OPERATOR FIELD GUIDE

From Noise to Decisions

An Operator's Guide to Decision-Grade Evidence

A practical method for deciding what to test, what to stop and what evidence is strong enough to act on.

Jörn Green

Green Bear Consulting

First edition | August 2026

How to use this guide

Monday's leadership meeting ends with three approvals: extend the pilot, add two features and enter another segment. Each decision sounds reasonable. None names the assumption it depends on, the evidence that would change it or what the company will stop to make room. Six weeks later all three initiatives are active and the original questions remain unanswered.

This guide is for that moment. Use it before approving a material initiative, during a stalled portfolio review, or when a leadership team cannot agree on what its results mean. Its purpose is narrower than strategy and broader than experimentation: make the decision explicit, identify what must be true, collect evidence that is fit for that decision and record what follows.

> **The promise.** You will not remove uncertainty. You will make it visible, locate the evidence gap that matters and decide how much uncertainty the organisation is prepared to carry.

The method at a glance

1. Define the decision before debating the initiative.
2. Map the causal chain from intervention to realised value.
3. Find the assumptions with the highest consequence and weakest evidence.
4. Grade each observation against the claim it is meant to support.
5. Design the smallest test capable of changing the decision.
6. Pre-commit four legitimate outcomes: proceed, revise, stop or inconclusive.
7. Record the decision, what remains uncertain and what is now closed.
8. Install a review rhythm so evidence compounds instead of resetting.

SignalWorks and the three shorter cases are hypothetical composites, not disguised client stories. This is an operating method, not scientific or specialist assurance.

Contents

1. When Activity Stops Creating Knowledge

Recognise the point at which more work is producing interpretation rather than a better decision.

A growing company can look healthy while its understanding is deteriorating. Customers are talking to the team. Features are shipping. The pipeline contains names. Dashboards move. Every initiative has an intelligent owner and a plausible reason to exist. Yet the same questions return because none of the activity has been designed to answer them.

This is not chaos. The problem is more specific: the organisation cannot show how recent activity changed a material belief or decision. Success creates stories, failure creates explanations, and neither reliably updates what the company will do next. Learning becomes a side effect rather than an operating requirement.

Noise is decision-relative

A customer interview is useful for understanding language and context. It is weak evidence of repeated use. A signed pilot can demonstrate commitment under negotiated conditions. It says little about scalable delivery. A payment can establish willingness and budget for a transaction. It does not prove retention or realised value. The same observation can be signal for one claim and noise for another.

Working definition. Noise is information or activity that cannot meaningfully update the decision currently in front of the organisation. Uncertainty is the part of that decision that remains unresolved after the available evidence is stated honestly.

Five warning signs

- The same initiative is approved repeatedly under different names.
- Teams report activity and outputs, but cannot name the behaviour or business outcome expected to change.
- A result is called positive without a threshold agreed in advance.
- Contradictory evidence is explained away rather than recorded.
- New work begins without an explicit decision about existing work.

Protect work that is not an experiment

A commercial evidence cycle must not be used to cut obligations that keep the company safe and credible. Contractual delivery, security, compliance, reliability, maintenance and legally required controls may create little new learning and still be mandatory. Separate them from discretionary work before asking what to stop.

Category	Purpose	Primary review question
Run	Deliver the existing business.	Is it meeting the service, cost and quality commitment?
Protect	Control material risk and fulfil obligations.	Is the control adequate for the exposure?
Learn	Resolve uncertainty that blocks a decision.	What decision will this evidence change?
Scale	Expand something already supported by evidence.	Are the mechanism and economics holding at greater volume?

The method in this guide focuses on Learn and Scale. Run and Protect work still require owners, measures and review, but their legitimacy does not depend on testing a commercial assumption.

Start with one initiative

Choose one initiative that leadership expects to matter within the next ninety days. Do not begin with the whole strategy or a catalogue of every assumption in the company. Ask one question: what decision is this work meant to inform? If nobody can answer in a sentence, the initiative is not yet ready for approval.

> **Chapter output.** A named initiative and the decision it is supposed to change. The next chapter turns that sentence into a Decision Brief.

2. Define the Decision

Make ownership, timing, reversibility and the required strength of evidence explicit before the team debates solutions.

Teams often begin with an idea: build the integration, extend the pilot, enter Germany, hire sales. That invites a discussion about whether the idea is attractive. Begin instead with the choice the organisation must make and the consequence of choosing poorly.

SignalWorks sells operational analytics to mid-market fleet companies. Three prospects have completed custom pilots. Users praised the reports, but delivery required direct founder involvement and manual data preparation. The proposed initiative is a twelve-week productisation programme. The actual decision is whether the company should commit a dedicated team and postpone two other roadmap items to pursue a repeatable fleet-operations offer.

The six fields of a decision

- Decision: State the choice in terms that allow a clear yes, no, staged commitment or redesign.
- Owner: Name the person with authority to close the choice. Consultation can be broad; accountability cannot be collective.
- Deadline: Set the latest useful decision date. Evidence that arrives after the commercial window may be accurate and worthless.
- Consequence: Describe the capital, opportunity, dependency or risk created by a wrong decision.
- Reversibility: Estimate how easily the commitment can be undone and what reversal would cost.
- Evidence threshold: Specify how strong the evidence must be for this stage, given the exposure and the cost of delay.

Reversibility changes the threshold

A cheap, reversible trial can proceed with limited evidence when delay is costly. A commitment that creates long contracts, regulatory exposure, migration risk or a large dedicated organisation requires a stronger case. The point is not to classify every decision perfectly. It is to prevent a small experiment and a one-way commitment from travelling through the same approval process.

Amazon's distinction between one-way and two-way doors is useful here: some choices can be reversed cheaply, while others deserve more deliberate scrutiny. In practice, many decisions can be staged. Approve a bounded test now while withholding the irreversible commitment until the next evidence gate.

SignalWorks - Decision Brief

Field	Record
Decision	Should SignalWorks fund a twelve-week productisation programme for the fleet-operations offer?
Owner	CEO, after input from Product, Delivery and Commercial leads.
Deadline	4 September, before the next roadmap cycle is locked.
Commitment	One product squad for twelve weeks; two roadmap items delayed.
Consequence if wrong	Lost roadmap capacity, continued founder dependency and delayed learning in the core product.
Reversibility	Staged. Four-week evidence gate before the remaining eight weeks are released.
Evidence required now	Evidence that target users return without founder prompting, onboarding can stay within a defined effort limit and at least two buyers accept the proposed commercial terms.

A decision is not a forecast

The Decision Brief does not require certainty or a detailed business case. It creates a boundary around the uncertainty. The owner can now ask whether the available evidence is sufficient for a four-week gate, rather than pretending to know whether the entire market will succeed.

Chapter output. A Decision Brief with one owner, one deadline, an explicit commitment and an evidence threshold proportionate to reversibility and exposure.

3. Build the Causal Chain

Expose the mechanism that is expected to connect the initiative to customer behaviour and realised value.

An initiative earns attention by sounding plausible. A causal chain forces the team to show where that plausibility could break. The chain is not proof. It is a shared model of how the proposed action is expected to produce value, including the boundary within which the claim will be tested.

The complete chain

Element	Question
Decision	What choice will the evidence inform?
Intervention	What will the company change or provide?
Mechanism	Why should that intervention alter behaviour?
Behaviour	What will a specific actor do differently?
Outcome	What operational or commercial result should follow?
Value	Where is value realised, by whom and over what period?
Boundary	For which population, context and time window is the claim being made?

The mechanism is the step most teams skip. “We add predictive alerts, customers use the product more, retention improves” contains an intervention, behaviour and outcome. It does not explain why the alert changes the user’s work. The missing mechanism might be earlier detection of costly maintenance events. If users cannot act on the alert, the feature can work technically while the mechanism fails operationally.

SignalWorks causal chain

Running example

Field	Record
Decision	Release the remaining eight weeks of productisation funding.
Intervention	A standardised weekly fleet-risk review with automated data ingestion and ranked exceptions.
Mechanism	The workflow reduces manual analysis and directs the fleet manager to a small number of preventable cost or safety issues.

Field	Record
Expected behaviour	Fleet managers review the ranked exceptions weekly and assign follow-up actions without prompting from SignalWorks.
Outcome	More issues are investigated earlier with less analyst effort.
Value	The customer avoids preventable loss and reduces review time; SignalWorks can deliver without bespoke founder work.
Boundary	Mid-market fleets with an existing telematics source, during a four-week pilot, with no live weekly facilitation.

Mark the measurement boundary

Each link needs an observable indicator, but the indicator must remain close to the claim. Opening a report can show access. It does not show that the user understood the exception, took action or realised value. A follow-up task can show workflow adoption. It does not by itself establish that the task prevented a cost. Record where observation ends and inference begins.

List competing explanations

A result rarely has one possible cause. Pilot users may return because the founder reminds them, because participation is free, because a consultant prepares the data or because the product genuinely improves their work. Write the most credible alternatives before the test. Then design the observation so at least one important alternative becomes less plausible.

Observed result	Possible explanation	How to distinguish it
Weekly report opened	Automated reminder created the activity.	Compare opening with and without personal follow-up.
Actions assigned	Founder interpreted the data live.	Remove live facilitation and track independent completion.
Pilot renewed	Relationship or discount drove the renewal.	Present standard terms and record who controls budget.

The chain does not need to be exhaustive. It needs enough precision to reveal which link carries the decision. For SignalWorks, the critical question is not whether people like the report. It is whether the workflow produces repeated independent action while delivery effort stays inside a scalable boundary.

Chapter output. A Causal Chain that names the intervention, mechanism, behaviour, outcome, value, boundary and credible competing explanations.

4. Find the Critical Assumptions

Select the few beliefs whose failure would change the decision and whose evidence is currently weakest.

Every causal chain depends on many conditions. Listing all of them creates a risk register too large to guide action. The job is to identify the critical few: assumptions with meaningful consequences if wrong, limited current support and a realistic path to learning before the decision deadline.

Select rather than score

False precision is tempting. A formula that multiplies impact by uncertainty can make weak judgements look objective. Use simple categories and expose the reasoning. For each assumption, ask: would this change the decision, what evidence currently supports it, what would contradict it, and can we learn in time?

Dimension	Low	High
Decision consequence	The decision survives if the assumption is wrong.	The initiative must be redesigned, staged or stopped.
Evidence weakness	Repeated direct evidence exists in the target context.	Support is indirect, one-off, prompted or absent.
Learning urgency	The answer can wait without material cost.	The next commitment will make learning slower or more expensive.

Use confidence labels carefully

Confidence is a judgement about the current record, not a probability of truth. Four labels are usually enough: unexamined, plausible, supported or contradicted. The label must link to evidence in the ledger. “Supported” means sufficient for the current decision stage; it does not mean permanently proven.

SignalWorks - Assumption Register

Field	Record
A1 - Problem recurrence	Fleet managers face the target review problem often enough to justify a weekly workflow. Confidence: plausible.
A2 - Independent use	Users will review and assign actions without founder prompting. Confidence: weak; current pilots were facilitated.
A3 - Delivery boundary	A new fleet can be onboarded in six hours or less with the standard connector. Confidence: unexamined.

Field	Record
A4 - Buyer and budget	The operational owner can access budget at the proposed annual terms. Confidence: plausible; one prior paid custom pilot.
A5 - Outcome relevance	Assigned actions address issues valuable enough to sustain renewal. Confidence: weak; realised outcomes were not tracked consistently.
Critical for next gate	A2 and A3. Failure of either blocks release of the remaining productisation commitment.

Write the consequence of being wrong

An assumption becomes useful when its consequence is explicit. If independent use fails, SignalWorks may still have a valuable analyst-led service. That is a different product and cost structure. If onboarding exceeds the boundary, the company can redesign ingestion before testing broader demand. The evidence does not merely approve or reject the idea; it reveals which business the company is actually building.

Give assumptions owners and expiry dates

The owner is responsible for maintaining the evidence record, not for defending the assumption. Add a review date. Market conditions, product design and customer context change; an assumption supported six months ago may no longer justify the next commitment.

Chapter output. An Assumption Register that identifies the critical few, current confidence, supporting and contradicting evidence, consequence if wrong, owner and review date.

5. Grade the Evidence

Judge each observation by the claim and decision it can support, rather than placing all signals on one universal ladder.

Organisations rarely lack data. They lack agreement about what a specific observation establishes. Interest, behaviour, commitment and payment answer different questions. None is automatically superior outside the claim being tested.

Six evidence dimensions

Dimension	Question
Directness	How closely does the observation measure the behaviour or outcome in the claim?
Cost	What meaningful time, money, reputation or opportunity did the actor commit?
Repetition	Did the behaviour occur once, or persist across relevant occasions?
Representativeness	Does the population and context resemble the decision target?
Independence	Could prompting, incentives, facilitation or the relationship explain the result?
Proximity to value	Does the evidence show exploration, use, retention, willingness to pay or realised outcome?

These dimensions are a profile, not a composite score. A small sample can be direct and independent but poorly representative. A signed contract can show cost and willingness to pay while remaining dependent on a founder relationship. Preserve the shape of the evidence instead of hiding it behind one number.

What common observations can establish

Observation	Can support	Does not establish by itself
Positive interview	Problem language, context and stated priorities.	Adoption, repeated use or budget.
Signup or demo request	Interest strong enough for a low-cost action.	Activation, retention or realised value.
Repeated use	Behaviour under the observed conditions.	Causality, willingness to pay or scalable delivery.
Resource commitment	Internal priority and switching cost.	Successful outcome or durable demand.
Payment	Willingness and budget under the transaction terms.	Retention, realised value, causality or scalable demand.

Observation	Can support	Does not establish by itself
Measured customer outcome	Change in the observed outcome.	That the product caused the change without further design.

Record contradiction as evidence

Averages and executive summaries often remove the cases that matter most. If four users return and two do not, record the difference in context. If the paid customer required bespoke delivery, preserve that fact next to the payment. Contradiction is not a defect in the story. It is a clue to the boundary of the mechanism.

SignalWorks - Evidence Profile for independent weekly use

Field	Record
Current observation	Three custom pilots completed; users opened weekly reports and assigned actions.
Directness	Moderate. The target behaviour occurred, but facilitation was present.
Cost	Moderate. Customers contributed data and staff time; one paid for a custom pilot.
Repetition	Limited. Behaviour repeated for four weeks in two pilots.
Representativeness	Weak to moderate. All three customers came through founder relationships.
Independence	Weak. The founder joined weekly reviews and prepared several interpretations.
Proximity to value	Use and some action assignment observed; realised operational value not measured consistently.
Decision implication	Insufficient to approve scale. Sufficient to justify a bounded test without facilitation.

Evidence is allowed to remain insufficient

The pressure to tell a coherent story can turn missing evidence into optimistic interpretation. The correct conclusion may be that the current record does not support the commitment. That is useful if it leads to a smaller next step, a redesigned proposition or an explicit decision to carry the uncertainty.

Chapter output. An Evidence Profile linked to one assumption and decision, including contradictions, limitations and the exact claim each observation can support.

6. Design the Smallest Decision Test

Create the least expensive credible observation capable of changing the next action.

A test is valuable when it can change a decision. It does not need to validate the company, prove the market or produce a statistically generalisable conclusion. It must answer a bounded question with enough credibility for the commitment at hand.

Start from the decision rule

Write the four possible actions before designing the activity. This prevents a convenient measure from becoming the goal and limits post-hoc reinterpretation. The thresholds can still require judgement, but any change after the test begins must be recorded as a change, not presented as the original rule.

Outcome	Use when	Required action
Proceed	The agreed threshold is met and no guardrail is breached.	Release the next staged commitment and name the next open assumption.
Revise	The mechanism has support, but the offer, workflow or test design needs a material change.	Document the change and permit a bounded retest only if the remaining uncertainty justifies it.
Stop	Evidence materially weakens the critical assumption or the economics no longer justify learning.	Close the initiative or reclassify it as a different proposition.
Inconclusive	The test could not answer the question because execution, sample, measurement or context failed.	Name the failure, limit repetition and decide whether another test is worth its cost.

The Decision Test Card

- Scope: the assumption, population and context to which the result may apply.
- Method: the intervention and, where feasible, a comparison or baseline.
- Decision signal: one primary measure, a threshold and a useful time window.
- Guardrails: safety, service, cost, ethics, legal and customer-experience boundaries.
- Owner: the person accountable for execution and record quality.
- Decision rules: proceed, revise, stop and inconclusive conditions agreed before launch.

SignalWorks - Decision Test Card

Field	Record
Assumption	Target users will run the weekly exception workflow without founder prompting, within a scalable onboarding boundary.

Field	Record
Population	Six mid-market fleet teams using a supported telematics source; no existing close founder relationship.
Intervention	Standard onboarding, automated ranked exceptions and written workflow guidance. No live weekly interpretation by SignalWorks.
Primary measures	Weekly review completed and follow-up actions assigned; onboarding hours recorded separately.
Threshold	At least four teams activate; at least three complete the workflow in three of four weeks; at least two request continuation at the stated annual terms.
Window	Four weeks after onboarding.
Guardrails	No more than six SignalWorks hours per onboarding; critical false alerts escalated within one business day; customer data handled under existing controls.
Proceed	Behaviour, commercial continuation and onboarding threshold all hold.
Revise	Independent behaviour holds, but onboarding or proposition design breaches a guardrail and has one credible bounded remedy.
Stop	Fewer than two teams repeat the workflow or no buyer accepts the stated commercial terms.
Inconclusive	Fewer than four teams receive usable data, a connector failure blocks exposure, or the primary measure cannot be observed. Maximum one retest.

Choose the smallest credible test, not the smallest activity

A single friendly customer may be faster but unable to distinguish the product effect from the relationship. A landing page can test message response but not operational use. A large pilot may collect more data while introducing bespoke support that destroys the mechanism being tested. Small means no larger than the decision requires; credible means the result can survive the most obvious competing explanation.

Protect the customer and the business

Experiments do not suspend obligations. Define consent, data handling, service risk, remediation and stop conditions where relevant. Do not expose customers to avoidable harm merely to create a cleaner test. If the necessary guardrail prevents the assumption from being tested directly, record that limitation and choose a safer proxy or specialist review.

Do not confuse a gate with a market verdict

Six teams cannot establish scalable market demand. They can provide decision-grade evidence for whether SignalWorks should release another eight weeks of productisation work. Each stage should name what remains unproven. That keeps a successful test from expanding into a claim the evidence never supported.

Chapter output. A pre-committed Decision Test Card whose measures, threshold, window and guardrails are strong enough for the next staged decision without supporting a stronger claim.

7. Make the Decision Hold

Turn the result into an owned action, preserve dissent and close the paths that the decision actually removes.

A test can be well designed and still fail organisationally. The result enters a meeting, everyone offers an interpretation and the initiative continues unchanged. The Decision Record prevents the evidence cycle from ending as another discussion.

Run the decision review in five moves

Step	Decision review move
1	State the original decision, assumption and pre-committed rule. Do not begin with a narrative of the work performed.
2	Present the observation, including missing data, breaches and contradictory cases.
3	Compare the result with the rule and test the strongest competing explanation.
4	Choose proceed, revise, stop or inconclusive. If the rule is overridden, record who overrode it and why.
5	Name the commitment released, the work paused or closed, the remaining uncertainty and the next review date.

The SignalWorks result

Five of six teams received usable data and activated. Three completed the workflow in at least three of four weeks. Two asked to continue at the stated annual terms. The behavioural and commercial thresholds were met. Onboarding, however, required between nine and fourteen SignalWorks hours for four teams, breaching the six-hour guardrail.

The correct outcome is revise. The workflow appears capable of producing independent repeated use, but the delivery model is not yet inside the approved boundary. SignalWorks releases two weeks of work to redesign ingestion, while the remaining eight weeks stay withheld. One retest is permitted. If onboarding still exceeds six hours, the company must either price and position the offer as a higher-touch service or stop treating it as a scalable product line.

SignalWorks - Decision Record

Field	Record
Decision date and owner	4 September; CEO.
Result	Behavioural and continuation thresholds met; onboarding guardrail breached.
Outcome	Revise and run one bounded retest.
Commitment released	Two weeks of ingestion redesign; remaining six weeks withheld.
What is closed	No broader rollout, new segment entry or additional feature scope before the onboarding gate.
Dissent	Commercial lead preferred immediate rollout to the two willing buyers; concern recorded that delay could reduce momentum.
Remaining uncertainty	Realised customer value and renewal beyond the initial continuation request remain unproven.
Next review	Two weeks after revised onboarding begins.

Closure should match reversibility

A reversible decision need not eliminate every option. It should close enough paths to make the chosen test or commitment real. An irreversible decision needs a higher evidence threshold and explicit acceptance of residual risk. Staged commitments often create the best discipline: they preserve options without pretending that keeping everything active is free.

Dissent belongs in the record

A decision record is stronger when it preserves the best objection. Dissent helps later reviewers distinguish a bad decision from a reasonable decision followed by an unexpected outcome. It also prevents unresolved disagreement from reappearing as selective memory.

Inconclusive is controlled, not comfortable

An inconclusive result does not automatically authorise another test. Record why the test failed, what a credible repetition would cost and how many repetitions are permitted. If the uncertainty is expensive to resolve and the decision is reversible, the owner may proceed while carrying it. If the exposure is high, the same gap may block approval.

Chapter output. A Decision Record that states the result, interpretation, action, dissent, residual uncertainty, commitments released and paths closed.

8. Install the Operating Rhythm

Make evidence review part of portfolio management so learning survives beyond one test and one leadership meeting.

A good record can still disappear into a folder. The operating rhythm connects decisions, tests and portfolio consequences. Its job is not to create more reporting. It is to ensure that material beliefs are updated, commitments are staged and unsupported work does not continue through inertia.

The weekly Evidence Review

Use a forty-five-minute review for active Learn items. Keep Run and Protect in their existing control forums unless the same issue affects the decision. The meeting should be small enough to decide and broad enough to expose cross-functional consequences.

Time	Focus	Output
5 min	Changes in decision deadlines, exposure or guardrails.	Reprioritised review order.
10 min	New observations and contradictions.	Evidence Ledger updated.
15 min	Tests reaching a gate.	Proceed, revise, stop or inconclusive decision.
10 min	Portfolio capacity and new requests.	Work admitted, paused or rejected.
5 min	Owners, commitments and communication.	Decision Records closed and distributed.

Four roles

- Decision owner: holds authority for the commitment and accepts residual uncertainty.
- Test owner: protects the method, observation quality and guardrails.
- Challenger: presents the strongest competing explanation and checks claim inflation.
- Recorder: maintains the Evidence Ledger and ensures the chosen action enters normal operating tools.

One person can hold more than one role in a small company, but the challenger should not be the person whose status depends most heavily on the initiative continuing.

Control portfolio admission

Every new Learn item must identify the decision, critical assumption, owner and capacity it will consume. Set an explicit work-in-progress limit based on the organisation's ability to run credible tests and make timely decisions. There is no universal number. If a new item exceeds the limit, leadership must pause, close or defer another item.

Scale items require an evidence record from the prior stage. When scale changes the population, channel, cost structure or operating mechanism, reopen the affected assumptions. Evidence from one context does not automatically travel to another.

The Evidence Ledger

The ledger is the institutional memory of the method. It links claims to observations and observations to decisions. A spreadsheet, CRM object, Notion database or product-operations tool can hold it. The system matters less than the discipline: one record per material assumption, visible contradictions, dated confidence changes and links to every decision that used the evidence.

Date	Assumption	Observation	Profile change	Decision affected
12 Aug	Users act without facilitation.	Three facilitated pilots completed.	Independent evidence remains weak.	Approve bounded test only.
2 Sep	Users act without facilitation.	Three of five usable teams repeated workflow.	Independent use supported for current gate.	Behaviour threshold met.
2 Sep	Onboarding stays within six hours.	Four teams required 9-14 hours.	Assumption contradicted.	Release redesign only.

Escalate changes in exposure, not every weak signal

A new observation deserves escalation when it changes the decision deadline, threatens a guardrail, materially alters reversibility or contradicts a critical assumption. Routine variation belongs in the ledger until the review gate. This protects leadership attention while preserving evidence that may matter later.

Chapter output. A weekly review with named roles, a Run/Protect/Learn/Scale portfolio, an explicit admission rule and an Evidence Ledger that allows learning to compound.

9. Three Worked Cases

See how the same method changes commercial, product and portfolio decisions without pretending that one signal settles the whole question.

The cases below are hypothetical composites. Their purpose is to show the decision mechanics and evidence boundaries. The numbers are illustrative, not benchmarks.

Case A - Interest was mistaken for willingness to pay

A B2B software company ran twenty discovery calls. Seventeen participants described the problem as important, nine joined a waitlist and four agreed to a free design partnership. Leadership proposed hiring two account executives and funding a full commercial launch.

Commercial case

Field	Record
Initial claim	Positive conversations and signups show sufficient demand to scale sales.
Available evidence	Strong problem language, low-cost interest and four time commitments.
Missing evidence	Budget ownership, willingness to accept standard terms, repeatable buying trigger and sales-cycle length.
Decision	Defer sales hiring. Offer a paid, tightly scoped implementation to six qualified buyers at stated terms.
Result	Two buyers accepted; two wanted the outcome but lacked current budget; two required integrations outside the target offer.
Action	Proceed with one founder-led commercial cycle in the funded segment; do not generalise demand across the full waitlist.
Limitation	Payment supports willingness and budget for two transactions. It does not establish retention, scalable acquisition or realised value.

Case B - Usage growth was produced by prompting

A collaboration product introduced a weekly planning feature. Weekly active use rose by 38 per cent among pilot teams. Product leadership attributed the increase to the feature and prepared a broad rollout. The Evidence Profile showed a serious independence gap: customer-success managers had run weekly sessions and sent personal reminders.

Product case

Field	Record
Initial claim	The feature creates a self-sustaining weekly planning habit.
Available evidence	Repeated use during a facilitated pilot and positive qualitative feedback.
Missing evidence	Behaviour without personal prompting and evidence that planning improved the target team outcome.
Decision test	Remove live facilitation for half the teams, retain automated reminders for all and observe four more weekly cycles.
Result	Use held in teams with an existing planning owner and fell sharply elsewhere.
Action	Revise. Position the feature for teams with an established planning role and redesign onboarding for teams without one.
Limitation	The test identifies a boundary condition. It does not prove that the role alone caused persistence or that usage produced business value.

Case C - Too much work prevented a decisive test

A sixty-person technology company had fourteen “strategic” initiatives. Each had a sponsor, and none consumed enough capacity to appear individually dangerous. Together they fragmented product, commercial and data teams. Leadership could not run a credible test of its new enterprise proposition because the same people were supporting three migrations, two partnerships and several feature commitments.

Operating case

Field	Record
Initial claim	All fourteen initiatives are necessary to preserve momentum and optionality.
Portfolio review	Three Run items, four Protect obligations, four Learn items and three proposed Scale items.
Evidence gap	Only one Scale item had an evidence record from a prior stage. Three Learn items had no named decision or test window.
Decision	Protect the seven Run/Protect items, admit two Learn items, continue one supported Scale item and pause the remaining four.
Result	The enterprise team completed a six-week decision test; one paused partnership later became unnecessary when the target segment changed.
Action	Keep the admission limit and require a Decision Brief for every new Learn or Scale item.

Field	Record
Limitation	Reduced work in progress created capacity; it does not prove that fewer initiatives always improve performance.

What the cases have in common

The method did not reward pessimism. Two cases proceeded in a narrower form and one changed the portfolio so a real test became possible. The common move was to limit each conclusion to what the evidence could support, then match the next commitment to the remaining uncertainty.

First exercise. Take one initiative expected to matter within ninety days. Write the decision. Map the causal chain. Circle the assumption with the highest consequence and weakest evidence. If the next leadership meeting cannot name the observation that would change the decision, the initiative is not ready for approval.

Toolkit

Use these seven records as a chain. Keep them short. Link to supporting material instead of turning the record into a presentation. Every completed record should make the next decision easier to inspect. The worksheets are designed for printing or manual recreation in your operating system; this edition is not a fillable PDF.

Tool 1 - Decision Brief

Use before approving a material Learn or Scale initiative.

Tool 1 - Decision Brief - blank record

Field	Record
Decision	
Owner and contributors	
Decision deadline	
Commitment requested	
Consequence if wrong	
Cost of delay	
Reversibility and cost of reversal	
Evidence threshold for this stage	
Next review date	

Tool 2 - Causal Chain

Show how the intervention is expected to create behaviour, outcome and value inside a defined boundary.

Tool 2 - Causal Chain - blank record

Field	Record
Decision	
Intervention	
Mechanism	
Expected behaviour	
Operational/commercial outcome	
Realised value	
Population and context	
Time window	
Primary measure	
Competing explanations	

Tool 3 - Assumption Register

Maintain one record per critical assumption; link observations from the Evidence Ledger.

Tool 3 - Assumption Register - blank record

Field	Record
Assumption	
Why it matters to the decision	
Consequence if wrong	
Current confidence	
Supporting evidence	
Contradicting evidence	
Evidence gap	
Owner	
Review/expiry date	

Tool 4 - Evidence Profile

Describe the shape and boundary of the evidence. Do not collapse the dimensions into one score.

Tool 4 - Evidence Profile - blank record

Field	Record
Claim being assessed	
Observation	
Directness	
Cost/commitment	
Repetition	
Representativeness	
Independence	
Proximity to value	
Contradictions and limitations	
Decision implication	

Tool 5 - Decision Test Card

Pre-commit the smallest credible test capable of changing the next action.

Tool 5 - Decision Test Card - blank record

Field	Record
Decision and assumption	
Population	
Intervention	
Comparison/baseline	
Primary measure	
Threshold	
Time window	
Guardrails	
Owner	
Proceed rule	
Revise rule	
Stop rule	
Inconclusive rule and retest limit	

Tool 6 - Decision Record

Close every evidence gate with an owned action and a record of residual uncertainty.

Tool 6 - Decision Record - blank record

Field	Record
Decision date and owner	
Original rule	
Observed result	
Competing explanation considered	
Outcome: proceed/revise/stop/inconclusive	
Commitment released	
Work paused or closed	
Dissent	
Residual uncertainty	
Next review	

Tool 7 - Evidence Ledger

Use one row for each material observation or confidence change. Link the supporting source in the live system.

Date	Assumption / claim	Observation	Source	Profile change	Decision affected	Owner

Facilitator checklist

- Is the decision written as a choice rather than an aspiration?
- Does the owner have authority to close it?
- Is the evidence threshold proportionate to reversibility and exposure?
- Does the causal chain include a mechanism and measurement boundary?
- Is the selected assumption both consequential and weakly supported?
- Can the test distinguish the strongest competing explanation?
- Are proceed, revise, stop and inconclusive rules recorded before launch?
- Are safety, legal, service and cost guardrails explicit?
- Will the result change capacity, commitment or scope?
- Has dissent and residual uncertainty been preserved?

Evidence boundaries and source basis

What this guide claims

The guide offers a practical operating method for making assumptions, evidence and decisions visible. It can improve the quality and traceability of a decision process. It cannot guarantee a correct outcome, establish scientific causality from weak operational data or replace domain-specific professional judgement.

The SignalWorks narrative and the three worked cases are hypothetical composites. Their numbers illustrate the method and should not be used as benchmarks. The evidence dimensions, record sequence and Run/Protect/Learn/Scale portfolio are Green Bear's synthesis of established ideas in experimentation, causal reasoning, reversible decisions, evidence-based management and organisational decision practice.

Source basis

1. Scrum.org - Evidence-Based Management Guide (2024). A practitioner guide to improving value delivery under uncertainty through goals, measures and experiment loops.
2. Amazon - 2016 Letter to Shareholders. Source for the distinction between reversible two-way-door decisions and more consequential one-way-door decisions.
3. Hernán and Robins - Causal Inference: What If. A rigorous reference for causal questions, competing explanations and the limits of observational inference.
4. Staw - "Knee-deep in the Big Muddy" (1976). Foundational experimental work on escalation of commitment after negative consequences.
5. The Lean Startup - Principles. A practitioner source for validated learning and iterative testing under uncertainty.

Use boundary

Apply higher standards where decisions affect health, safety, employment, regulated conduct, security, privacy, financial reporting or material contractual commitments. This guide can structure the question and evidence record; it does not substitute for legal advice, statistical design, security assurance, financial audit or engineering validation.

About the author

Jörn Green

Technology and operations leader · Green Bear Consulting

Jörn Green is a technology and operations executive with more than 25 years of experience across software, gaming, telecommunications and IoT. He has led product and technology organisations in Europe and Asia, with responsibility for strategy, delivery, commercial outcomes and organisational scale.

At Sony, he led mobile software and firmware-over-the-air operations reaching approximately 100 million devices, built an organisation in China from one person to around 200, and took an IoT business from concept to its first automotive customer and more than EUR 2 million in annual revenue. At Ubisoft, he scaled Ubisoft Connect from approximately 40 to 240 people while its user base grew from 40 million to 60 million. At Play'n GO, he led a 300-350-person games organisation with more than EUR 35 million in P&L responsibility and increased annual output from 52 to 64 games, a 22% improvement.

Through Green Bear Consulting, Jörn works with technology leadership teams when growth, complexity or uncertainty has outpaced the way the organisation operates. His work focuses on decision rights, operating discipline, commercial evidence and leadership structures that convert strategy into coordinated execution.

Start with one stuck decision. The free 45-minute Green Bear Decision Diagnostic clarifies the decision, exposes the most important evidence gap and identifies the next useful step. It is an initial diagnostic conversation, not a completed set of consulting records.

Start here: Free Decision Diagnostic

Read more: jorngreen.online

Connect: Jörn Green on LinkedIn