Title

DIAGNOSTIC TAXONOMY OF UNCERTAINTY IN SUSTAINABILITY METRICS. Appendix to Delmas et al., 2026. Improving the credibility of corporate sustainability Metrics. Academy of Management Perspectives.

Author
Delmas, Magali

Publication Date
2026-03-12

APPENDIX: DIAGNOSTIC TAXONOMY OF UNCERTAINTY IN SUSTAINABILITY METRICS

This appendix provides a diagnostic framework for identifying and assessing the two foundational forms of uncertainty that undermine the credibility of sustainability metrics: effect uncertainty and measurement uncertainty. While the main text explains how these uncertainties affect metric selection and credibility, this appendix equips practitioners with systematic tools to diagnose specific manifestations, trace them to their root causes, and assess their severity within their own organizations.

Cite as: Delmas, M. A., Clark, K., Minor, D., & Timmer, T. (2026). *Appendix: Diagnostic taxonomy of uncertainty in sustainability metrics. Supplemental material to "Improving the credibility of corporate sustainability metrics."* Academy of Management Perspectives.

A1. Understanding Effect Uncertainty

Effect uncertainty concerns whether a sustainability initiative will produce its intended outcome. Unlike measurement problems that affect how we quantify results, effect uncertainty reflects fundamental ambiguity about the causal relationship between actions and outcomes. Even when measurement is precise, effect uncertainty means we cannot confidently predict whether our intervention will succeed, fail, or produce unintended consequences.

Three Dimensions of Effect Uncertainty

Effect uncertainty manifests across three interrelated dimensions that shape both the predictability and interpretability of sustainability outcomes:

Valence (Direction of Impact) refers to whether an intervention produces positive effects, negative effects, or no discernible change. The same action can yield benefits in one context but cause harm or prove ineffective in another.

Diagnostic question: Could this initiative reasonably produce outcomes opposite to our intentions, or have null effects in some contexts?

Example: A supplier engagement program designed to improve environmental practices might successfully upgrade capabilities at large suppliers while inadvertently pushing smaller suppliers out of the supply chain due to increased compliance costs, creating both positive and negative externalities simultaneously.

Magnitude (Scale of Impact) captures the range of possible effect sizes, where identical interventions can yield anywhere from trivial to transformative results depending on context.

Diagnostic question: If we implemented this initiative across ten different sites or business units, would we expect similar effect sizes, or could results vary by a factor of 5x or 10x?

Example: Two facilities implementing identical waste-reduction protocols might experience a 5% reduction in one location and a 40% reduction in another due to differences in baseline waste levels, operational complexity, workforce engagement, or pre-existing informal practices.

Timing (Temporal Dynamics) concerns when effects materialize—immediately, after delays, or unpredictably—and whether they persist, diminish, or reverse over time.

Diagnostic question: Can we confidently predict when measurable outcomes will appear, or might effects emerge across widely different timeframes (months vs. years) depending on conditions we don't fully control?

Example: A regenerative agriculture program might show measurable soil health improvements within two years at one farm but require five to seven years at another due to differences in soil type, climate patterns, prior land management practices, and farmer adoption behavior.

Contextual Sources of Effect Uncertainty

System Complexity and Interdependence. Sustainability outcomes typically result from multiple interacting factors in various configurations, with different causal pathways leading to similar results (equifinality) and identical actions producing different results across contexts (multifinality). This configurational causality makes it difficult to isolate which factors drive success.

Diagnostic indicators:

- The initiative's success depends on alignment of 5+ interdependent factors
- Causal pathways involve feedback loops or nonlinear relationships
- Similar past initiatives produced highly variable results across contexts
- Outcome depends on factors outside the organization's direct control

Example: A water stewardship initiative's success depends on the interaction of technological interventions, stakeholder engagement quality, regulatory compliance, hydrological conditions, agricultural practices in the watershed, and climate patterns—making it nearly impossible to attribute measured water savings definitively to any single intervention.

External Volatility and Shocks. Unforeseen external changes—market shifts, regulatory reforms, technological disruptions, or societal events—can dramatically alter outcomes independent of the initiative's inherent effectiveness.

Diagnostic indicators:

- Initiative operates in domains subject to rapid regulatory change
- Outcomes are sensitive to commodity prices, exchange rates, or market conditions
- External stakeholder behavior (competitors, suppliers, customers) significantly influences results
- Recent precedents exist of similar initiatives derailed by external shocks

Example: Dramatic carbon emissions reductions during COVID-19 lockdowns stemmed from government mandates and economic disruption, not corporate sustainability programs, yet many companies' emissions targets appeared to be met—creating attribution confusion about which reductions reflected genuine program effectiveness versus external circumstances.

Temporal Delays Between Action and Outcome. Extended timeframes between implementation and measurable results allow numerous intervening variables to emerge, making it progressively harder to attribute outcomes to specific initial actions.

Diagnostic indicators:

- Expected time to observable outcomes exceeds 3-5 years
- Multiple organizational changes (leadership, strategy, structure) likely before outcomes materialize
- Technology, regulations, or market conditions may fundamentally shift during the intervention period
- No reliable interim indicators exist to validate the causal pathway

Example: Net-zero commitments targeting 2050 involve 25-30 year horizons during which technologies, regulations, markets, and organizational capabilities will transform in ways that cannot be reliably forecast, creating what scholars call "future washing"—using uncertainty about the future to make pledges that may never materialize or whose fulfillment cannot be credibly attributed to current actions.

Behavioral Sources of Effect Uncertainty

Cognitive Biases in Design and Implementation. Mental shortcuts and perceptual distortions systematically skew how managers and employees interpret information, design interventions, and assess results. Confirmation bias, optimism bias, anchoring effects, and availability heuristics can obscure genuine process-outcome relationships.

Diagnostic indicators:

- Initiative design relied heavily on success stories from other organizations without rigorous analysis of contextual differences
- Teams interpret ambiguous results as validating their approach rather than questioning assumptions
- Negative feedback or contrary evidence is dismissed as "implementation issues" rather than design flaws
- Post-hoc explanations for setbacks emphasize external factors while successes are attributed to the initiative

Example: A sustainability team committed to a particular supplier engagement model might selectively focus on the 30% of suppliers showing improvement while discounting the 70% showing no change as "not yet ready" rather than questioning whether the intervention model itself is effective.

Organizational Resistance and Implementation Gaps. Even well-designed initiatives fail when they lack authentic buy-in from leadership and employees. Resistance manifests as passive non-compliance, resource diversion, competing priorities, or active undermining.

Diagnostic indicators:

- Initiative lacks visible, consistent executive sponsorship
- Middle managers view the initiative as "extra work" rather than core business
- Resource allocation (budgets, staff time, technology) remains inadequate despite stated commitments
- Implementation varies widely across business units without clear strategic rationale
- Employee surveys show low awareness or skepticism about the initiative's importance

Example: A corporate renewable energy mandate might be systematically undermined by facility managers who view it as threatening operational flexibility and budget control, leading to delayed procurement, minimal-compliance approaches, or cherry-picking only the most economically favorable projects—ensuring weak overall results that appear to validate initial skepticism.

Internal Political Dynamics. Conflicting interests, departmental silos, and competition over resources can prevent coherent execution of sustainability initiatives, as individuals or units protect their interests at the expense of collective goals.

Diagnostic indicators:

- Key departments (operations, finance, procurement) not involved in initiative design
- Success metrics conflict with existing departmental KPIs or incentives
- Resource allocation decisions favor traditional business priorities over sustainability
- Cross-functional coordination relies on goodwill rather than formal governance structures
- Initiative "ownership" is unclear or contested among organizational units

Example: A circular economy initiative requiring product redesign might face resistance from engineering teams measured on time-to-market and cost minimization, from procurement teams with supplier contracts incompatible with circular materials, and from finance teams concerned about upfront investments—resulting in fragmented implementation that undermines the initiative's systemic logic.

A2. Understanding Measurement Uncertainty

Measurement uncertainty arises when the quantification of sustainability actions or outcomes is imprecise, unstable across measurement approaches, or dependent on contested assumptions. Unlike effect uncertainty, which concerns whether our actions work, measurement uncertainty concerns whether we can accurately determine what happened. Even when causal relationships are clear, measurement uncertainty can make it impossible to credibly report on outcomes.

Two Primary Manifestations

Variability (Random Error) reflects the extent to which repeated measurements of the same phenomenon yield inconsistent results. High variability signals that measurements are unreliable —influenced by measurement conditions, data collection procedures, or instrument precision rather than genuine differences in the underlying phenomenon.

Diagnostic question: If three independent teams measured this outcome using the same stated methodology, how much would their results differ?

Example: A company measuring water consumption across facilities might obtain results varying by 15-30% depending on meter calibration schedules, whether measurements include cooling tower evaporation, how consumption during equipment maintenance is handled, and whether data collection happens daily versus weekly.

Bias (Systematic Error) occurs when measurement approaches consistently skew results in a particular direction, either through flawed methodologies that structurally favor certain outcomes or through organizational routines that systematically overlook certain impacts.

Diagnostic question: Does our measurement approach have built-in assumptions or structural features that would consistently over- or under-estimate the true value?

Example: Scope 3 emissions calculations using industry-average emission factors systematically underestimate actual supply chain impacts for companies sourcing from emissions-intensive suppliers while overestimating impacts for companies with cleaner-than-average suppliers—creating systematic bias that persists across reporting periods.

Contextual Sources of Measurement Uncertainty

Lack of Standardization. The proliferation of competing frameworks, methodologies, and certification schemes without clear dominance or interoperability enables organizations to select approaches that present their performance most favorably while claiming technical compliance.

Diagnostic indicators:

- Multiple accepted methodologies exist with no clear "gold standard"
- Different methodologies can yield 20%+ differences in reported values
- Competitors in the same industry report using different frameworks
- Framework selection is driven by favorable results rather than technical appropriateness
- The organization has changed reporting frameworks across periods without clear technical justification

Example: Carbon accounting standards differ substantially in how they treat biogenic emissions, renewable energy certificates, market-based versus location-based reporting, and supply chain boundaries—enabling companies to report dramatically different emissions profiles for identical operations depending on methodology choices.

Data Quality Gaps. Measurement credibility requires accurate data (reflecting true values within known margins of error) and reliable data (producing consistent results across repeated measurements). Gaps in either dimension introduce uncertainty.

Diagnostic indicators for accuracy:

- Reliance on estimates, industry averages, or modeled values rather than direct measurement
- Use of outdated conversion factors or emission factors (>3 years old)
- Measurement instruments not calibrated to recognized standards
- No documentation of expected measurement error or confidence intervals
- Significant data gaps filled through interpolation or proxy variables

Diagnostic indicators for reliability:

- No documented, standardized data collection protocols
- High turnover in teams responsible for data collection
- Different business units use different measurement approaches for the same variable
- Historical data shows unexplained volatility or discontinuities

- No regular audit or validation of measurement processes

Example: A retailer reporting Scope 3 emissions from purchased goods using financial spend data and industry-average emission factors faces both accuracy issues (emission factors may not reflect actual supplier practices) and reliability issues (small changes in categorization of spend can create large swings in reported emissions).

Boundary and Allocation Ambiguities. Many sustainability metrics require defining system boundaries (what's included/excluded) and allocating impacts across complex processes—decisions that involve judgment and can dramatically affect results.

Diagnostic indicators:

- Boundary definitions rely on materiality thresholds that involve subjective judgment
- Allocation methodologies (economic vs. physical vs. causal) would yield substantially different results
- Boundary definitions have changed across reporting periods
- Inconsistent treatment of joint ventures, franchises, or outsourced operations
- No sensitivity analysis showing how results change with alternative boundary choices

Example: A food company's water footprint varies by an order of magnitude depending on whether it includes only direct operations, extends to tier-1 suppliers, includes agricultural water use, or attempts to capture full lifecycle impacts—and no single boundary is unambiguously "correct."

Temporal Lags and Reporting Delays. Timeliness in reporting is essential for drawing accurate causal inferences and enabling timely decisions. Delays between data collection and disclosure can render metrics obsolete, especially in rapidly changing domains.

Diagnostic indicators:

- Sustainability data reported 12+ months after the period it describes
- Significant operational changes occurred between measurement and reporting
- Data collection cycles (annual) don't align with decision cycles (quarterly/monthly)
- Historical revisions are common and substantial (>10% changes to previously reported values)

Example: A sustainability report covering operations from 18 months prior may no longer reflect current reality if the company has since divested business units, changed suppliers, implemented new technologies, or experienced regulatory changes—making the data more historical artifact than decision-support tool.

Behavioral Sources of Measurement Uncertainty

Strategic Selection and Disclosure Bias. When sustainability metrics are tied to incentives, reputational stakes, or competitive positioning, teams face pressures to emphasize favorable metrics while downplaying or omitting less desirable information.

Diagnostic indicators:

- Reporting emphasizes metrics showing improvement while omitting metrics showing deterioration
- Denominators or baselines change across periods in ways that favor reported trends
- Positive results highlighted in absolute terms, negative results presented in normalized or contextual terms
- Voluntary disclosures become less detailed when performance weakens
- Peer comparisons use metrics where the company performs well while avoiding metrics where it lags

Example: A company might highlight impressive absolute emissions reductions driven by divesting emissions-intensive assets while avoiding discussion of increasing emissions intensity in remaining operations—technically accurate but systematically biased toward favorable interpretation.

Obfuscation and Complexity. The use of technical jargon, complex data presentations, changing denominators, or selectively reported baselines can obscure genuine performance trends even when reported numbers are technically accurate.

Diagnostic indicators:

- Metrics require specialist knowledge to interpret
- Presentation formats emphasize favorable comparisons (favorable baseline periods, peer groups, or normalizations)
- Methodological changes coincide with performance inflection points
- Explanatory narratives are vague about methodologies, assumptions, or limitations
- Graphical presentations use truncated axes, logarithmic scales, or selective time periods

Example: Reporting water efficiency improvements using revenue-normalized metrics during a period of significant price increases can show improving efficiency even when absolute water consumption increased—a presentation choice that obscures rather than illuminates environmental impact.

Organizational Culture and Motivated Reasoning. Teams genuinely committed to sustainability initiatives may unconsciously overestimate impacts or underestimate implementation challenges through cognitive biases rooted in shared mental models and organizational culture.

Diagnostic indicators:

- Consistent pattern of actual results falling short of projected impacts
- Favorable assumptions accepted readily while skepticism applied to cautionary data
- Post-hoc explanations for measurement challenges emphasize external factors
- Reluctance to invest in measurement improvement that might reveal lower impacts
- Internal conversations systematically more optimistic than external expert assessments

Example: A sustainability team deeply invested in a regenerative agriculture program might consistently accept optimistic assumptions about soil carbon sequestration rates, discount measurement challenges as temporary, and interpret ambiguous data as validating their approach—not through deliberate misrepresentation but through unconscious biases that influence which measurement approaches and assumptions seem most reasonable.

A3. Diagnostic Process

Managers can diagnose uncertainty through a two-stage process. Stage 1 involves initial screening using six simple tests: for effect uncertainty, assess whether the causal pathway is clear and evidence-based, whether the initiative would work consistently across different contexts, and whether effects can be isolated from confounding factors; for measurement uncertainty, determine whether quantification relies on direct observation versus modeling or estimates, whether repeated measurements would vary by more than 10-30%, and whether results are sensitive to methodological choices. When initial screening suggests moderate to high uncertainty or when reputational stakes are significant, Stage 2 employs deeper assessment methods including multi-stakeholder workshops to map initiatives onto the uncertainty matrix, expert elicitation protocols to document confidence levels and plausible outcome ranges, quantitative variability analysis where historical data permits, and theory-of-change mapping to explicitly document causal pathways and assess evidence strength for each link.

A3. Summary Table: Uncertainty Taxonomy

The following table distills the diagnostic framework into a quick reference guide that enables managers to identify uncertainty types, recognize warning signs, and determine appropriate next steps.

Uncertainty Type	Core Question	Key Manifestations	Primary Sources	Diagnostic Red Flags
Effect Uncertainty	Will the initiative work as intended?	• Valence (direction) • Magnitude (scale) • Timing (when)	**Contextual:** • System complexity • External volatility • Temporal delays **Behavioral:** • Cognitive biases • Resistance • Internal politics	• Similar initiatives show highly variable results • Causal pathway unclear or contested • Outcomes depend on many uncontrolled factors • Long delay to observable results
Measurement Uncertainty	Can we accurately quantify what happened?	• Variability (inconsistency) • Bias (systematic error)	**Contextual:** • Lack of standardization • Data quality gaps • Boundary ambiguities • Temporal lags **Behavioral:** • Selection bias • Obfuscation • Motivated reasoning	• Results vary >20% across measurement approaches • Heavy reliance on estimates vs. direct measurement • Contested boundaries or allocation rules • Metrics tied to high-stakes incentives

This diagnostic framework enables managers to systematically assess uncertainty in their sustainability initiatives, trace it to specific sources, and make informed decisions about metric selection, uncertainty reduction strategies, and disclosure approaches aligned with the credibility-centered roadmap presented in the main text.