. Use them as practical decision words:

Table 2 What is the difference between generalizability, transportability, and replication?

Decision	Source and target relationship	Evidence still needed	Honest result
Replication	Population, treatment, outcome, and context are intended to remain the same, with a new observation or sample	Repeat the source design and check whether the result recurs	The source result was reproduced or not reproduced
Transport	At least one target field is intentionally different	A bridge through target data, measured effect modifiers, designed variation, or a new target study	The target inference is supported, limited, or unresolved
Analogy	A product, market, or mechanism merely sounds similar	A matched comparison has not been supplied	A hypothesis or prior, not a transfer result

Table from this essay. Sources and interpretation are given in the article.

Replication can strengthen confidence in the source claim without answering a different target question. Transport is the explicit attempt to answer that different question. Analogy is what remains when the similarity is described but the bridge is missing.

How do effect-generalization and sign-generalization differ?

Sometimes a reader wants the size of the effect. “Will the intervention raise first-value completion by about eight percentage points in the target?” That is an effect-generalization question.

Sometimes the reader needs only the direction. “Does the intervention improve the outcome in the target, even if the size is uncertain?” That is a sign-generalization question. The second goal is weaker than transferring a precise magnitude, but it is still a claim about a defined target and a defined outcome.

Weaker does not mean assumption-free. A sign claim still requires the target effect not to reverse across the relevant dimensions. Purposive variation can support the direction only when moderator interactions do not cross zero over the target-relevant range. If effects change sign in an unobserved or unsupported stratum, a positive source result cannot identify a positive target effect.

Egami and Hartman separate these goals because they need different assumptions and methods. For effect-generalization, they discuss weighting-based, outcome-based, and doubly robust estimators. For sign-generalization, they develop a multiple-testing procedure that can use purposive variation across relevant dimensions. The existence of a method does not mean that a reader has run it. The design and data still have to support the target question.

This distinction gives a useful stop signal. If a study can plausibly support only the direction, do not report a target magnitude as though it survived. If even the direction depends on an unmeasured context mechanism, the honest result may be “not identified.”

What does a transfer matrix look like?

The matrix below is a synthetic operating example. The source row imagines an illustrative result of plus eight percentage points for first-report export. No row is a real experiment or customer result. Each target row changes one field so the claim boundary can be seen before a team argues about the number.

Figure 1 The external-validity transfer matrix

ID	SOURCE OR TARGET OBJECT	TREATMENT AND COMPARATOR	OUTCOME AND WINDOW	CONTEXT	CHANGED FIELD	VERDICT
Stable row identifier.	Source result or named target.	What was done, and against what?	Which event or construct, and when?	Which operating or institutional setting?	Population, treatment, outcome, context, none; bridge?	Replicate, transport, limit or not?

Six rows are deliberate: one synthetic source result, one replication candidate, and four one-field target changes. The illustrative plus-eight-point effect is not a benchmark.

Author's synthetic transfer matrix grounded in Egami and Hartman (2023) and Campbell (1979). All rows, effects, and decisions are illustrative.

Table 3 What does a transfer matrix look like?

ID	Source or target object	Treatment and comparator	Outcome and window	Context	Changed field	Verdict
S-01	US self-serve SMB trial accounts	Email onboarding versus holdout	First report exported by day 14	US self-serve, synthetic source setting	None	Source result: +8 percentage points, illustrative only
R-01	New sample from the same US self-serve SMB target	Same email onboarding versus holdout	Same first-report event by day 14	Same source setting	None intended	Replication candidate; repeat the source design
T-01	US enterprise accounts	Same email onboarding versus holdout	Same first-report event by day 14	Same source setting	Population	Do not transport without target population evidence
T-02	Same US self-serve SMB accounts	Sales-led implementation versus holdout	Same first-report event by day 14	Same source setting	Treatment	New treatment question; source effect does not identify it
T-03	Same US self-serve SMB accounts	Same email onboarding versus holdout	90-day renewal	Same source setting	Outcome and time	New outcome question; activation result does not identify renewal
T-04	Same customer profile in a DACH operating context	Same email onboarding versus holdout	Same first-report event by day 14	Different regulatory and operating context	Context	Context bridge and target evidence required

Table from this essay. Sources and interpretation are given in the article.

The matrix does not say that every difference destroys transfer. It says that every difference has to be named. A population change may be addressed with target covariates and a credible overlap assumption. A treatment change may require a new experiment. An outcome change may require a new measurement and follow-up horizon. A context change may require evidence about mechanisms or effect-modifying conditions.

Which target data can support a transfer claim?

Start with the plain-language estimands:

Source claim: the effect of the studied treatment versus comparator on the declared outcome for the source population, under the source context and time horizon.

Target claim: the effect of the target treatment versus comparator on the declared target outcome for the target population, under the target context and time horizon.

The bridge between them is not a narrative sentence. It is evidence about the fields that changed. Relevant pre-treatment variables may include customer type, baseline capability, channel, geography, implementation model, or another effect modifier. The variable matters only if it is measured in a way that relates the source and target objects and if the study design supports the intended interpretation.

Egami and Hartman describe three broad estimator families for effect-generalization: weighting-based, outcome-based, and doubly robust approaches. They also emphasize that the target data and assumptions must make the chosen goal achievable. A reader should therefore record whether the target variables are observed, whether the source and target overlap, and whether the treatment and outcome are actually the same objects.

That overlap has a stricter name: positivity, or common support. For every measured effect-modifier stratum represented in the target, the source data must contain units with non-zero probability of representation, and the treatment and comparator needed for the target contrast must each be observed with non-zero probability. Operationally, check $P(S = 1 \mid X = x) > 0$ and $P(A = a \mid X = x, S = 1) > 0$ for every target-relevant x and each treatment value a. Weighting, outcome-modeling, and doubly robust estimators handle the bridge differently, but without this support the target effect is not non-parametrically identified from the source result. If the target contains feature combinations absent from the source, such as enterprise deal sizes or legacy security constraints, transport requires new design or explicit extrapolation assumptions.

If a target variable is missing, the target is outside the observed range, or the mechanism changes with context in a way the data cannot address, the transfer is not identified by the source result. That is a valid finding. A methods page should make it possible to stop without inventing confidence.

Why does one replication not prove external validity?

Replication and transfer are related but not interchangeable. A replication can repeat the source population, treatment, outcome, context, and analysis in a new sample. If it recurs, the source claim has stronger support under that design. It still does not answer what happens when the population, treatment, outcome, or context changes.

Campbell's discussion of planned social change is useful as historical framing because it keeps assessment tied to the setting in which a change was planned and to the practical need for replication and cross-validation. The lesson here is modest: a result should not be detached from the setting and then treated as portable by default.

This is also why "the same mechanism" is not enough. Mechanisms are embedded in people, institutions, incentives, measurement, and timing. If the target changes one of those conditions, the mechanism is a hypothesis to examine, not a certificate of transfer.

What is external validity not?

External validity is not:

- A universal quality score. A study can be excellent for a narrow target and uninformative for another.
- Statistical significance. A small p-value describes evidence within the specified design. It does not name a new target.
- A sample-size threshold. More observations can improve precision without creating population, treatment, outcome, or context overlap.
- A product or market similarity label. “B2B,” “onboarding,” or “activation” can conceal different units and interventions.
- A substitute for measurement checks. The outcome must still represent the same construct or event in the target.
- A current-company result. A published study does not validate a claim about the owner’s customers, offer, or market.
- A causal effect of the transfer process. The matrix organizes an inference; it does not itself create a counterfactual.

The defensible evidence review is the adjacent owner for search closure and extraction. It does not remove the need to define the target when the conclusion is moved beyond the reviewed sources.

How should a team review a transfer claim?

Use the following sequence before a study result enters a business case, product decision, or research conclusion:

1. Write the source estimand. Name the source population, treatment, comparator, outcome, measurement, context, and time horizon.
2. Name the target estimand. Write the population, treatment, comparator, outcome, measurement, context, and horizon the reader actually cares about.
3. Mark the four dimensions. Label every change as population, treatment, outcome, context, or a combination.
4. Choose the goal. Decide whether the claim needs effect magnitude or only effect direction.
5. List the bridge. Record effect modifiers, target data, overlap, measurement checks, and assumptions that could connect source to target.
6. Classify the decision. Call it replication, transport, or analogy. Do not use “generalizes” as a conclusion without the classification.
7. Write the smallest verdict. Use supported, limited, not identified, or new evidence required. Record what would change the verdict.

Table 4 How should a team review a transfer claim?

Pattern	First question	Smallest honest verdict
Same source objects and setting, new sample	Was the source design repeated without changing the estimand?	Replication candidate
New population with measured source-target overlap	Are the variables that modify the effect observed in both populations?	Transport may be analyzable under stated assumptions
New treatment implementation	Is the target intervention equivalent in the causal sense, not only in its label?	New treatment evidence required
New outcome or horizon	Does the target event measure the same construct at the same time boundary?	Do not infer the target outcome from the source result
New context or institution	Could the mechanism change with regulation, channel, language, or time?	Context bridge required
Similar label with no comparison record	Which fields are actually matched, and which are assumed?	Analogy only
Target data or overlap missing	Can the target claim be evaluated with the available design and data?	Not identified

Table from this essay. Sources and interpretation are given in the article.

This card does not force every result into a transport analysis. It gives the reviewer a visible way to decline a larger claim, request a new study, or narrow the decision to the source setting.

Can external validity make a decision more confident?

It can make the decision more precise, but it cannot manufacture evidence. The useful output may be a supported transfer, a bounded transfer, a replication request, a target-data request, or a clear stop.

Egami and Hartman explicitly note that credible effect- or sign-generalization can be impossible given the design, available data, or nature of the target problem. That is not a failure of writing. It is the correct result when the target claim outruns the bridge.

The sentence to carry forward is therefore specific: for this source result, this target changes these fields, this is the transfer goal, these assumptions connect the two, and this is what remains unidentified. That is external validity as an operating discipline. It is more useful than calling a result “generalizable” and leaving the target unnamed.

Campbell, D. T. (1979). Assessing the impact of planned social change. Evaluation and Program Planning, 2(1), 67-90. https://doi.org/10.1016/0149-7189(79)90048-X

Egami, N., & Hartman, E. (2023). Elements of external validity: Framework, design, and analysis. American Political Science Review, 117(3), 1070-1088. https://doi.org/10.1017/S0003055422000880

END OF ESSAY

HOW TO CITE THIS ESSAY

Isoglu, S. (2026, September 6). What is external validity? A result needs a transfer boundary. Sinan Isoglu. https://isoglu.com/insights/journal/what-is-external-validity/

KEEP READING

From the research bench

Staggered difference-in-differences needs a cohort-time estimand

https://isoglu.com/insights/journal/staggered-difference-in-differences-needs-a-cohort-time-estimand/

From the research bench

An uplift claim needs a specification before it needs a number

https://isoglu.com/insights/journal/an-uplift-claim-needs-a-specification/

From the research bench

What is reproducibility? Can another analyst reconstruct the result?

https://isoglu.com/insights/journal/what-is-reproducibility/

Sinan Isoglu, MBA (Quantic)
Commercial growth leader, lecturer and doctoral researcher

Online version: https://isoglu.com/insights/journal/what-is-external-validity/