STANDARD

The Risk Taxonomy of Evaluation Failure

How to identify evaluation risks before commitment

5	4	3
DESIGN	CONTEXT	ETHICS

Twelve risks in three families. A sign is a reason to investigate, not proof.

The Risk Taxonomy of Evaluation Failure

Evaluations can fail through problems in design, operating context or ethics. A pre-mortem imagines that the planned work has failed and asks what could explain it. This catalog gives recognizable signs and possible responses. A sign is a reason to investigate, not proof that a failure will occur.

Document expertise, incentives, financial ties, access and reporting control. Reviewers may need methodological, operational, ethical and affected-community perspectives. Separation from delivery can help impartial review; it cannot substitute for these checks.

DESIGN	CONTEXT	ETHICS
Underpowered for a decision-relevant effect	Political timing	Unjustified or harmful allocation
The wrong outcome	Partner capacity	Consent or refusal that is not meaningful
Interference or an unclear treatment contrast	Data that does not exist in the form required	Data use without adequate standing or safeguards
Missing outcomes and attrition	Ambition beyond the budget	
Open-ended analysis or selective reporting		

The catalog at a glance. It is nonexhaustive, and sector-specific threats may need specialists.

© 2026 Prova LLC · All rights reserved · PP-000248-10

01 Design: the planned evidence may not answer the question

Underpowered for a decision-relevant effect. A design may be unable to distinguish an important effect from sampling uncertainty. An assumed effect larger than plausible experience can conceal this risk; a small minimum detectable effect can also be irrelevant to the choice. *The tell:* inadequate justification for effect size, variability, clustering, enrollment, follow-up or precision. *The response:* examine realistic scenarios and the smallest difference relevant to action. Rescope, extend or redesign where feasible. If precision remains inadequate, state that limitation. A non-significant estimate does not establish no effect; low power does not guarantee a false null.

The wrong outcome. The instrument may record something adjacent to the intended benefit, or omit a consequential harm. *The tell:* outcome meaning, population or timing does not fit the question, or the theory of change substitutes for evidence of fit. *The response:* justify the construct and source, test comprehension and feasibility, and specify primary and additional outcomes. Document changes before relevant results are examined. Pre-specification alone does not make a measure valid.

Interference or an unclear treatment contrast. Shared markets, classrooms or information can transmit effects across groups. Receipt and substitutes can differ from offers. *The tell:* plausible spillover routes or unrecorded crossover. *The response:* define the intended contrast and relevant interference structure; consider cluster allocation or a design that estimates spillovers where feasible. Check the consequences for power and scope. Measuring crossover alone does not repair the inference, and complete isolation may be impossible or contrary to the question.

Missing outcomes and attrition. Missingness can affect either group, even at similar rates. *The tell:* long follow-up, mobile participants, absent records or no feasible tracking plan. *The response:* establish comparable collection and authorized follow-up, budget realistic retention, examine reasons and patterns, and use justified sensitivity analyses or bounds. Equal response rates or similarity on measured

© 2026 Prova LLC · All rights reserved · PP-000248-10

baseline characteristics do not establish unbiased outcomes. Keep unobserved outcomes unknown; assumptions for imputation or bounds must be explicit.

Open-ended analysis or selective reporting. Many choices about outcomes, exclusions or subgroups can make a favorable finding easier to obtain. *The tell:* unclear primary questions, undocumented analyst access, undisclosed changes or unreported outcomes. *The response:* specify the principal analyses before relevant results are inspected, retain dated changes and report exploration as exploration. Address multiple testing and material deviations. Registration supports transparency; it neither guarantees validity nor prohibits justified new inquiry.

© 2026 Prova LLC · All rights reserved · PP-000248-10

02 Context: operating conditions can change delivery, evidence or use

Political timing. An election, leadership change or policy shift may alter delivery, comparison or use. *The tell:* a timetable spanning a relevant event, or results arriving after the actual decision. *The response:* identify scenarios, feasible reporting points and authorized responses to disruption. A timing buffer alone does not repair a changed counterfactual. Record whether the current question must be narrowed, deferred or redesigned.

Partner capacity. A partner may be unable to sustain enrollment, delivery or research records as planned. *The tell:* untested workflows, thin systems, staff load or a protocol dependent on undocumented exceptions. A missing research track record alone is not evidence of failure. *The response:* test the actual workflow with frontline staff, resource the work and reduce avoidable burden. Resolve assignment, delivery and recording responsibilities before launch.

Data that does not exist in the form required. A plan assumes records, linkage, coverage or permissions that have not been established. *The tell:* required fields, time points or people are absent, or access is only promised. *The response:* verify a sample of actual records, definitions, linkage quality, legal and ethical permission, and cost. Build a feasible channel or rescope. Existing records do not imply free or authorized reuse.

Ambition beyond the budget. Outcomes, sites, arms or follow-up exceed feasible resources. *The tell:* staffing, access, tracking, analysis or review costs are absent from the plan. *The response:* compare scoped alternatives against the evidence job, learning purpose and deadline. Resource the core or narrow it openly. Ethics, required protections and critical outcome validity cannot be traded away to preserve an ambitious title.

© 2026 Prova LLC · All rights reserved · PP-000248-10

03 Ethics: rights, burdens, consent and data use

Unjustified or harmful allocation. Random allocation can affect urgent needs, entitlements, alternatives and fair access. *The tell:* uncertainty about rights or safety, no defensible allocation rationale, or affected people excluded from the account of burden. *The response:* establish applicable duties and the proper ethical review before changing access. Genuine uncertainty or scarcity does not alone authorize a lottery. A phase-in can still delay necessary service and needs its own justification. Where appropriate, preserve priority routes and scope learning around permissible allocation.

Consent or refusal that is not meaningful. A signed form can conceal poor understanding or pressure. *The tell:* inaccessible explanation, fear of losing service, concealed reuse, or no way to ask questions or withdraw where applicable. *The response:* use a comprehensible, truthful process appropriate to the setting; distinguish research participation from service eligibility and explain the actual consequences of refusal. Revisit material changes. Establish applicable requirements and any authorized exceptions through the proper review; do not treat a form or this catalog as that authorization.

Data use without adequate standing or safeguards. People may bear data collection risks while having little influence over use or benefit. *The tell:* unclear access, retention, disclosure, complaints or representation, especially for people who could be harmed. *The response:* establish permissions and proportionate governance, limit access and disclosures, and provide meaningful routes for affected perspectives and challenge. A committee or a claim of shared ownership does not establish consent, legal rights or representative authority.

© 2026 Prova LLC · All rights reserved · PP-000248-10

04 A worked pre-mortem

The people, places, organizations and figures in this document's examples are fictional.

Consider a multi-site youth-employment trial. Its power model assumes a large effect without supporting evidence. Enrollment and follow-up workflows remain untested. A scheduled state policy change may alter access in comparison sites before the outcome is measured.

The review records three tasks before commitment. The analyst reruns plausible precision scenarios, including clustering and attrition. The operations lead tests enrollment, priority routes and authorized outcome access, with staff time costed. The policy lead maps which groups the change affects and asks whether the proposed causal contrast remains defensible.

The actual commissioning authority receives the unresolved conditions and scoped alternatives. Methodological and ethical reviews address their respective questions. Approval is conditional on specified evidence and responsibilities; a review checklist does not grant it. If the required contrast or protections cannot be established, redesign or defer the study. A bounded existing account may serve the present decision.

The example shows a response to identifiable risks.

© 2026 Prova LLC · All rights reserved · PP-000248-10

05 The discipline

For each risk, record **failure pathway / evidence and contrary evidence / likelihood judgment and its basis / consequence / proposed response / owner / cost and permission dependencies / due date / residual uncertainty / pause condition / responsible decision authority**. An unresolved entry remains unresolved. Likelihood categories are qualitative judgments, not calibrated probabilities. Do not multiply checklist scores into an approval rule.

Some conditions, including unlawful allocation or unaddressed serious harm, require refusal or proper resolution regardless of a reassuring average rating. A less likely failure can also be severe enough to change the plan. Record disagreements and review again when allocation, staffing, source access, policy or outcomes change.

The catalog is nonexhaustive. Sector-specific threats may need specialists. Use the same scrutiny for Prova's own plans.

RELATED: The Honesty Layer Rubric, for reviewing claims against evidence; The Identification Atlas, for comparison patterns; Experimental Design Preflight, for checking a randomized design before enrollment.

© 2026 Prova LLC · All rights reserved · PP-000248-10

06 Further reading

1 **World Bank/IDB, Impact Evaluation in Practice (2016)**, power, assignment and attrition passages: design feasibility and uncertainty.

2 **American Evaluation Association, Guiding Principles**, inquiry, integrity, respect and equity principles.

3 Gary Klein, the pre-mortem technique.

QUESTION IT

Found an error, or want to question this document? Contact Prova at hello@byprova.com.

© 2026 Prova LLC · All rights reserved

Cite as: Prova (2026). The Risk Taxonomy of Evaluation Failure. byprova.com

Examples are fictional unless sourced. Not legal or financial advice.