The ASA Measurement of Effect Protocol

Measure first. Estimate second. Grade each.

Treadstone 71 LLC · East 71st Tradecraft Publishing

Version 1.0 · Draft for comment 0.1 · November 2, 2026

Field	Entry
Document	The ASA Measurement of Effect Protocol (ASA-MOE), Version 1.0, draft for comment 0.1
Issued	November 2, 2026
Comment window	November 2 to December 2, 2026 (30 days)
Version 1.0 final	January 2027, with a published disposition register for every comment received
Extends	The ASA Evidence Standard, Public Edition 1.2, Treadstone 71 LLC, October 4, 2026. DOI 10.5281/zenodo.23137881
Cite as	Treadstone 71 LLC. *The ASA Measurement of Effect Protocol*, Version 1.0, draft for comment 0.1. Half Moon Bay, CA: Treadstone 71 LLC, November 2, 2026. https://www.treadstone71.com/asa-measurement-of-effect. DOI assigned at Zenodo deposit.
License	Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), with the internal-use grant in Section 21
Supersedes	Working draft 0.3 (September 8, 2026) and the two working drafts numbered 0.1 (September 2026). The change record sits in Annex I.

Table 0. Document control.

Treadstone 71

Contents

Tearlines 7

Bottom Line [Tearline 1] 7

Part I · The Problem and the Evidence 8

1. The problem the protocol answers [Tearline 1] 8

Failure 1. Reach reported as effect 8

Failure 2. No baseline and no comparison 9

Failure 3. The operator grades the operator 9

Failure 4. The method stays inside the vendor 9

2. Comparable methods: what exists, and where each falls short [Tearline 1] 9

2.1 How the review ran 9

2.2 The comparables matrix 10

2.3 What the matrix shows 12

2.4 Nine gaps, ranked by consequence 13

2.5 Keep, extend, build new 14

3. The T71 Standard inside the protocol [Tearline 2] 15

4. Definitions [Tearline 2] 16

Part II · Measurement 18

5. Component 1 · Pre-registration [Tearline 3] 18

5.1 What the registration fixes 19

5.2 Timestamp, hash, escrow 19

5.3 Amendments, stopping, and the deviation log 20

5.4 Sealed registrations 20

5.5 Registration discipline from the Advanced SATs 20

6. Component 2 · Comparison design [Tearline 2] 21

6.1 Choosing a design 21

6.2 National survey series as donor pools 22

6.3 Randomization that qualifies 22

6.4 Contamination 22

6.5 What lowers a grade after the fact 23

6.6 What the grade does not decide 24

7. Component 3 · Attitudinal instruments [Tearline 3] 24

7.1 What the instrument measures 24

7.2 Validated or published 24
7.3 Fielding 24
7.4 Sample size and the minimum detectable effect 24
7.5 Attrition 25
7.6 Respondent protection 25
8. Component 4 · Behavioral and network indicators [Tearline 3] 27
8.1 Indicator families 27
8.2 What qualifies 28
8.3 The indicator register 28
8.4 Windows, leakage, and interested feeds 28
8.5 Bot traffic and adversary shaping 28
8.6 Indicators as primary measures 28
9. Component 5 · Context: population profiles and adversary effort [Tearline 3] 28
9.1 Population profiles 29
9.2 Adversary effort series 29
9.3 Adversary profile 29
9.4 Indicators of Change scoring for signposts 29
10. Component 6 · The grading rubric [Tearline 2] 29
10.1 Grading the inputs 30
10.2 Grading the effect claim 30
10.3 Two ceilings 31
10.4 The combination rule 31
10.5 Analysis of Competing Hypotheses on the causal claim 32
Part III · Estimation 33
11. Component 7 · The estimate [Tearline 2] 33
11.1 Four estimate questions 33
11.2 Estimate rules 33
11.3 The estimate names an outcome 34
11.4 Transport and the profile match 34
11.5 Signposts 35
11.6 Persistence arithmetic 36
11.7 Scale 36
11.8 Adversary response 36

11.9 The Estimate Statement 37
11.10 The calibration ledger 37
Part IV · Reporting, Marks, and Audit 38
12. The Effect Statement and the Estimate Statement [Tearline 2] 38
12.1 Effect Statement fields 39
12.2 Rules the statements obey 39
12.3 Cost per unit of effect 40
13. Publication and restricted publication [Tearline 1] 40
14. Conformance marks [Tearline 1] 41
15. Independent audit [Tearline 1] 41
15.1 Five checks 42
15.2 Findings and severity 42
15.3 Determination, conflict, and terms 42
16. Worked example [Tearline 2] 43
16.1 The campaign and the decision 43
16.2 Registration 43
16.3 Execution 43
16.4 Measurement and grading 44
16.5 Estimates 45
16.6 The same campaign, measured the usual way 46
17. Contract language [Tearline 1] 46
18. Linkage to the Adversary Index [Tearline 2] 47
Part V · Validity, Limits, and Use 50
19. Validation program and limits [Tearline 4] 50
20. Relationship to evaluation practice, and the sources [Tearline 4] 51
20.1 Law, regulation, and official standards 51
20.2 Military doctrine 52
20.3 Reporting and professional standards 53
20.4 Peer-reviewed method 53
20.5 Peer-reviewed evidence on corrections, prebunking, and research ethics 55
20.6 Research institute and international organization reports 55
20.7 National programs read at the issuing government 55
20.8 Registries 56

21. Use, citation, and license [Tearline 4] 57
22. Comments and requests [Tearline 1] 57
Annexes 57
Annex A · Registration Form [Tearline 3] 57
Annex B · Indicator Register [Tearline 3] 59
Annex C · Effect Statement and Estimate Statement [Tearline 3] 60
Annex D · Data-room checklist [Tearline 3] 62
Annex E · Audit Statement [Tearline 3] 63
Annex F · Signpost Register [Tearline 3] 63
Annex G · Calibration ledger [Tearline 3] 64
Annex H · Probability language [Tearline 3] 64
Annex I · Change record and verification notes [Tearline 4] 65

Tearlines

The protocol has four readers. Each section heading carries its tearline.

Tearline	Reader	Sections
Tearline 1, Strategic	Funders, ministers, oversight committees, contracting officers	Bottom Line; 1; 2; 13; 14; 15; 17; 22
Tearline 2, Operational	Evaluation leads, StratCom desk leads, program managers	3; 4; 6; 10; 11; 12; 16; 18
Tearline 3, Tactical	Analysts, survey managers, data handlers, campaign operators running the measurement	5; 7; 8; 9; Annexes A to H
Tearline 4, Technical	Methodologists, review boards, ethics committees, counsel	19; 20; 21; Annex I

Table 1. Tearline map.

Bottom Line [Tearline 1]

Who. Governments, ministries, foundations, and the vendors they pay to counter hostile information operations. Treadstone 71 publishes the method and audits against it. The firm does not run the campaigns it audits.

What. ASA-MOE is an open protocol in two layers. The measurement layer establishes, with a graded design and graded sources, what a counter-influence campaign changed in what a defined population believed or did. The estimation layer then projects from that measurement to the questions a funder asks next: whether the effect lasts, whether it travels to a new population, whether it scales, and how the adversary answers. No estimate prints without a measurement under it, and no estimate prints above the band its measurement earned.

So what. Counter-disinformation spending has outrun the evidence that it works. A Home Office-funded study published in May 2026 opens with the plain finding that “there is currently no established or agreed upon framework for measuring the impacts of foreign state information, influence and interference threats” (Innes, Orpen and Ashford, CREST, 2026). RAND, in a study for the Office of the Secretary of Defense, found “limited efforts to document the impact of individual programs on Ukrainian attitudes and behavior” and quoted a practitioner: “The only measurable [indicator of success] is reach” (Helmus and Holynska, RAND, 2024). Of the 24 frameworks, doctrines, and national programs reviewed in Section 2, none built for counter-influence work does all four of the following: require a comparison population, grade its data sources, check the measurement for adversary shaping, and bound the causal claim with a probability band.

Why now. Russia, China, and Iran run influence operations at industrial volume. The European External Action Service logged 540 foreign information manipulation and interference incidents in 2025 (EEAS, 4th Report on FIMI Threats, March 2026). The US Global Engagement Center, the main US civilian counter-disinformation office, closed on December 23, 2024, and a search of its archived pages turned up no published monitoring and evaluation framework. Budgets that show a graded effect are likely (55 to <85) to survive the next review cycle better than budgets that show reach.

Impacts so far. Working draft 0.3 circulated inside Treadstone 71 on September 8, 2026. No campaign has yet been measured under the protocol. The design grades, the rubric, the Effect Statement, and the IO Effectiveness Audit have run only on the synthetic campaign in Section 16.

Outlook. Funders will likely (55 to <85) begin writing measurement clauses into counter-influence contracts within two budget cycles, because oversight bodies in the United Kingdom, the European Union, and Ukraine now ask for impact evidence that vendors cannot supply. Whether ASA-MOE becomes the clause they write depends on the validation results in Section 19. Until those results exist, the protocol is a documented method and not a validated instrument.

Recommendations. Funders should require MOE-Registered status before launch and MOE-Audited status before final payment (Section 17). Operators should register a comparison design of DS-3 or stronger. Oversight committees should read the two statements side by side and refuse any estimate that has no measurement under it.

Gaps. Two gaps remain open in the draft. No inter-rater data on design grades exists yet. The estimation layer has no resolved estimates to calibrate against. Section 19 states the work plan for each. Current US joint planning doctrine (JP 5-0, 2025) sits behind controlled access, so the protocol quotes the 2020 edition; JP 3-04 (2022) was read in full from a held copy and is cited by page without a link.

Part I · The Problem and the Evidence

1. The problem the protocol answers [Tearline 1]

A ministry funds a counter-narrative campaign. Six weeks later the vendor delivers a report. The report counts impressions, shares, and comments. A chart climbs. The vendor judges its own work a success, and the ministry renews the contract.

Nobody in that chain measured what changed in what people believed.

Doctrine has named the gap for years.

US joint doctrine separates a measure of performance, “an indicator used to measure a friendly action that is tied to measuring task accomplishment,” from a measure of effectiveness, “an indicator used to measure a current system state, with change indicated by comparing multiple observations over time” (JP 5-0, *Joint Planning*, 2020, glossary). NATO’s agreed definition ties the measure of effectiveness to “the attainment of an end state, achievement of an objective, or creation of an effect” (AJP-5, Edition B, 2026, Lex-8). Civilian evaluation law draws the same line. The US Government Accountability Office calls performance measurement “ongoing monitoring and reporting of a program’s accomplishments” and reserves impact evaluation for estimating “what would have happened in the absence of the program” (GAO-21-404SP).

Practice lags the definitions. Badly.

Four failures recur, and each one is a failure of measurement before it is a failure of the campaign.

Failure 1. Reach reported as effect

Impressions, views, and engagement count exposure. They say nothing about belief. Ukraine’s Center for Countering Disinformation, a working body of the National Security and Defense Council, publishes careful network analysis, yet its headline figures describe the adversary’s reach: more than 85,400 publications

on mobilization and 419.7 million views from October through December 2025 (CCD analytical report). Those numbers measure the adversary's effort and exposure. Neither one measures effect, the adversary's or the defender's.

Failure 2. No baseline and no comparison

Doctrine requires the baseline.

"Before changes in the OE can be observed, a baseline or initial assessment is required," JP 5-0 states, and its assessment appendix tells planners to record "correlations versus causation" among the limits of their indicators. The NATO CIMIC Centre of Excellence handbook admits the harder half of the problem: "One of the biggest challenges with MoEs is the difficulty in isolating variables and establishing a direct cause-and-effect relationship" (CIMIC handbook, 8.4.15).

The 2022 joint doctrine on information goes further and concedes the ground. "Distinguishing between correlation and causation makes information activity assessment difficult," JP 3-04 states, and then reports that "correlation between indicators and events has proven more accurate than efforts to establish concrete cause and effects relationships" (JP 3-04, *Information in Joint Operations*, September 14, 2022, page VI-2; limited release, cited from a copy held by Treadstone 71, no public link). The same page lists "Lack of universal measures and indicators" among nine factors that complicate assessment. The Marine Corps puts the baseline problem in one sentence: "Without a baseline it is difficult to determine whether an information activity has had its desired impact" (MCWP 8-10, *Information in Marine Corps Operations*, February 29, 2024, page 5-1, approved for public release).

A baseline without a comparison population cannot separate the campaign from the news cycle, an election, or the adversary's next move.

Failure 3. The operator grades the operator

The vendor runs the campaign, picks the metric after seeing the data, and reports to the funder. Under Rule 1 of the ASA Evidence Standard, a source speaking in its own cause grades no higher than C, fairly reliable. A vendor grading its own campaign is that source.

Failure 4. The method stays inside the vendor

Proprietary methods stay locked inside the vendor. The funder never checks one. No committee audits one. Nobody hands one to the next vendor. Without reproduction, a result is an opinion with a chart.

The protocol does not make a campaign work. Measurement is not rescue.

The protocol turns "did it work" into a graded, dated, reproducible statement, and then turns "will it keep working" into a graded estimate that stands or falls on registered signposts. A failed campaign gets the same treatment as a successful one.

2. Comparable methods: what exists, and where each falls short [Tearline 1]

2.1 How the review ran

Treadstone 71 reviewed 24 doctrines, standards, research frameworks, and national programs that address measuring the effect of communication or counter-influence work. Each source was read at its originator: the issuing government, standards body, research institute, or journal. Where only a secondary

account existed, the source was excluded or labeled. Each comparable was tested against nine questions drawn from the four failures above and from the ASA Evidence Standard.

Test	Question
T1	Does it separate effect from output and reach?
T2	Does it require a measured baseline before launch?
T3	Does it require a comparison population?
T4	Does it fix the design before launch (registration)?
T5	Does it grade each data source for reliability and credibility?
T6	Does it require independent verification of the result?
T7	Does it bound the causal claim with stated probability language?
T8	Does it check whether the adversary shaped the measurement?
T9	Does it discipline forecasts with signposts and score them when they resolve?

Table 2. The nine tests applied to every comparable.

2.2 The comparables matrix

Y means the source requires the practice. P means the source describes or recommends it without requiring it. N means the source does not address it. The grades describe each document as written, not the quality of the people who use it.

Comparable	T1	T2	T3	T4	T5	T6	T7	T8	T9
US JP 5-0, *Joint Planning* (2020), assessment	Y	Y	N	N	N	N	N	N	P
US JP 3-04, *Information in Joint Operations* (2022), Chapter VI	Y	Y	N	P	N	N	N	N	P
US MCWP 8-10, *Information in Marine Corps Operations* (2024), Chapter 5	Y	Y	N	P	N	N	N	N	P
US DoD Strategy for Operations in the Information Environment (2023)	N	N	N	N	N	N	N	N	N
NATO AJP-5 Ed B (2026) and AJP-10.1 (2023), assessment	Y	Y	N	N	N	N	N	N	P
NATO CIMIC handbook 8.4.15	Y	Y	N	N	N	N	N	N	N
UK JDP 2-00 (2023), intelligence grading	N	N	N	N	Y	N	P	P	N
ICD 203 (2015), analytic standards	N	N	N	N	Y	P	Y	P	N
GAO-21-404SP and OMB M-20-12	Y	P	P	P	N	Y	N	N	N

Comparable	T1	T2	T3	T4	T5	T6	T7	T8	T9
HM Treasury Magenta Book (2020; updated 2026)	Y	Y	Y	P	N	P	N	N	N
OECD DAC criteria (2019)	Y	P	P	N	N	P	N	N	N
UK GCS Evaluation Framework 2.0 and Evaluation Cycle	Y	P	N	N	N	N	N	N	N
UK RESIST 2 toolkit (2021)	P	N	N	N	N	N	N	N	N
RAND RR-809, IIP assessment series (2015 to 2017)	Y	Y	P	N	N	P	N	N	P
CREST 3E and harm-footprint model (2026)	Y	P	N	N	N	N	N	P	N
Carnegie evidence guide (2024)	Y	n/a	P	n/a	N	n/a	N	N	P
CONSORT 2025 and SPIRIT 2025	Y	Y	Y	Y	N	N	N	N	N
Roozenbeek et al. field trial (2022)	Y	P	Y	n/a	N	N	N	N	N
Ukraine: Information Security Strategy, media-literacy strategy, Filter index	P	P	N	N	N	P	N	N	N
Estonia: Ministry of Defence opinion series	P	Y	N	N	N	P	N	N	N
Latvia: national strategic communication concept 2023 to 2027	P	P	N	N	N	N	N	N	N
Lithuania: Vilnius University comprehensive-approach paper (2024) and National Threat Assessment 2026	N	P	N	N	N	N	P	N	N
Germany, Canada, United States: federal counter-disinformation pages	N	N	N	N	N	N	N	N	N
EEAS FIMI threat reports and response framework	N	N	N	N	P	N	N	Y	N

Table 3. Comparables matrix. Sources and reading notes in Section 20 and Annex I. n/a marks a test that does not apply to the source's form, for example a single trial with no forecast.

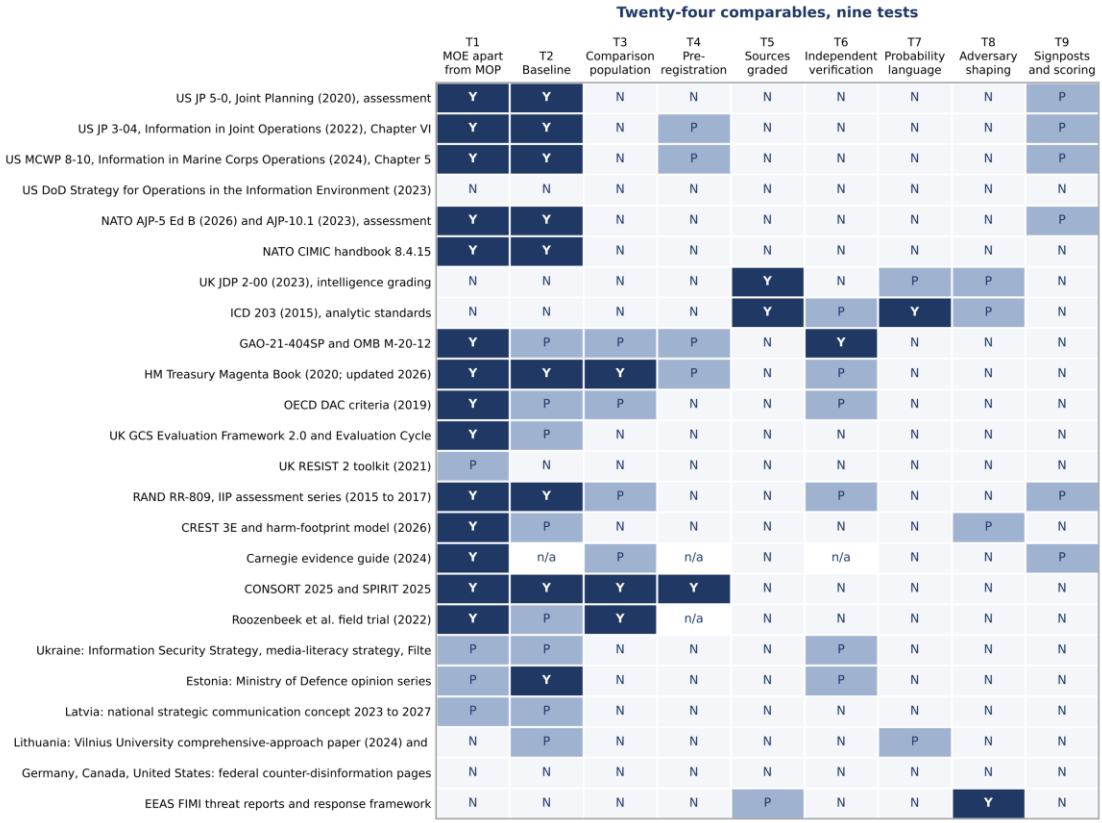


Figure 3. The comparables matrix as a heat map. Dark cells require the practice, mid-tone cells describe it, light cells are silent. Column T9 has no dark cell. Source: Table 3, Treadstone 71.

2.3 What the matrix shows

Read the columns, not the rows. Gaps hide there.

Every column but the last has a Y somewhere, and no row has a Y in every column. Column T9 has none: no source reviewed scores its own forecasts. The parts of a sound measurement already exist in separate traditions that rarely meet.

Military doctrine owns the vocabulary and the baseline. Joint and NATO assessment doctrine separates effect from output and demands a starting measurement, then stops. Neither JP 5-0 nor AJP-5 requires a comparison population, and neither grades the data an assessment rests on, even though NATO grades every intelligence report on the Admiralty scale (JDP 2-00, Table 3.1).

JP 3-04 is the most candid of the set. Read it twice. Its assessment chapter requires a baseline “for identifying changes in conditions” (page VI-5), asks for “a plan for how data will be analyzed” with “specific statistical analyses” (page VI-8), wants analysis by “personnel trained or qualified in statistical analysis techniques” (page VI-10), and closes on a demand for “at least the minimum requirements of reliability and validity” with every compromise described (page VI-12). Then it settles for correlation, on the stated ground that correlation has worked better for planners than causal claims. ASA-MOE disagrees with that one sentence on purpose. Correlation is cheaper. A comparison population is what turns a correlation into a causal claim, and where no comparison exists the protocol prints the correlation as a correlation, at DS-5, instead of upgrading it by habit. The Marine Corps doctrine adds two pieces the protocol keeps: a

written theory of change, and the split between program failure, where the activity never ran as planned, and theory failure, where it ran and did not work (MCWP 8-10, pages 5-4 to 5-5). The exposure log in Section 8 exists to tell those two apart.

Demand for the measurement is on the record. Twice. In November 2023 the Pentagon's Strategy for Operations in the Information Environment called for "assessments to evaluate and monitor OIE and understand the return on investment in campaigning" (page 14) and offered no method for producing one. A year earlier, the Government Accountability Office had found the gap in the field: 13 of 20 Department of Defense components named funding among their five most important institutional challenges in the information environment, and US Africa Command told GAO that the shortfall "limits its ability to develop the means to assess and measure the effectiveness of activities undertaken in the information environment" (GAO-22-104714, September 2022, pages 21 to 22).

Intelligence tradecraft owns source grading and probability language. ICD 203 forces a likelihood term onto every judgment (ODNI, ICD 203). That discipline never crossed into campaign assessment. An analyst who grades a liaison report B2 grades nothing when the same desk reports that a campaign "worked."

Evaluation science owns the counterfactual. The Magenta Book calls a genuinely comparable counterfactual "essential," and the trial-reporting standards require registration and a flow diagram (HM Treasury, Magenta Book; CONSORT 2025). None of those standards expects an adversary inside the data. A clinical trial has no opponent seeding the control arm.

The national programs own the data series. Estonia's Ministry of Defence has fielded an annual opinion survey on national defense since 2000. The March 2026 wave reached 1,372 respondents aged 15 to 74 through a random sample, 73 percent by web panel and 27 percent by telephone (Kaitseministeerium). Ukraine's media-literacy index has run since 2020, and Latvia's concept sets the only numeric outcome target found in the review: at least 75 percent of adults with minimal media and information literacy by 2027 (likumi.lv). None of the series compares reached and unreached populations, so none of them isolates a campaign. Each one is usable as pre-period data for a synthetic control (Section 6).

Lithuania shows the pattern at its clearest. Good program. No measurement. A 2024 Vilnius University paper on the country's whole-of-society approach reports that 82.5 percent of respondents in the 2022 Civic Attitudes Survey (n = 1,212, fielded by Norstat) had not watched a Russian television channel in the preceding month, then reports in the next paragraph that 84 percent had not watched the independent Russian-language alternatives either (Maliukevičius, 2024, page 7). The paper credits the first number to the 2022 broadcast ban and to a declining Russian-speaking population, and it is honest about the second. Neither figure has a baseline wave printed beside it, and neither has a comparison. The paper describes a program. Nothing in it measures one.

2.4 Nine gaps, ranked by consequence

Rank	Gap	Evidence	Confidence the gap is general
1	No comparison population required in any counter-influence framework reviewed	Table 3, column T3; CIMIC handbook admission on isolating variables; JP 3-04 preference for correlation over causation	Highly likely (85 to <95) across the reviewed set

Rank	Gap	Evidence	Confidence the gap is general
2	Reach and effort reported as effect	RAND RR-A2771-1 practitioner quote; CCD reach metrics; German, Canadian, and US pages report activity only; GAO-22-104714 on components unable to measure effectiveness	Highly likely (85 to <95)
3	Self-measured results go unchecked	No reviewed counter-influence framework requires independent verification	Likely (55 to <85)
4	Data sources go ungraded	Only intelligence tradecraft grades sources (T5)	Highly likely (85 to <95)
5	No check for adversary shaping of the measurement	Only EEAS threat analysis and CREST address adversary manipulation, and neither applies it to the defender's measurement	Likely (55 to <85)
6	Causal claims carry no probability language	Only ICD 203 binds a likelihood term, and it governs intelligence products, not campaign reports	Highly likely (85 to <95)
7	Targets and forecasts carry no signposts and no scoring	Latvia's 75 percent target has no registered measurement design; doctrine lists indicators without calibration	Likely (55 to <85)
8	No cost per unit of effect	No reviewed framework prices an effect, only an activity	Likely (55 to <85)
9	Respondent protection absent where respondents meet the false claim	A scoping review found "fewer than 30% of misinformation papers report any debriefing" (Greene et al., 2023)	Likely (55 to <85)

Table 4. Gap findings. Confidence applies to the 24 sources reviewed. A framework outside the review that closes a gap lowers the band.

2.5 Keep, extend, build new

Treatment	Element	Origin	What ASA-MOE does with it
Keep	Measure of effectiveness and measure of performance	JP 5-0; AJP-5; AJP-10.1	Uses the doctrinal definitions unchanged and adds a test: a measure of effectiveness counts only with a comparison population
Keep	Baseline before action	JP 5-0; JP 3-04; MCWP 8-10; CIMIC handbook	Baseline becomes Component 1, hashed and escrowed
Keep	Theory of change; program failure apart from theory failure	MCWP 8-10; JP 3-04 "expected change" mapping	The registered hypothesis chain; the exposure log separates a campaign that never reached its audience from one that reached it and failed

Treatment	Element	Origin	What ASA-MOE does with it
Keep	Admiralty grading and Kent-style bands	JDP 2-00; ICD 203; ASA Evidence Standard	Applied to every data input and to the effect claim
Keep	Trial registration and flow reporting	CONSORT 2025; SPIRIT 2025; OSF	Shapes the registration form and the sample field
Extend	Effort, exposure, effect	CREST 3E model	Effort and exposure become graded covariates; effect requires a design grade
Extend	Counterfactual designs	Magenta Book; Abadie; Hemming	Ranked DS-1 to DS-5, and the rank caps the claim
Extend	National opinion series	Estonia, Ukraine, Latvia	Registered as donor pools for synthetic control
Build new	Two-ceiling combination rule	ASA-MOE	Design grade and evidence quality each cap the band; the lower governs
Build new	Deception check on the measurement	ASA C8 applied to the defender's data	Indicator flooding, comparison seeding, and timed counter-messaging score as D-Score markers
Build new	Estimation layer with signposts and calibration	ASA-MOE	Section 11
Build new	Cost per unit of effect with interval	ASA-MOE	Printed at the point estimate and at each interval bound
Build new	Conformance marks and independent audit	ASA-MOE	Section 14 and Section 15

Table 5. What the protocol keeps, extends, and adds.

3. The T71 Standard inside the protocol [Tearline 2]

ASA-MOE runs on methods Treadstone 71 already publishes and uses. No new dialect. Nothing below adds a second vocabulary.

Treadstone 71 method	Role in ASA-MOE	Where
ASA Evidence Standard, Public Edition 1.2	Grades every data input and the effect claim: Admiralty anchor, 30 factors in nine clusters, T-Score and D-Score, nine rules	Sections 4, 10
The Adversary Index (TAI)	Supplies the adversary profile for the narrative the campaign answers. Cognitive Warfare Velocity (CWV), Narrative Volume (NV), and Attribution Evasion (AEI) set the expected pressure on the comparison population	Sections 9, 11, 18

Treadstone 71 method	Role in ASA-MOE	Where
Cognitive Warfare Tradecraft Reference (CWTR)	Classifies the adversary narrative by maneuver and DISARM crosswalk in the registration, so a shift in adversary tactics registers as a signpost	Sections 5, 11
Cognitive Warfare Monitor	Weekly graded narrative series, run in treatment and comparison areas alike, as the adversary-effort covariate	Sections 9, 11
STEMPLES Plus and Indicators of Change	Profiles each population before launch; the 0 to 6 indicator discipline scores signposts each wave	Sections 9, 11
Cultural Nexus Framework	Names the moderators registered in advance and sets the similarity test when an estimate travels to a new population	Sections 5, 11
Three-source triangulation	Gates any adversary-effort indicator before it enters the analysis as a covariate	Section 9
Advanced Analytic Dominance (66 Advanced SATs)	Key Assumptions Check on the registration; Analysis of Competing Hypotheses on the causal claim; Indicators and Signposts of Change on every estimate; Premortem before launch	Sections 5, 10, 11
Cognitive Bias Validator	Optional input to the bias disclosure graded as factor F19	Section 10
Read-back rule and publish/withhold firewall	Governs restricted publication: effect-side results publish, cheap-to-defeat collection details stay restricted	Section 13
Public Attribution Series score-link discipline	Model for how an audited Effect Statement enters the Adversary Index: as a graded evidence item at a quarterly release, never directly	Section 18

Table 6. Treadstone 71 methods used by the protocol.

4. Definitions [Tearline 2]

Term	Definition
Campaign	A bounded set of communications, run by an operator for a funder, intended to change what a defined population believes or does in response to an adversary narrative
Funder	The organization paying for the campaign and answerable for the spend. The funder owns the registration and the Effect Statement
Operator	The organization designing and delivering the campaign's communications
Measurer	The organization fielding the instruments and holding the raw data. Graded as a source under ASA cluster C1
Auditor	A party with no role in the campaign that verifies design, data, computation, and estimates, and signs the Audit Statement

Term	Definition
Effort	The resources and actions a party puts into communication: messages produced, spend, channels used. Effort is a measure of performance. The CREST 3E model uses the same word for adversary activity
Exposure	Documented delivery of content to a member of a population. Exposure is reach, and reach is never effect
Effect	The difference in a registered effect measure between the treatment population and the comparison population after the campaign, net of their difference before it
Measurement	A graded statement of an effect that has already happened, computed from data collected under a registered design
Estimate	A graded projection from a measurement to a time, population, scale, or adversary condition that was not measured. Every estimate names the measurement it rests on
Signpost	A registered, observable, dated event that, if seen, raises or lowers confidence in an estimate. Scored 0 to 6 each wave
Falsifier	A registered observation that, if made, defeats a claim or an estimate outright
Treatment population	The defined population the campaign reaches, stated in the registration with size, frame, languages, and exclusions
Comparison population	A defined population the campaign does not reach during the measurement window. How it is chosen sets the design grade
Attribution window	The registered interval in which a change in a behavioral indicator counts toward the campaign
Leakage	Exposure of the comparison population to campaign content or its echo during the window
Deviation	Any departure from the registered design, instrument, schedule, or analysis plan
Minimum detectable effect	The smallest effect the design and sample distinguish from no effect at the registered confidence
Design grade	DS-1 to DS-5, the grade for the comparison design as executed. Enters the ASA path as factor F20
Effect Statement	The one-page report of a measurement (Section 12)
Estimate Statement	The half-page report of one or more estimates, attached to an Effect Statement (Section 12)
Interested party	Any organization whose standing, payment, or reputation depends on the result: the operator always; the funder when it measures; a ministry supplying its own indicator feed

Table 7. Definitions.

The protocol keeps one distinction the CREST authors draw. Their report prefers "'estimating' rather than 'measuring' information effects, reflecting the limitations in precision and accuracy of currently available

data and methodologies." ASA-MOE takes the opposite order on purpose. A campaign's effect is measured first, with its interval printed and its design graded, because an estimate needs something solid to stand on. Estimation follows, and it inherits every limit the measurement carries.

Part II · Measurement

The measurement layer answers one question: what changed, in whom, and how sure the evidence lets anyone be that the campaign caused it. Six components do the work. Five gather and fix the evidence. The sixth grades it.

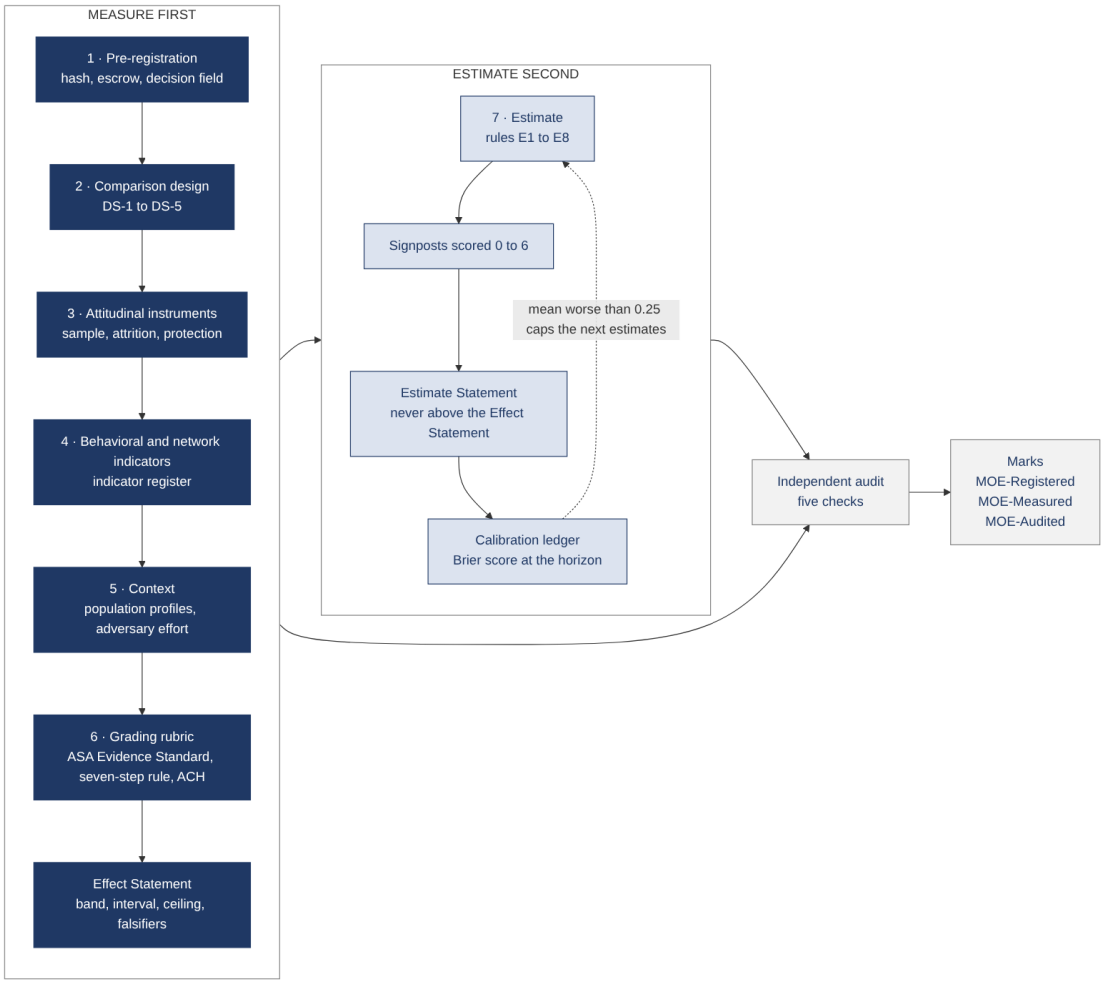


Figure 1. The protocol from registration to graded statement. Measurement runs first and closes with a graded Effect Statement. Estimation runs second, rests on that statement, and never prints above it. Source: Treadstone 71.

5. Component 1 · Pre-registration [Tearline 3]

The registration carries the weight of the whole protocol. Nothing else weighs as much.

Everything measured after launch is checked against it, and nothing observed after launch changes it. A campaign without a registration still gets measured, but it carries no design grade: F20 scores 0, the statement prints Voided in the design field, and no mark attaches.

5.1 What the registration fixes

Annex A is the form. Before the first message goes out, the registration states:

- the parties: funder, operator, measurer, and auditor if engaged;
- the adversary narrative addressed, with its CWTR maneuver identifier and Cognitive Warfare Monitor reference where one exists;
- the primary hypothesis, directional and measurable, with the minimum effect it claims;
- numbered secondary hypotheses, each marked secondary;
- the treatment population, its sampling frame, languages, and exclusions;
- the comparison design and its assignment or matching procedure;
- the instrument, item by item, and each behavioral indicator with its source, attribution window, and leakage rule;
- the moderators expected to change the effect, drawn from the population's STEMPLES Plus and Cultural Nexus profile;
- the minimum detectable effect and the sample that supports it;
- the analysis plan: model, covariates, missing-data rule, attrition rule, and multiple-comparison rule;
- the stopping rule and the deviation policy;
- the follow-up waves, and the signposts and falsifiers for every estimate the funder intends to request;
- the budget, the publication commitment, and the signatures.

One field matters more than the rest. The registration names the decision the result feeds: what the funder does differently on a positive result and on a null one. A registration with no decision attached describes a research study, and the protocol is not built for research studies.

5.2 Timestamp, hash, escrow

The completed form is saved as a PDF. Its SHA-256 hash goes on the form's cover and into escrow with a dated third party before launch. Acceptable escrow includes a public registry, a timestamping service, or the auditor's engagement file. Three registries meet the test, each in its own way:

Registry	What its own documentation states	Fit
OSF Registries	A registration is "a time-stamped, read-only version of your study plan"; once submitted, "you will not be able to edit or make changes to it"; an embargo runs up to four years	Public or sealed registrations; holds the hashed PDF and files
AsPredicted	Single-page time-stamped PDFs; a registration "remains private until an author makes it public"; public registrations "cannot be modified"	Short registrations; sealed by default
AEA RCT Registry	Built for randomized trials in the social sciences; marks registrations made before the intervention starts	DS-1 and DS-2 designs only

Table 8. Escrow options. The EGAP registry stopped accepting registrations on October 15, 2023, and points researchers to OSF.

The auditor checks the hash first. Then the date. A registration whose hash fails to match the file presented, or whose escrow date falls after the first exposure, is not a registration.

5.3 Amendments, stopping, and the deviation log

Before launch, an amendment is hashed and escrowed again, and the two hashes print on the Effect Statement. After launch, any change to design, instrument, indicators, or analysis plan is a deviation. The only permitted post-launch change is the registered stopping rule firing.

Every deviation is logged with its date, reason, and the field it touched. A logged deviation is a fact about the campaign and is graded on its merits. An unlogged deviation that the auditor finds counts as concealment and drops the design grade one step per instance. Presenting an unregistered analysis as the primary result voids the design grade: F20 scores 0 and the statement prints Voided with the reason. Voided is not DS-5. DS-5 is a registered design with no comparison; Voided is a registration that was not kept.

An analysis that the registration did not name still prints, under the label exploratory, beside the registered result, under Table 8a. Every exploratory analysis counts, including the ones that found nothing, because the number of analyses tried is what sets the false-positive rate (Simmons, Nelson, and Simonsohn 2011).

Exploratory finding	Band ceiling	Condition
Logged, with the full count of exploratory analyses run and their results	Roughly even chance (45 to <55)	Prints beside the registered result under the label exploratory; never sets the band
Logged, but the count of analyses run is missing or the null results were dropped	Unlikely (15 to <45)	The statement prints the ceiling and the reason; a Major finding at audit
Contradicts the registered primary result	No band of its own	The registered result stands; the contradiction prints as a watch item and a registered falsifier for the next wave; second reader mandatory
Replicated in a later registered wave or campaign at DS-3 or stronger	The replication's own ceiling	The replication is a measurement; the exploratory label drops
Presented as the primary result without registration	Voided (F20 = 0)	The paragraph above

Table 8a. Exploratory analyses.

5.4 Sealed registrations

Some campaigns cannot publish their design in advance without warning the adversary. A sealed registration publishes the hash before launch and holds the content with the funder and the auditor until the Effect Statement issues. Sealing earns the same design grade as a public registration, because the hash proves the design was fixed. The statement records that it was sealed.

5.5 Registration discipline from the Advanced SATs

Three structured analytic techniques run on the draft registration before it is hashed. A Key Assumptions Check lists every assumption the hypothesis rests on and marks each one as supported, unsupported, or

untested. The Premortem asks the team to assume the campaign failed and write down why; each reason becomes a registered falsifier or a design fix. Red-teaming the comparison design looks for the easiest way contamination enters. The three outputs file with the registration. None of them changes the design after the hash.

Registrations carry the identifier MOE-REG-YYYY-NNNN, assigned by the funder or the auditor and never by the operator alone.

6. Component 2 · Comparison design [Tearline 2]

The comparison design separates the campaign from everything else that happened in the same weeks. Five designs qualify, ranked by how well each does that separation. The rank is the design grade, and the grade caps the causal claim.

Grade	Design	Minimum requirements	Ceiling on the causal claim	Method reference
DS-1	Randomized assignment of audiences or areas	Documented random procedure run before launch by a named party, draw recorded; each arm measured pre and post	Almost certain (95 to 99)	CONSORT 2025; cluster extension, Campbell et al. 2012
DS-2	Stepped-wedge rollout across areas or waves	Rollout order drawn at random and recorded; every area measured before its own start; waves at least one measurement interval apart; all areas reached	Highly likely (85 to <95)	Hemming et al. 2018
DS-3	Matched comparison population, pre-specified	Matching variables registered; none added or dropped after launch; comparison never exposed; each population measured pre and post	Likely (55 to <85)	Rosenbaum and Rubin 1983; Stuart 2010
DS-4	Synthetic control built from pre-period data	Donor pool and weighting registered; at least three pre-period points; pre-period fit reported; no post-period tuning	Likely (55 to <85)	Abadie, Diamond and Hainmueller 2010; Abadie 2021
DS-5	Pre/post measurement, no comparison	Baseline measured before launch with the registered instrument	Roughly even chance (45 to <55)	None. DS-5 records a change, not a cause

Table A. Design strength grades. The ceiling is a maximum, not an entitlement.

6.1 Choosing a design

Three questions drive the choice.

Is any part of the population free to hold back? Randomize. Must everyone be reached eventually, but not at once? Stagger the rollout at random; many counter-narrative campaigns fit a stepped wedge, because the design delays areas rather than denying them. Is everyone reached at once? Then the

comparison comes from outside, either a matched population the campaign cannot reach or a synthetic one built from history.

Difference-in-differences estimation under DS-3 and DS-4 follows current practice for staggered timing (Callaway and Sant'Anna 2021; Roth et al. 2023). A funder who accepts DS-5 accepts, before launch, that the strongest statement the campaign will ever earn is roughly even chance. That is a legitimate choice for a small or urgent campaign. As a surprise afterward, DS-5 is not legitimate.

6.2 National survey series as donor pools

DS-4 needs history: at least three pre-period measurements of the effect measure, or a close proxy, across many units. Several governments already hold it. Estonia's Ministry of Defence opinion series runs back to 2000, Ukraine's media-literacy index to 2020, and the Detector Media index fielded on the ministry's Filter project reports annual waves. A registration that draws donors from such a series names the series, the waves, and the items, and grades the series as a source under ASA cluster C1. Donor series run by an interested party, such as a ministry measuring its own program, grades no better than C, and Rule 1 caps the claim.

6.3 Randomization that qualifies

The random procedure is documented: unit of assignment, method (a seeded generator with the seed recorded, a public lottery, a witnessed coin sequence), who ran it, when, and where the output sits. A procedure the operator ran alone without a record counts as if it never happened. The design falls to DS-3 if the data still support a matched comparison, or to DS-5 if not.

6.4 Contamination

A comparison population reached by the campaign or its echo is contaminated. The registration states how contamination will be found: an exposure-check item in the comparison post wave, a check for campaign content shared into comparison areas, and the leakage rule on each indicator. The contamination fraction, c, is the share of the comparison population that reports exposure to the campaign on the exposure-check item, the same quantity Torgerson's formula uses: the proportion of the control group that received the intervention. Exposure in the reached population prints beside it as the check on the item, not as a denominator.

Contamination does two things. Each has a number. First, it shrinks the observed difference toward zero: with contamination c, the observed effect is the true effect multiplied by $(1 - c)$, so the sample needed to detect the same effect grows by $1 \div (1 - c)^2$, and "around 30% contamination can be sustained before the sample size has to be doubled" (Torgerson 2001, page 356 and web appendix). A simulation study of educational trials reached the same point from the other side: with contamination up to 30 percent, an individually randomized design analyzed for compliance still beat a cluster design on power (Keogh-Brown et al. 2007, executive summary). Torgerson's cancer-trial reference set puts typical crossover at 18 percent, with a range of 10 to 36 percent. The ladder sits on those figures: 10 percent is the floor of the observed crossover range, 30 percent is the point where the required sample has doubled, and 50 percent is where half the comparison has received the treatment and the comparison stops being one.

Contamination fraction c	Consequence
c ≤ 0.10	None. The statement prints c
0.10 < c ≤ 0.30	The registered intention-to-treat estimate prints as attenuated, beside a complier or dose-adjusted estimate from the registered plan; the minimum detectable effect is recomputed at 1 ÷ (1 − c)² and the statement says whether the sample still clears it
0.30 < c ≤ 0.50	The design drops one grade. Past 30 percent the sample needed has doubled and the comparison is no longer the population the registration described
c > 0.50	DS-5. Half or more of the comparison received the treatment; what remains is a before-and-after study

Table 8b. Contamination ladder. Thresholds derive from the Torgerson dilution formula, not from a published downgrade rule; the protocol states that origin.

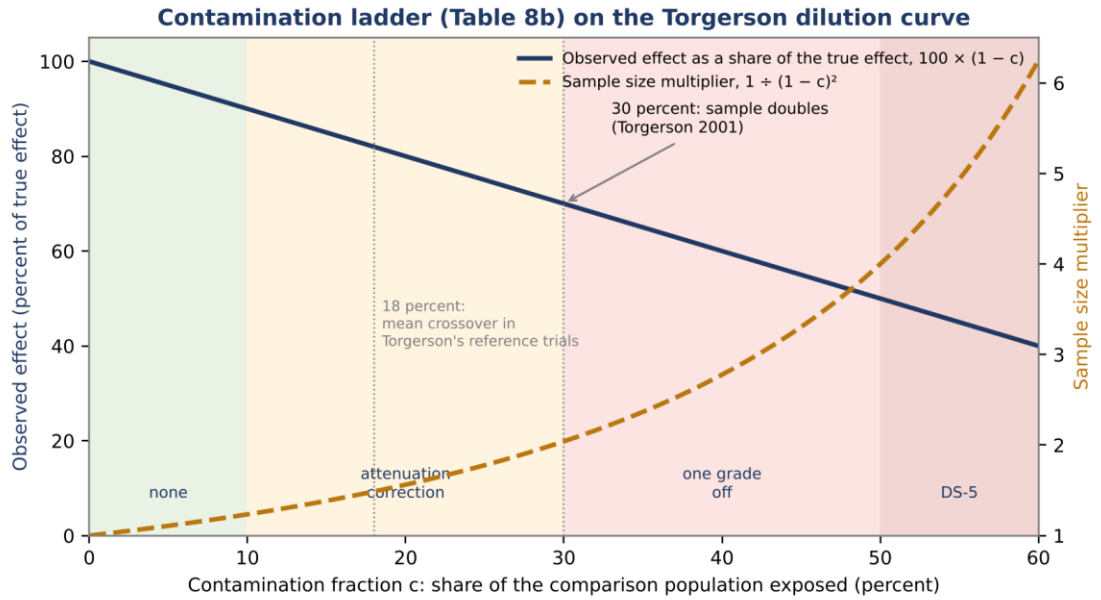


Figure 4. The contamination ladder drawn on the dilution curve. The solid line is the observed effect as a share of the true effect, 1 − c; the dashed line is the sample-size multiplier, 1 ÷ (1 − c)². Shaded zones are the four rows of Table 8b. Sources: Torgerson 2001, BMJ 322:355, Table 1 and web appendix; Treadstone 71.

Contamination found by the auditor and not reported by the measurer is an unlogged deviation.

6.5 What lowers a grade after the fact

The grade comes from the design as run. Plans do not count. One step comes off for each of the following: assignment changed after the draw; a comparison area reached before its scheduled date; a matching or donor variable added or dropped after launch; a wave missed or moved outside its registered window in a way that touches the comparison; a stop for an unregistered reason. The auditor sets the final grade and states each step.

6.6 What the grade does not decide

Grades cap the causal claim. The grade does not decide whether the effect was large, cheap, or worth the money. A DS-1 campaign that moved belief one point is a well-measured small effect. Under DS-5, a campaign that appears to have moved belief 15 points is an unmeasured claim. The protocol keeps the two from being confused.

7. Component 3 · Attitudinal instruments [Tearline 3]

The instrument measures what the campaign set out to change: belief in the adversary's claim, trust in the sources on each side, and intended behavior. Identical items run before and after, in each population.

7.1 What the instrument measures

Primary measure: belief in the specific claim the campaign answers, on a 0 to 100 agreement scale or a labeled scale converted by a registered rule. Conversion for a k-point labeled scale defaults to linear: the lowest label scores 0, the highest 100, and each step 100 divided by k minus 1. A seven-point item scores 0, 16.7, 33.3, 50, 66.7, 83.3, and 100. Secondary measures include trust in the claim's carriers, trust in the counter-sources, intended behavior, and recognition of campaign content. Recognition is an exposure check, never an effect. Exposure-check items name the campaign content, so they are withheld from any pre-launch publication and release with the Effect Statement.

An instrument that measures attitude toward the funder is off the protocol. Gratitude is not belief.

The question is whether the narrative moved, not whether the audience likes the ministry.

7.2 Validated or published

The instrument is either validated, with a citation to its validation, or new and published in full with the registration. A new instrument is weak only when its items are ambiguous, double-barreled, leading, or changed between waves. Items stay identical in wording and order across waves, except that exposure-check items run only post. Translated instruments are back-translated and the two versions filed. A change to the primary item voids the pre/post comparison for that measure.

7.3 Fielding

Fielding runs blind: neither respondent nor interviewer nor panel vendor knows which population the respondent belongs to or that a campaign is under test. Mode stays the same across waves and populations. Sampling is probability-based, or quota-based with quotas and weights registered. Response rates print per population per wave. The rule for speeders, straight-liners, and failed attention checks is registered; a rule applied after seeing the results is an unregistered analysis.

7.4 Sample size and the minimum detectable effect

For two groups compared on a 0 to 100 scale at the post wave, the conventional floor is n per group = $2 \times (z_{1-\alpha/2} + z_{1-\beta})^2 \times \sigma^2 \div \delta^2$. With $\sigma = 25$, $\delta = 5$, $\alpha = 0.05$ two-sided, and power 0.80, the floor is $2 \times (1.96 + 0.84)^2 \times 625 \div 25 = 392$ per group per wave. A design using the pre-wave score as a covariate needs fewer. Stepped wedges need the calculation at the area level, and a wedge with few areas registers a small-sample correction for the interval and reports the interval from it (Hemming et al. 2018). A behavioral primary on a proportion uses the two-proportion floor, n per group = $(z_{1-\alpha/2} + z_{1-\beta})^2 \times [p_1(1 - p_1) + p_2(1 -$

$p_2)] \div (p_1 - p_2)^2$. Results smaller than the minimum detectable effect print as smaller than the minimum detectable effect, never as “no effect.”

7.5 Attrition

People drop out between waves. Some more than others. When dropout differs between populations, the surviving samples stop being comparable, and the effect estimate tilts. The protocol uses the attrition standard of the US Institute of Education Sciences What Works Clearinghouse, which judges overall and differential attrition together against a cautious or an optimistic boundary (WWC Procedures and Standards Handbook, Version 5.0). The cautious boundary is the default, because exposure to counter-messaging plausibly changes who keeps answering. A second trigger stands beside it: differential attrition above five percentage points between populations. When either trigger fires, the statement prints Lee bounds on the effect (Lee 2009). Once the lower bound crosses zero, the causal band drops one step.

7.6 Respondent protection

The instrument shows respondents the adversary’s false claim.

Measuring belief in a claim requires naming it. That exposure carries an obligation, and the protocol writes the obligation into the design rather than leaving it to the measurer’s conscience.

Obligation	Rule	Source
Consent	Where US federal rules apply, studies that withhold the purpose run under an IRB-approved alteration of consent, which requires minimal risk, impracticability without the alteration, no harm to rights and welfare, and, "whenever appropriate," additional pertinent information after participation	45 CFR 46.116(f)
Debriefing	Every respondent who saw the false claim receives a debriefing that states the claim is false, explains why, and offers withdrawal of the data	APA Ethics Code 8.07(c) and 8.08(a); ICC/ESOMAR Code 2025, Article 4
Debrief timing	Each respondent is debriefed at the point of exit, not at the end of the study. A respondent interviewed once, in a cross-sectional wave, is debriefed at the end of that interview. A panel respondent due for a later wave is debriefed after that later wave and leaves each earlier wave with a contact line and a notice that fuller information follows. Anyone who withdraws, is screened out, or shows distress is debriefed at once. Comparison respondents follow the same rule; the design protects the comparison by sampling it, not by withholding the debrief from people who are done	APA Ethics Code 8.08(a) and (b): debrief "as early as is feasible," and where delay is justified, "take reasonable measures to reduce the risk of harm"; Greene et al. 2023
Evidence that debriefing works	A detailed debriefing retracted planted misinformation at intervals from one week to six months; a general debrief "is typically insufficient"	Greene and Murphy 2023; Murphy et al. 2020
Truthful sponsorship	No false claim about who sponsors the study or why; researchers are free to withhold some information when disclosure biases answers	AAPOR Code, 2026
Political opinion data	In the EU, answers revealing political opinions are special-category data under GDPR Article 9; research processing relies on Article 9(2)(j) with Article 89(1) safeguards, including pseudonymization. In the UK, the research condition runs through DPA 2018 Schedule 1, paragraph 4; the ICO marks its research guidance "under review" after the Data (Use and Access) Act 2025	GDPR; ICO

Obligation	Rule	Source
Minors	No respondent under 18 without guardian consent and ethics approval; most campaign measurements exclude minors outright	ICC/ESOMAR Code 2025, Article 2

Table 9. Respondent protection. A registration without a debriefing plan does not qualify for MOE-Registered.

8. Component 4 · Behavioral and network indicators [Tearline 3]

The instrument records what people say. Indicators record what they did. Words and deeds diverge. JP 3-04 names three outcomes for any message: the recipient never saw it, saw it and disregarded it, or "internalizes or processes the message or activity to some degree" (page VI-7). The exposure log answers the first. The instrument and the indicators answer the other two, and they answer differently often enough that the protocol collects each. Each corrects the other: an audience reports a changed belief and changes nothing, or changes a behavior it never reports.

8.1 Indicator families

Family	Examples	Effect or context	Typical source grade
Behavior toward the claim	Regional search share for the claim's terms; sharing of the adversary content; reports of the content to a platform or hotline	Effect	B to C, by provider record
Behavior toward the counter-source	Traffic to the countering source by area; uptake of a requested action, such as reporting a scam or registering to vote	Effect	B to C
Network spread of the claim	Reproduction rate of the claim across accounts and networks; reach of the claim inside each population; resonance with local grievances; reactance after the campaign	Effect, when compared across populations	B to C, with a C6 check
Adversary effort	Weekly item counts and new variants of the narrative in each population, from the Cognitive Warfare Monitor; CWTR maneuver tags	Context covariate, never effect	Graded per item
Campaign delivery	Exposure log by population and date; spend	Measure of performance	Operator data: C at best, interested

Table 10. Indicator families. The network-spread family follows the reproduction, reach, resonance, and reactance dimensions of the CREST harm-footprint model, applied to the gap between reached and unreached populations.

The adversary-effort family deserves a word.

CREST's 3E model separates effort, exposure, and effect, and its authors warn that adversary operators "likely perceive and believe their influence actions are having an impact," so they keep investing. ASA-MOE measures that effort in each population as a covariate. If the adversary doubled its output in the

comparison regions during the window, the effect estimate must account for it, and the D-Score must ask whether the doubling targeted the measurement.

8.2 What qualifies

An indicator qualifies when it records an action by members of the population, runs by the same method in the comparison population, has a registered attribution window, and comes from a source graded on the ASA anchor. Impressions, reach counts, follower growth, and sentiment scores from an unpublished model fail the test. So does any indicator with no comparison population or no named source.

8.3 The indicator register

Annex B is the register: one row per indicator, completed in the registration, with source and grade, effort level, attribution window, leakage rule, expected direction, interested-party flag, corroborating source, C6 screen, and primary or secondary status. An indicator added after launch is exploratory and carries no weight in the causal judgment.

8.4 Windows, leakage, and interested feeds

The attribution window does not move after launch. Moved after launch, an indicator turns exploratory and prints without weight.

Search share and traffic leak most, because a countering source reached from one area is a click away from the next; the usual leakage rule reports the indicator with and without leaked units and names in advance which figure the judgment rests on. An indicator supplied by the operator, the funder, or a ministry with a stake in the result is flagged interested, grades C at best, and is reported without weight unless an uninterested source corroborates it.

8.5 Bot traffic and adversary shaping

Adversaries who shape a narrative also shape indicators.

Coordinated accounts flood a search term, inflate sharing counts, or push traffic to a countering source to make a campaign look effective, or to a rival source to make it look useless. Cluster C6, factor F17, screens each platform-derived indicator for coordinated inauthentic behavior. C8 carries the markers observed into the measurement's D-Score. An indicator that fails the C6 screen is reported and not weighted.

8.6 Indicators as primary measures

The primary measure is attitudinal unless the registered purpose is a behavior. A campaign built to move people to report a scam, register to vote, or call a line registers that behavior as primary, and the instrument becomes secondary.

9. Component 5 · Context: population profiles and adversary effort [Tearline 3]

Measurement without context answers "how much" and leaves "under what conditions" blank. The estimation layer needs the second answer. Component 5 records it before launch and every wave after.

9.1 Population profiles

Each population, treatment and comparison, receives a pre-launch profile on the STEMPLES Plus domains (social, technical, economic, military, political, legal, educational, security, plus the domains the case requires) and on the Cultural Nexus dimensions that bear on the narrative. The profile does two jobs. First, the profile tests whether the comparison population matches the treatment population on the registered moderators. Second, the profile sets the baseline for transport: an estimate that the effect travels to a third population is graded on how closely that population's profile matches the measured ones (Section 11.4).

9.2 Adversary effort series

The adversary-effort series runs weekly through the window in every population. Items enter the series only after three-source triangulation, the same gate that governs each indicator in the Adversary Index. Each item carries its CWTR maneuver tag. The series answers three questions at the close: did the adversary press harder in one population than the other, did its tactics shift, and did any shift line up with the measurement waves.

9.3 Adversary profile

The registration records the adversary's current Adversary Index profile. Cognitive Warfare Velocity (CWV), Narrative Volume (NV), and Attribution Evasion (AEI) set the expected pressure. A campaign answering a high-velocity, high-volume actor faces a harder test than one answering a slow and narrow one, and the Estimate Statement says so.

9.4 Indicators of Change scoring for signposts

Signposts (Section 11.5) are scored each wave on a 0 to 6 scale adapted from the Treadstone 71 Indicators of Change discipline.

Score	Reading
6	Signpost fully met in the registered direction
5	Signpost nearly met; one registered condition outstanding
4	Signpost strongly active
3	Signpost active, with concrete observations
2	Signpost detectable
1	Signpost faint
0	No observation

Table 11. Signpost scale. Each score carries the source grade of the observation behind it and a direction: supports or opposes the estimate.

10. Component 6 · The grading rubric [Tearline 2]

The rubric runs the ASA Evidence Standard twice: once on each data input, and once on the effect claim itself. Nothing in it is new to a reader of that standard except the object and the two ceilings.

10.1 Grading the inputs

Every input enters on the Admiralty anchor, source reliability A to F and information credibility 1 to 6, printed apart and never averaged.

Grade	Reliability of the source	Grade	Credibility of the information
A	Completely reliable	1	Confirmed by other sources
B	Usually reliable	2	Probably true
C	Fairly reliable	3	Possibly true
D	Not usually reliable	4	Doubtful
E	Unreliable	5	Improbable
F	Reliability cannot be judged	6	Truth cannot be judged

Table 12. The NATO Intelligence Grading System, as printed in JDP 2-00, 4th edition (August 2023), Table 3.1, and used as the ASA entry anchor.

Inputs behind the primary measure (the measurer, and the source of any behavioral indicator registered as primary) grade at Level 3, the full factor set, with C9 marked N/A. Secondary inputs grade at Level 2: the anchor plus C1, C2, and C8. Context items grade at Level 1, the anchor only. Each input receives its own ledger row. An interested measurer caps at C.

10.2 Grading the effect claim

The effect claim is an item graded at Level 3. Full effort, every time.

- **Factor F20 carries the design grade.** Confidence calibration and reproducibility, in cluster C7, is where the comparison design enters the ASA path.

Design grade	DS-1	DS-2	DS-3	DS-4	DS-5	Voided
F20 score	90	75	60	55	35	0

Table 13. Design grade to F20. The auditor adjusts by up to five points for reproducibility: up when a second analyst recomputes the same result from raw data, down when the computation cannot be reproduced. Per-factor thresholds are in the ASA Evidence Standard, Public Edition 1.2; the public computation is the unweighted mean, so any auditor reproduces the T-Score from the public text.

- **The T-Score follows the ASA formula unchanged:** the mean of C1, C2, C3, C4, and C7, capped at 100 minus the higher of the C5 and C6 risk scores.
- **The D-Score belongs to the measurement.** Cluster C8 runs on the measurement itself and scores five kinds of marker: indicator flooding, seeding of the comparison population, timed counter-messaging during the post wave, synthetic or automated survey responses, and an operator claim that outruns the data. Floor 25, as in ASA Public Edition 1.2.
- **Cluster C9 grades estimates, not measurements.** C9 is marked N/A on the effect claim. Section 11 states how the estimate layer uses it.

10.3 Two ceilings

An effect claim is bounded twice. The lower bound wins.

The design grade sets one ceiling (Table A). ASA Table 5 sets the other through the T-Score band: Strong (80 to 100) permits any band, Adequate (60 to <80) caps at Likely, Weak (40 to <60) caps at Roughly even chance, and Insufficient (0 to <40) permits no forward judgment. The lower governs. A DS-2 design with a T-Score of 74 prints no higher than Likely. So does DS-3 with a T-Score of 90. A strong design cannot carry weak data, and strong data cannot carry a weak design.

10.4 The combination rule

Step	Rule	Effect on the printed band
1	Design ceiling (Table A) and T-Score ceiling (ASA Table 5)	The lower of the two is the maximum
2	Corroboration (ASA Rule 7)	One channel caps the claim at Likely. Highly likely needs a second channel that is independent of the first, not interested, and registered, moving in the registered direction. Almost certain needs that and a replication: a second measured campaign or a second measured population at DS-3 or stronger
3	D-Score on the measurement	Likely or above: one band lower, deception line printed, second reader mandatory. Highly likely or above: the band also prints under a decision-hold line, counter-deception collection is tasked, and MOE-Audited is withheld until that collection resolves (ASA Table 5)
4	Input grades on the primary measure	An input at C or below on reliability, or with a T-Score below Adequate, drops the band one step. A reliability grade that step 6 already caps counts once
5	Attrition bounds	A Lee lower bound that crosses zero drops the band one step
6	Self-measured cap (ASA Rule 1)	A claim measured only by an interested party carries anchor reliability C until an independent audit lifts it
7	The surviving band prints	With its interval, the ceiling it sits under, the step that set it, and its falsifiers

Table B. The combination rule.

Step 2 reads straight off the credibility axis of the entry anchor. Nothing new there. On the Admiralty scale a single-source item sits at credibility 2, “Probably true,” at best; credibility 1, “Confirmed by other sources,” requires independent confirmation (JDP 2-00, Table 3.1), and the ASA Evidence Standard carries the same logic as a corroboration penalty on single-source items. “Probably true” is Likely. “Confirmed” is what Highly likely means. A channel, under step 2, is a data type with its own collection path: the registered instrument is one channel, a behavioral indicator from a provider with no stake is another. Two survey waves from the same panel are not two channels, and the comparison population is not one either; the comparison belongs to the design, which step 1 already graded.

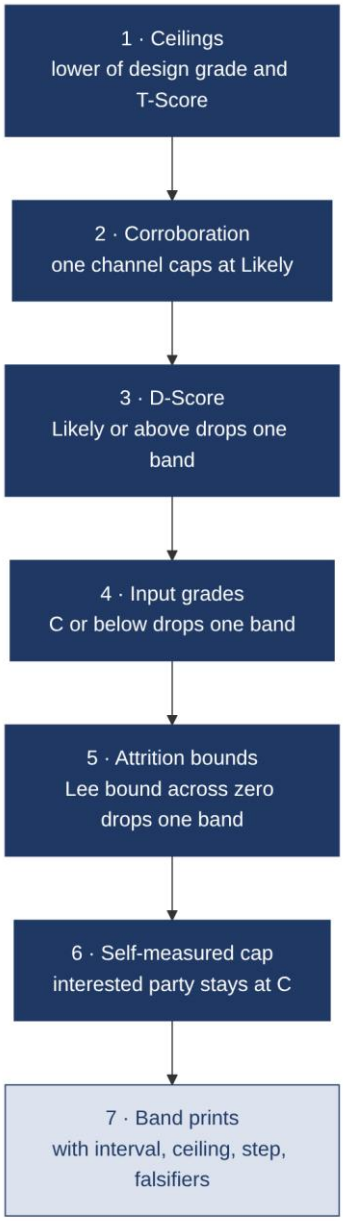


Figure 2. The combination rule. Each step only lowers the band. Nothing after step 1 raises it. Treadstone 71.

10.5 Analysis of Competing Hypotheses on the causal claim

Before the band prints, the analyst runs Analysis of Competing Hypotheses against at least four rival explanations of the observed difference: a news event that reached the populations unevenly; a change in adversary effort between populations; differential attrition or measurement drift; and the campaign. Each piece of evidence is scored against each hypothesis for consistency. The campaign hypothesis must survive with the fewest inconsistencies. If a rival hypothesis survives as well or better, the band drops to Roughly even chance or lower, and the statement names the rival. The matrix files in the data room.

Part III · Estimation

11. Component 7 · The estimate [Tearline 2]

A measurement describes the past. Funders decide about the future. The gap is time.

Will the drop in belief hold through the winter? Will the same campaign work in the eastern regions? And does doubling the spend double the effect? What happens when the adversary answers? Every one of those questions goes beyond the data, and every vendor answers them anyway, usually in the closing paragraph of the report, usually without a band.

The estimation layer brings those answers under the same discipline as the measurement. Estimation runs second, and the order is the point.

11.1 Four estimate questions

Question	What it projects	What it rests on
Persistence	Whether the effect holds at a registered later date	The measured effect, any later waves, and the decay arithmetic in 11.6
Transport	Whether the effect appears in a population that was not measured	The measured effect and the profile match in 11.4
Scale	Whether the effect changes at a higher or lower dose or spend	The exposure log and the dose response inside the observed range (11.7)
Adversary response	Whether the effect survives a change in adversary effort or tactics	The adversary-effort series, the Adversary Index profile, and the CWTR maneuver record (11.8)

Table 14. The four estimate questions.

11.2 Estimate rules

Rule	Text
E1 · Measurement first	No estimate without an Effect Statement of at least MOE-Measured standing. The estimate cites that statement by registration identifier.
E2 · The ceiling	An estimate never prints above the band of the measurement it rests on.
E3 · One step per reach	Each reach beyond the data drops the band one step: time past the last measured wave; a population that was not measured; a dose outside the observed range; an adversary condition that was not observed. Steps add.
E4 · Horizon	No estimate projects past twice the measured interval, counted from first exposure to the close of the last measured wave.
E5 · Signposts	Every estimate carries at least two registered signposts, one that supports it and one that opposes it, each with an observable, a source and its grade, and a date window.
E6 · Falsifier	Every estimate names at least one observation that defeats it outright.

Rule	Text
E7 · Scoring	Every estimate resolves at its horizon and enters the calibration ledger (11.10). An estimate that never resolves counts as a miss.
E8 · Revision	Estimates re-issue under ASA Rule 9 when a signpost moves or a later wave lands. Prior bands stay on the record with their dates.
E9 · Outside view first	Before the inside-view band prints, the estimate states the reference-class rate: what published studies or the house ledger show for effects of the same kind at the same horizon. An inside view that departs from the outside view by more than one band names the reason.

Table 15. The nine estimate rules.

Each rule answers one way an estimate goes wrong. Nine ways, and a rule for each. The first two stop an estimate from outrunning its evidence; together they are the measurement-first order the protocol exists to enforce. Rule E3 prices every reach beyond the data in the same unit, one band, so two analysts facing the same reach land in the same place, and E4 stops extrapolation past the point where the fitted decay has nothing to stand on. Signposts and falsifiers, E5 and E6, make the estimate falsifiable before it is wrong. Scoring and revision, E7 and E8, keep score and keep history. The last rule is the reference-class forecast: planners who estimate from the inside of their own case, without the base rate, run optimistic, and the fix is to state the outside view first (Kahneman and Lovallo 1993). Rule E9 is new in the draft and the one rule Treadstone 71 flags for comment: the house ledger that supplies the reference class does not exist until estimates resolve, so for the first year the outside view comes from the published correction literature in Section 11.6.

Rules E3 and E4 do most of the work. Watch them closely. A campaign measured as Highly likely to have cut belief by at least five points supports, one step down, a Likely estimate that the cut holds at the next registered date inside the horizon. Past the horizon the measurement supports nothing until another wave lands. A Highly likely measurement taken in one country supports, at best, a Likely estimate for a neighbor with a matching profile, and Roughly even chance for a neighbor that differs on a registered moderator.

11.3 The estimate names an outcome

The estimate layer borrows the discipline of the ASA F-Score, not its factors. That score asks an analyst to name the most likely achieved endgame and print the alternatives. The Estimate Statement does the same at the horizon: it names the most likely outcome (for example, “the reduction holds at four points or more”), prints two alternatives with their bands (for example, “the reduction decays below the minimum detectable effect” and “the reduction reverses after an adversary counter-campaign”), and states which signpost moves the reader from one to another. Cluster C9 stays N/A on the measurement and applies here, to the estimate, in that limited sense.

11.4 Transport and the profile match

An effect measured in one population travels to another only as far as the two resemble each other on the things that drive the effect. The registration names those moderators in advance from the STEMPLES Plus profile and the Cultural Nexus dimensions (Section 9.1). Examples include trust in national media, exposure to the adversary’s state outlets, language mix, and prior belief in the claim.

The match test borrows the ten-point tolerance the Adversary Index uses in peer review. A target population matches when each registered moderator, scored 0 to 100, falls within ten points of the range observed across the measured populations. Matching populations take one step under E3. A population outside tolerance on any registered moderator takes two. Replication in the new population at DS-3 or stronger removes the steps, because the replication is a measurement, and measurements outrank estimates.

The test is a plain-language version of what the generalizability literature does with more machinery. Transport of a causal effect to a new population is licensed only when the populations agree on the variables that modify the effect, and the modifiers have to be named before the comparison is run (Pearl and Bareinboim 2014). Where trial and target populations are each measured on those variables, a propensity-style comparison of the two gives a numeric distance (Stuart et al. 2011). The protocol asks for the moderators in the registration and uses a fixed ten-point band in place of a fitted distance, because a fixed band is reproducible by a second reader with a pencil. Two things about the band are open for comment. Ten points is the house tolerance from Rubric 7.6 of the Adversary Index, chosen for consistency across Treadstone 71 instruments and not derived from transport data; once the calibration ledger holds 20 resolved transport estimates, the band is re-set from the hit rate. And the band is the same for every design grade; a tighter band for estimates resting on DS-1 and DS-2 measurements, where the measured effect is sharp enough to make a mismatch visible, is the first candidate revision.

11.5 Signposts

A signpost is the estimate's early warning. Think tripwire, not forecast.

Each one names a thing an observer sees, not a thing an analyst concludes.

Signpost type	Supporting example	Opposing example	Typical source
Persistence	Search share for the claim's terms stays below the comparison regions for four consecutive weeks	Search share converges with the comparison regions within two weeks of the campaign's end	Platform data, graded
Transport	A partner's own survey in the target region shows prior belief within the measured range	The target region's trust in the counter-source runs 20 points below the measured regions	National survey series, graded
Scale	Exposure per head rises with no fall in the per-exposure effect inside the observed range	Recognition of campaign content rises while belief stops moving	Exposure log; instrument
Adversary response	The Cognitive Warfare Monitor records no new variant of the narrative directed at the reached regions	The Monitor records a new CWTR maneuver against the counter-source, triangulated across three sources	Monitor; CWTR record

Table 16. Signpost examples.

Every wave, each signpost scores 0 to 6 on the scale in Table 11. An opposing signpost at 4 or above drops the estimate one band at the next re-issue. A supporting signpost at 6, from a source graded B2 or better, recovers one step lost under E3. Nothing lifts an estimate above the measurement it rests on.

11.6 Persistence arithmetic

With two or more post waves, the protocol fits a simple decay. If the effect at the first post wave, measured at time t_1, is E_1, and at a later wave t_2 is E_2, then the decay rate is $\lambda = \ln(E_1 \div E_2) \div (t_2 - t_1)$, and the half-life is $\ln 2 \div \lambda$. The fit runs only when the two intervals exclude zero and the two effects share a sign. When the later interval includes zero, the statement prints "effect not detected at t_2" and no half-life. A half-life projected past the horizon in E4 does not print.

The published literature supplies context, not the number.

Meta-analyses find that corrections reduce belief in misinformation and that the misinformation still persists in part afterward (Chan et al. 2017), and that correction is harder on political topics than on health topics (Walter and Murphy 2018). A persistence estimate for a political narrative that runs against that grain needs a signpost to carry it.

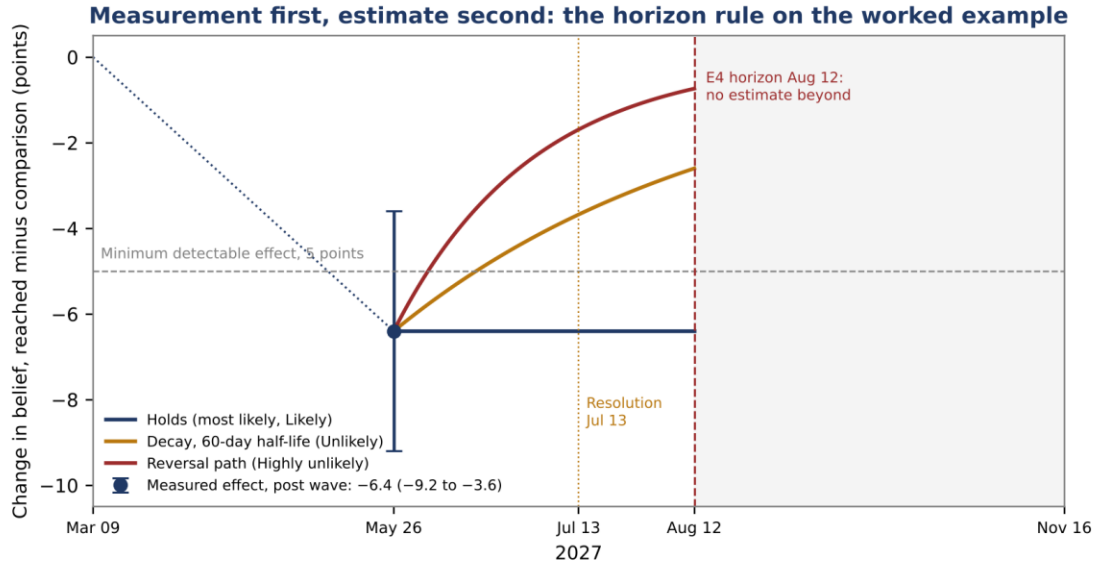


Figure 6. The horizon rule on the worked example. The measured effect at the post wave is the only point on the chart that is a measurement. The three paths after it are the estimate's most likely outcome and its two alternatives; the shaded region past August 12, 2027 is where E4 permits no estimate until a later wave lands. Values are the synthetic ones in Section 16. Source: Treadstone 71.

11.7 Scale

The exposure log records dose: how many campaign items each area received. Inside the observed range, the registered analysis plan fits a dose response, and a scale estimate reads from the fit. Outside the observed range, the estimate loses one E3 step and states the range it left. A change in delivery channel at scale, for example from video to text or from paid placement to partner outlets, counts as a separate reach and takes its own step. The cost per unit of effect at scale prints as a range, never as a point.

11.8 Adversary response

An adversary who notices a working campaign answers it. Expect that.

The estimate layer treats that answer as a condition to project, not a surprise to explain later. Registration also fixes the adversary's Adversary Index profile and the narrative's CWTR maneuver. During the window,

the adversary-effort series shows whether the adversary shifted volume or tactics. A Red Team exercise, one of the Advanced SATs, writes the adversary's two most likely counters before launch, and each counter becomes an opposing signpost. An estimate made under an adversary condition that was not observed during the measurement takes a step under E3. The D-Score of the original measurement carries forward through Table 16a, which runs the full ASA Table 5 scale rather than one cut-off. The logic is that a shaped measurement supports a projection even less than it supports the measurement itself: the measurement at least has the shaped data in hand, while the estimate extends the shape into a period nobody observed.

D-Score of the measurement	Deception band	Effect on the estimate
0 to <45	Unlikely or below	None beyond E2. The estimate inherits the measurement's band as its ceiling
45 to <55	Roughly even chance	Deception warning line printed on the Estimate Statement; second reader signs; the Red Team counter that best explains the measured pattern becomes a mandatory opposing signpost
55 to <85	Likely	The measurement already dropped one band under step 3; the estimate drops one further band and caps at Roughly even chance (45 to <55). Second reader mandatory
85 to <95	Highly likely	No estimate is issued. The Estimate Statement prints "measurement under decision hold" with the D-Score, and counter-deception collection is tasked under ASA Table 5 before any re-issue
95 to 100	Almost certain	No estimate is issued, and none re-issues from that measurement. A fresh measurement with its own D-Score is required

Table 16a. Deception ladder for estimates. Bands follow ASA Evidence Standard Table 5.

11.9 The Estimate Statement

Each estimate prints in a fixed block attached to the Effect Statement (Annex C): the question, the horizon date, the reference-class rate under E9, the most likely outcome and its band, two alternatives and their bands, the steps taken under E3 with reasons, the signposts with their latest scores, the falsifier, and the resolution date.

11.10 The calibration ledger

Estimates that never get scored teach nothing. Score them all.

Every estimate resolves on its horizon date as hit or miss against its most likely outcome. The ledger scores each resolved estimate with the Brier score: the squared difference between the probability stated and the outcome, using the band midpoint as the probability.

Band	Almost certain	Highly likely	Likely	Roughly even chance	Unlikely	Highly unlikely	Almost no chance
Midpoint used for scoring	0.97	0.90	0.70	0.50	0.30	0.10	0.03

Table 17. Band midpoints for the calibration ledger.

A forecaster who says 50 percent every time scores 0.25 (Brier 1950); a forecaster who is always certain and always wrong scores 1.0. Once a measurer or auditor has ten or more resolved estimates, the running mean over the most recent 20 sets that party's cap on the ladder in Table 17a. The cap applies to every estimate the party issues until the running mean recovers, and the party's ledger line prints with each Estimate Statement.

Running mean Brier score (last 20 resolved, minimum 10)	Reading	Cap on the party's next estimates
≤ 0.25	At or better than a coin	None
> 0.25 to 0.35	Worse than a coin	One band lower than the rules otherwise permit
> 0.35 to 0.45	Confidently wrong	Two bands lower; no estimate above Roughly even chance
> 0.45	Anti-calibrated	No estimates issued to any reader. The party enters ten shadow estimates in the ledger at the bands it believes, scored as usual, and returns once the running mean reaches 0.35 or better

Table 17a. Calibration ladder.

The ledger publishes annually by measurer and by auditor, with campaign identities withheld when a statement was restricted. A party with fewer than ten resolved estimates prints "calibration pending" on every Estimate Statement.

Part IV · Reporting, Marks, and Audit

12. The Effect Statement and the Estimate Statement [Tearline 2]

The Effect Statement is the product sentence of a measurement. One page. No appendix rescues it. A minister or a committee reads it in four minutes and knows what was claimed, what was measured, what the evidence supports, and what proves it wrong. The Estimate Statement adds a half page for each estimate requested.

12.1 Effect Statement fields

Field	Content
Header	Campaign; registration identifier and hash with escrow date; funder, operator, measurer, auditor; sealed or public
Registered hypothesis	Verbatim, with the minimum effect
Design	The comparison design as run and its grade, with each step taken and why
Sample	Per population per wave: size, response rate, removals, attrition, and Lee bounds when triggered
Effect	Primary measure with point estimate, interval, and units; the estimand, intention-to-treat on the registered population, with the complier estimate printed beside it whenever recognized exposure in the reached population is under 100 percent; the model and interval method as registered; secondary measures marked secondary; indicators with grades and interested flags
Context	Adversary effort by population across the window; profile match on registered moderators
Verified	One sentence: what the independent evidence shows (ASA Rule 3)
Claimed	One sentence: what the operator asserts beyond that, graded as an interested claim (ASA Rule 3)
Rival explanations	The strongest rival from the Analysis of Competing Hypotheses and why it fell short
Causal judgment	The band with its interval, the ceiling it sits under, and the combination step that set it
D-Score	Score, band, and the deception line when required
Falsifiers	Each registered falsifier and its status
Deviations	Every logged deviation and its treatment
Disclosures	Who paid, who ran, who measured, who audited, and the campaign cost
Cost per unit of effect	Cost divided by the point estimate, and by each interval bound, so the price range shows
Marks	MOE-Registered, MOE-Measured, MOE-Audited as earned, with dates and the auditor's name
Publication	What publishes, what is withheld, and the auditor's confirmation for each withheld item

Table 18. Effect Statement fields.

12.2 Rules the statements obey

- **Verified and Claimed sit in separate fields** so they cannot share a sentence.
- **No band, no product.** A causal judgment without a band is a draft. A band without its ceiling named is a draft.

- **No headline rounding.** The point estimate prints with its interval. “Reduced belief by six points” without “95 percent interval −9 to −4” does not conform.
- **Revision under Rule 9.** When a follow-up wave lands or a falsifier fires, the statement re-issues with the prior band and the new one on the same line, dated.
- **Estimates stay attached.** An Estimate Statement never circulates without the Effect Statement it rests on.
- **Harm prints as harm.** An effect in the wrong direction, belief in the claim rising in reached populations, is graded by the same rule and prints under the label adverse effect with its band. An adverse effect at Likely or above fires the registered stopping rule and goes to the funder within five business days.

12.3 Cost per unit of effect

The figure the funder asked for. Campaign cost divided by the point estimate of the primary effect, in the units of the measure: dollars per belief-scale point, dollars per percentage point of search share, dollars per registered action. Printed three times: at the point estimate and at each bound. A campaign costing $183,000 that moved belief 6.4 points, interval 3.6 to 9.2, prints $28,600 per point, in a range from $19,900 to $50,800. The range is the honest number. A point estimate alone invites a vendor comparison that the intervals do not support.

13. Publication and restricted publication [Tearline 1]

The Effect Statement, registration, instrument, and analysis code publish together. Where an item cannot publish, the statement says which item and why, and the auditor confirms the item was examined.

Some of what the method measures is sensitive.

Which counter-narratives worked, in which regions, on which audiences, is information a ministry needs to keep from the adversary it counters. The protocol follows the Treadstone 71 read-back rule: an indicator publishes only where the adversary cannot cheaply defeat it once it is public. Effect-side results publish, because staging an effect costs the adversary the very capability it fakes. Collection details that an adversary reads back and games stay restricted.

Always publishes	Withheld at the funder’s election, confirmed by the auditor	Always restricted
Registration hash and escrow date; design grade; sample and response rates; effect with interval; causal band and its ceiling; D-Score; disclosures of who paid, measured, and audited	Campaign content; population geography; instrument items; indicator sources; operator identity, with the reason for withholding printed	The settings of bot and coordination screens; account-level collection methods; any detection rule an adversary tunes against

Table 19. Publication tiers. A statement that withholds an item from the first column is not an Effect Statement.

The MOE-Audited mark certifies verification, not disclosure, so a restricted statement earns it exactly as a full one does.

14. Conformance marks [Tearline 1]

Mark	Condition	Granted by	What it says
MOE-Registered	Registration completed, hashed, and escrowed before first exposure, with a debriefing plan	The funder, on the escrow record; or an auditor after a Design Audit	The design was fixed before the campaign ran
MOE-Measured	Baseline and post measurement with the registered instrument and indicators in treatment and comparison populations, deviation log kept	The funder or operator, self-declared; carries the self-measured cap	The campaign was measured to the protocol by a party with a stake in the result
MOE-Audited	Design, data, computation, statement, and any estimates verified by a party with no role in the campaign; signed Audit Statement	The auditor only	An independent party checked the work; the cap lifts

Table 20. Conformance marks.

The marks attach to a campaign, never to an organization. No exceptions.

"MOE-Audited vendor" is a misuse. A mark used without its condition draws a public notice naming the campaign and the mark. An auditor withdraws MOE-Audited when it learns that data presented in the audit were altered or withheld, and issues a revised Audit Statement. A withdrawn mark stays on the record under Rule 9 with its prior state and date. Nobody pays for a mark. The audit is a service with a price; the mark is a fact the audit establishes.

15. Independent audit [Tearline 1]

The audit is how the self-measured cap lifts.

Any party with no role in the campaign is free to audit to the protocol. Treadstone 71 offers the service at published prices through the IO Effectiveness Audit: a Design Audit before launch, a Campaign Audit after, and a Program Audit across a funder's portfolio. The Program Audit applies the Campaign Audit to each campaign in the portfolio and adds the calibration ledger across them; a rubric for portfolio-level findings is scheduled for Version 1.1 and is not part of the draft. The protocol does not require Treadstone 71 as auditor. ASA-MOE requires the five checks, the severity scale, the conflict rule, and the signed statement.

15.1 Five checks

Check	What the auditor does
Design	Verifies the hash against the escrow record and the file presented; sets the design grade from the design as run; inspects the random-procedure record for DS-1 and DS-2
Data	Grades every input on the anchor at its effort level; checks the instrument as fielded item by item; confirms response, removal, attrition, and weighting records
Computation	Recomputes the effect from raw data with the registered plan and prints the recomputed figure beside the reported one; a gap beyond rounding is a finding with its size stated
Statement	Checks the Effect Statement field by field against Section 12
Estimate	Checks each estimate against rules E1 to E9: the measurement under it, the ceiling, the steps, the horizon, the signposts, the falsifier, the resolution date, and the reference-class rate

Table 21. The five audit checks.

15.2 Findings and severity

Severity	Definition	Consequence
Minor	A logged deviation, or an omission repaired in the draft-findings window, with no effect on assignment, measurement, or the primary computation	Reported; no change to grade or band
Major	An unlogged deviation; unreported contamination; a secondary analysis presented beside the primary without its label; an input misgraded; an estimate above its ceiling or past its horizon	Grade drops one step per instance; band re-set; an offending estimate withdrawn
Critical	An unregistered analysis presented as the primary result; a recomputation gap beyond rounding with no logged cause; a hash or escrow failure; evidence of altered data	Design grade voided (F20 = 0); MOE-Audited withheld; MOE-Measured withheld where data were altered

Table 22. Severity scale. The three terms match the published IO Effectiveness Audit.

15.3 Determination, conflict, and terms

The Audit Statement ends in one of three determinations. Granted: minor findings at most; MOE-Audited earned. Granted with findings: major findings exist, each consequence applied, MOE-Audited earned on the re-set band. Withheld: a Critical finding exists; the cap stays, and the campaign re-presents once the cause is repaired where repair is possible.

No auditor audits a campaign it designed, ran, advised on, or measured. Conflict ends eligibility. A prospective auditor with any such role says so at the conflict check, declines, and names an independent auditor working to the same protocol. The rule sits in the engagement letter, and the Audit Statement carries the auditor's declaration under it.

Audits run under mutual non-disclosure. The data-room checklist (Annex D) issues at engagement, and the clock starts when the data room is complete. The client receives draft findings and has five business days to supply corrections or context. Facts change a finding.

Argument does not. The auditor keeps the working file for the life of the Effect Statement and re-opens it under Rule 9 whenever new waves arrive, a falsifier fires, or an estimate resolves.

16. Worked example [Tearline 2]

One campaign, end to end. Synthetic data, real arithmetic. The scenario, parties, and numbers are synthetic and exist to show the format. Nothing here describes a real campaign. The same synthetic campaign appears on the ASA-MOE page and the IO Effectiveness Audit page.

16.1 The campaign and the decision

An adversary narrative claims that allied military exercises will cause fuel shortages. Ministry X funds a four-week counter-narrative campaign by Vendor Y across four regions. The decision the result feeds: renew Vendor Y for the next exercise cycle on a positive result; change approach on a null one.

16.2 Registration

MOE-REG-2027-0000, public, hashed and escrowed March 2, 2027. Primary hypothesis: agreement with the claim falls by at least five points on a 0 to 100 scale in reached regions relative to regions not yet reached. Secondary: search share of the hoax term falls by at least 20 percent relative to comparison regions. Design: stepped wedge, four regions, each exposed for four weeks, rollouts two weeks apart, order drawn at random March 2, 2027, draw recorded by the auditor; a step wave fielded in every region in the week before each rollout, so each region not yet reached stands as the comparison at that step. Analysis: mixed model on region-period means with a small-sample correction, as registered. Adversary Index mapping under Section 18: parent actor and the resonance indicator for the fuel-shortage narrative named in the registration. Instrument: a five-item belief scale fielded online by survey house Z, contracted by the ministry. Indicators: regional search share (platform data) and information-line calls (ministry data, flagged interested). Minimum detectable effect five points; 392 per region per wave. Registered moderators: trust in national broadcast news and pre-campaign agreement with the hoax. Follow-up wave eight weeks after close. Debriefing at each respondent's exit, per Section 7.6. The Design Audit returned minor findings, and MOE-Registered was granted.

16.3 Execution

Baseline fieldwork ran in all four regions the week before the first rollout. Rollouts ran March 9, March 23, April 10, and April 20, 2027; the third was due April 6 and slipped four days for an unrelated event, logged April 11, 2027, with assignment unchanged. Step waves ran in the week before each rollout, and the campaign closed in the last region on May 18, 2027. The post wave ran May 20 to 26, 2027, with exposure-check items. Across the step waves, recognized exposure ran 64 percent in reached regions against four percent in regions not yet reached, so c = 0.04, under the 10 percent line in Table 8b. A second deviation: the operator extended the information-line attribution window from seven days to ten, and logged the change. The adversary-effort series showed no difference in narrative volume between reached and not-yet-reached regions beyond week-to-week noise.

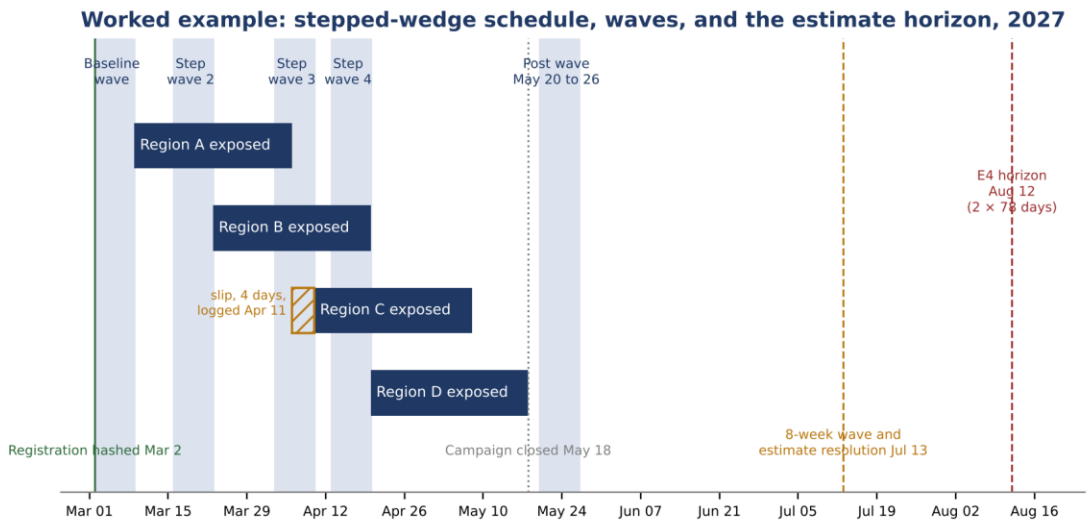


Figure 5. The worked example's schedule. Four regions, four weeks of exposure each, rollouts two weeks apart with one logged slip; a step wave before every rollout supplies the not-yet-reached comparison; the post wave, the eight-week resolution date, and the E4 horizon follow. Dates are synthetic. Source: Treadstone 71.

16.4 Measurement and grading

Belief in the claim fell 6.4 points more in reached regions than in regions not yet reached, 95 percent interval −9.2 to −3.6, intention-to-treat from the registered mixed model with its small-sample correction. With recognized exposure at 64 percent against four percent, the complier estimate prints beside it: −10.7 points, interval −15.3 to −6.0. Search share of the hoax term fell 27 percent more. Information-line calls rose 41 percent, reported without weight. Attrition stayed inside the cautious boundary, and differential attrition was under five points, so no bounds were needed.

Cluster	Factor scores	Cluster score
C1 Source credibility (measurer)	F1 reliability B: 70 · F2 authority: 80 · F3 independence, contracted by funder: 85 · F4 access, 31 percent response, quota-matched: 75	77.5
C2 Information validity	F5 corroboration, effect in all four regions and in search: 80 · F6 consistency across waves: 75 · F7 specificity, interval per region: 85 · F8 credibility 2: 70	77.5
C3 Relevance	F9 answers the registered hypothesis: 95 · F10 inside registered windows: 85	90
C4 Technical integrity	F11 raw data escrowed and hashed: 85 · F12 auditor recomputed within rounding: 90 · F13 timestamps and hashes agree: 80	85
C7 Tradecraft fit	F19 bias disclosure, AI use disclosed: 75 · F20 DS-2, 75, plus 5 for reproduction: 80	77.5
C5 Synthetic media risk	F14 attention checks and speeder removal: 10 · F15 platform data provenance: 15 · F16 file manipulation: 5	10

Cluster	Factor scores	Cluster score
C6 Influence-op risk	F17 coordinated behavior in search data, filtered and cross-checked: 15 · F18 propaganda markers in the countering content: 10	12.5
C8 Counterdeception	No flooding, no comparison seeding, no timed counter-messaging observed; modest operator overstatement present	D 30
C9 Endgame	N/A on the measurement; applies to the estimate	N/A

Table 23. Worked grading of the effect claim. Cluster scores are means of their factor scores.

T-Score: the mean of 77.5, 77.5, 90, 85, and 77.5 is 81.5, which rounds to 82. The cap is 100 minus 12.5, or 87.5, and does not bind. T is 82, Strong. D is 30, deception Unlikely, ledger note only.

Combination: Table A gives DS-2 a ceiling of Highly likely; T Strong adds no ceiling. Step 2 holds, because search share (B2, not interested) corroborates the instrument. D is Unlikely, so step 3 takes nothing. Step 4 stays idle; the primary inputs grade B2 and B2, and the C2 ministry feed carries no weight. Attrition stays inside the boundary, so step 5 does not fire. No self-measured cap applies under step 6; the measurer is independent and the campaign is audited. The three estimate bands in Section 16.5 leave room for probabilities that sum to 100, for example 70, 20, and 10. Analysis of Competing Hypotheses left the news-event and adversary-effort rivals with more inconsistencies than the campaign hypothesis. **Band: Highly likely (85 to <95) that the campaign caused a reduction in belief of at least five points.**

Cost: $183,000. Per point: $28,600 at the point estimate, $19,900 to $50,800 across the interval. The Audit Statement recomputed −6.3, interval −9.1 to −3.5, within rounding, and $29,000 per point at the audited figure. Determination: Granted, MOE-Audited, signed June 30, 2027.

16.5 Estimates

The ministry asked two questions: will the reduction hold at eight weeks after the campaign ends, and will it hold at 26 weeks?

From first exposure on March 9 to the close of the post wave on May 26, 2027, the measured interval runs 78 days. The horizon under E4 is twice that, 156 days from first exposure, which ends August 12, 2027. Under E9 the outside view prints first: the correction literature finds that corrections reduce belief and that part of the original belief returns at later follow-ups (Chan et al. 2017), so a partial hold is the reference-class outcome; the house ledger holds no resolved estimate yet.

Estimate	Horizon date	Inside horizon	Steps under E3	Band
Reduction of at least five points holds at eight weeks	July 13, 2027	Yes	One: time past the last wave	Likely (55 to <85)
Reduction holds at 26 weeks	November 16, 2027	No	n/a	Not issued. E4 bars it until the eight-week wave lands and extends the measured interval

Table 24. Estimates on the worked example.

Eight-week estimate. Most likely outcome: the reduction holds at five points or more, Likely. Alternatives: the reduction decays below the minimum detectable effect without reversing, Unlikely (15 to <45); the reduction reverses under an adversary counter-campaign, Highly unlikely (5 to <15), given the flat adversary-effort series. Supporting signpost: search share for the hoax term stays below its pre-campaign level in every region for four consecutive weeks after May 18, platform data B2. Opposing signpost: the Cognitive Warfare Monitor records a new variant of the fuel narrative directed at the reached regions, triangulated across three sources. Falsifier: the eight-week wave returns the reached regions to baseline. Resolution date: July 13, 2027, entered in the calibration ledger.

16.6 The same campaign, measured the usual way

No registration. One post-campaign survey, no baseline, no comparison, fielded by the vendor, reported to the ministry with 11.2 million impressions and a post-campaign belief figure. F3 independence scores 20, and C1 lands near 45. Factor F20 carries DS-5 at 35, and with no reproducible path C7 lands near 40. C2 loses corroboration and consistency and lands near 45. T comes out in the low fifties: Weak, capped at Roughly even chance. Table A gives DS-5 the same ceiling. The measurer is interested, so the anchor reads C and the step 6 cap stays; step 4 keys on the same grade and takes nothing more. The strongest sentence the protocol permits is roughly even chance (45 to <55) that the campaign changed anything. No estimate issues, because E1 requires MOE-Measured standing, and a design with no comparison population falls short of it. Same money. Same weeks.

17. Contract language [Tearline 1]

Eight model clauses for a grant agreement or service contract. A contracting officer uses them without the rest of the document. Bracketed terms are the funder’s to fill.

Clause	Model text
17.1 Registration	The Contractor shall complete the ASA-MOE Registration Form (Annex A of the Protocol), hash the completed form, and place the hash in escrow with [the Funder / the Auditor / a named registry] before the first exposure. The first payment is conditional on the Funder’s receipt of the escrow record.
17.2 Measurement	The Contractor shall measure the Campaign to the ASA Measurement of Effect Protocol, Version [1.0], with a comparison design of grade [DS-2 or stronger], the registered instrument fielded before and after in all populations, and the registered behavioral indicators.
17.3 Respondent protection	The Contractor shall obtain ethics approval where required and debrief every respondent exposed to the adversary claim, in accordance with Section 7.6 of the Protocol.
17.4 Independent audit	The Funder shall commission an independent audit by an auditor with no role in the Campaign. Final payment is conditional on a signed Audit Statement with a determination of Granted or Granted with findings.
17.5 Data room	Within [ten] business days of the post wave, the Contractor shall deliver the items in Annex D to the Auditor. The audit clock starts when the Auditor confirms the data room complete.

Clause	Model text
17.6 Deviations	The Contractor shall log every departure from the registration with its date and reason. Unlogged deviations lower the design grade, and an unregistered analysis presented as primary voids it.
17.7 Estimates	Any projection of the Campaign's effect beyond the measured data shall take the form of an Estimate Statement conforming to Section 11 of the Protocol. Projections in any other form carry no weight in renewal decisions.
17.8 Publication and marks	The Effect Statement, registration, instrument, and analysis code shall publish within [60] days of the Audit Statement, in full or in restricted form under Section 13. The Contractor shall describe the Campaign as MOE-Registered, MOE-Measured, or MOE-Audited only where the Effect Statement carries that mark, with the Campaign named.

Table 25. Model contract clauses.

18. Linkage to the Adversary Index [Tearline 2]

An audited Effect Statement says something about the adversary as well as the defender.

A narrative that a DS-2 campaign cut by six points in four regions has less grip there than the adversary hoped. The Adversary Index should hear about it, but carefully.

The protocol adopts the Public Attribution Series score-link rule, approved October 13, 2026, and adds two gates of its own. An Effect Statement never moves an Adversary Index score by itself. Each statement enters as an evidence item, graded under the ASA Evidence Standard, at the next quarterly release, and passes eight gates in order.

Gate	Rule	Why
1 Standing	Only an MOE-Audited Effect Statement at Likely or above, with primary inputs graded B2 or better, enters	The same confidence floor the Public Attribution Series applies to an attribution
2 Direction	An effect measured on a defender's campaign is evidence about the adversary narrative's grip on a named population, nothing more. The statement informs the indicators behind the Narrative Volume and Cognitive Warfare Velocity sub-indices that score resonance and uptake. The statement never moves the tempo, OPSEC, attribution-evasion, or AI-adoption indicators, which measure what the adversary does, not how well it lands	Keeps effect-side evidence from being read as effort-side evidence, the same error in reverse that Section 1 names
3 Mapping	The registration names, before launch, the indicator or indicators in the parent actor's Indicators of Change set that the result bears on, and the parent actor. One parent actor per narrative	No post hoc choice of which indicator to move

Gate	Rule	Why
4 Step size	A qualifying statement moves any one indicator by no more than one step on the 0 to 6 scale per quarter, in the direction the evidence supports	Matches the score-link rule's one-step discipline
5 Tolerance	Any resulting move of three or more points on a sub-index, or any move across a tier boundary, requires formal written justification and principal sign-off	The score-link rule's sign-off line. Rubric 7.6 sets the ten-point peer-review tolerance above it; the three-point line is the stricter test that applies to a single evidence item
6 Timing	Moves land only at the quarterly release. Interim unit notes and draft statements never move a score	The score-link rule's release discipline
7 No double count	Evidence already carried into the index through a Public Attribution Series dossier, the Cognitive Warfare Monitor, or the adversary-effort series of Component 5 is not counted again through the Effect Statement. The ledger names the evidence item once	One fact, one entry
8 Null results	An audited null result at DS-2 or stronger with T-Score Strong is evidence that the narrative resisted a competent counter; it moves the resonance indicator up by at most one step. A null result at any weaker grade is no evidence either way and moves nothing	A failed campaign proves the narrative held only when the campaign was measured well enough to have seen it move

Table 25a. Gates for carrying an Effect Statement into the Adversary Index. Gates 2 and 8 are new; the rest follow the Public Attribution Series score-link rule.

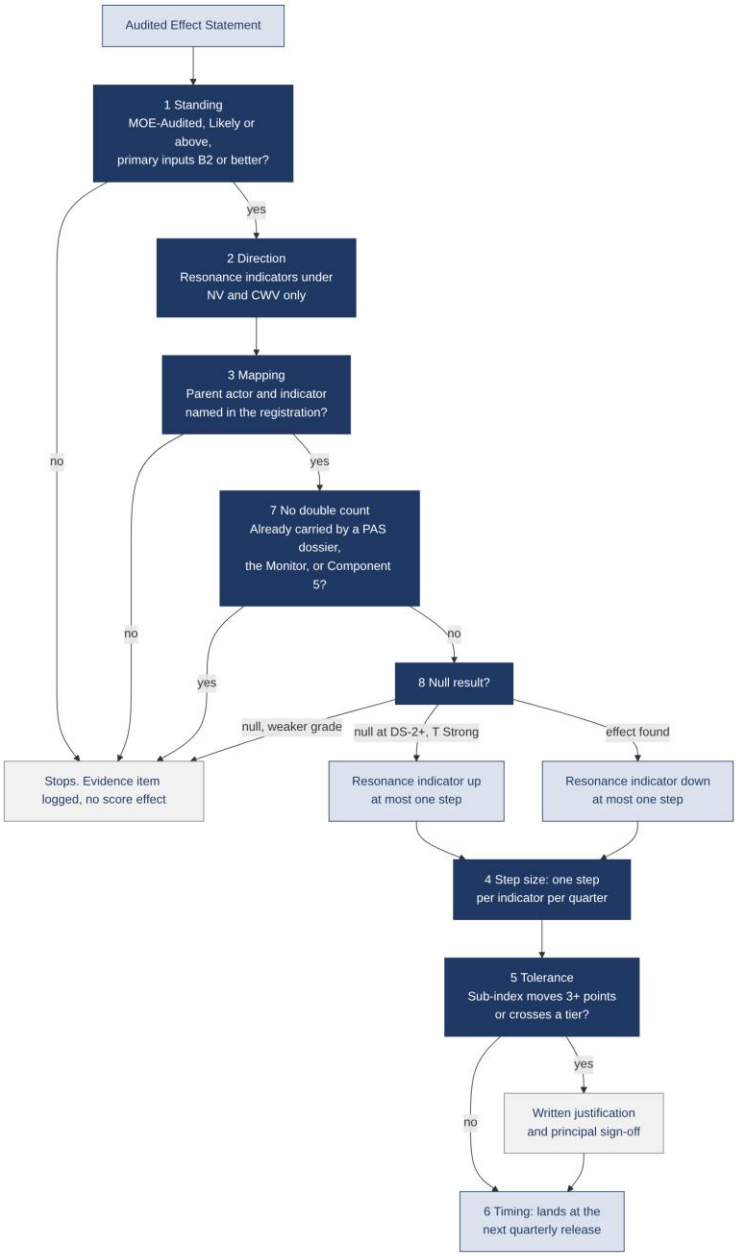


Figure 8. The eight gates in the order an auditor runs them. Three exits lead to the ledger with no score effect; the two that reach gate 4 move one indicator by one step at most, and land only at a quarterly release. Source: Table 25a, Treadstone 71.

Worked through on the example in Section 16. The statement qualifies under gate 1. The registration mapped the fuel-shortage narrative to the resonance indicator in the parent actor's set under gate 3. The adversary-effort series was flat, so nothing from Component 5 moves a tempo indicator. On the October 15, 2027 release (the July 15 release closes its evidence before the June 30 signature), the resonance indicator for that population drops one step, the Narrative Volume sub-index moves within tolerance, and the composite holds. One well-measured campaign against one narrative in four regions is a small fact about a large adversary, and the gates are built to keep it small.

Part V · Validity, Limits, and Use

19. Validation program and limits [Tearline 4]

What has been exercised. Nothing yet on a real campaign. Not one. The protocol publishes as a draft for comment with a synthetic worked example, and it says so.

What the validation program measures. Three things, in order.

Study	Design	Result reported
Design-grade agreement	A fixed set of 20 published campaign and intervention evaluations, each graded DS-1 to DS-5 blind by three raters	Krippendorff's alpha by grade, and every disagreement with its reason
Recomputation agreement	For evaluations with released data, two analysts compute the primary effect independently from raw data and the registered plan	The gap between the two figures and between each figure and the published one
Calibration	Every estimate issued under Section 11 resolves at its horizon and enters the ledger	Brier score by band, by measurer, and by auditor

Table 26. Validation program.

Design-grade and recomputation results publish with the Q1 2027 release of The Adversary Index, beside the ASA Evidence Standard validation results. Calibration results follow the first release after ten estimates have resolved. Until those results land, ASA-MOE stands as a documented method, not a validated instrument.

Limits. The protocol measures campaigns with a defined population, a bounded period, and a measurable outcome. Research studies fall outside its design. So do covert operations that cannot register even under seal, and effects slower than a follow-up wave catches. A measurement is only as good as its measurer; the protocol grades that and cannot repair it. Small campaigns will not reach the minimum detectable effect at any affordable sample, and the protocol's answer is to say so rather than lower the bar. No estimate has resolved yet, so the calibration rule in Section 11.10 has no history to act on. The protocol grades evidence for an effect. Whether the effect was worth the money stays the funder's call, informed by the cost-per-unit line.

Known open items in the draft.

Item	Status	Reason	Treatment in the draft
JP 5-0, *Joint Planning*, 2025 edition glossary	Not read	Distributed on the Joint Electronic Library Plus, which requires a Common Access Card	Quotes the 2020 edition, which carries the same definitions in the CJCSM 3108.01A reproduction
JP 3-04, *Information in Joint Operations* (2022)	Read in full from a copy held by Treadstone 71	Limited release under its Appendix E; distributed on JEL+ to Common Access Card holders only; no public copy exists on any originator site	Cited by chapter and page without a link; its assessment chapter grades on the comparables matrix
Lithuanian evaluation material	Read	National Threat Assessment 2026 and the 2024 Vilnius University comprehensive-approach paper, with its 2022 Civic Attitudes Survey figures	Section 2.3; Table 3; Table 28. No Lithuanian campaign evaluation with a comparison population was found
Inter-rater data on design grades	Not yet collected	Validation program starts on publication	Section 19

Table 27. Open items.

20. Relationship to evaluation practice, and the sources [Tearline 4]

The protocol borrows from four traditions and departs from each in one respect.

Social-science pre-registration fixes the design before the data (Nosek et al. 2018; Chambers 2013); the protocol adds the decision the result feeds and grades the design rather than only recording it. Trial reporting fixes the fields and prints the interval with every estimate (CONSORT 2025); the protocol adds source grades and a deception score, because no opponent shapes a trial's data. Program evaluation separates performance from effectiveness and demands a counterfactual (GAO-21-404SP; OMB M-20-12); the protocol refuses to print an effectiveness claim without a design grade and a band. Military assessment doctrine supplies the vocabulary; the protocol adds the comparison population that doctrine leaves optional. Where the ASA Evidence Standard and ASA-MOE appear to differ, the standard governs and the protocol is in error.

Sources below are ranked by authority. Law first. Each entry was opened at the issuing body while the draft was prepared.

20.1 Law, regulation, and official standards

Source	Use in the protocol
US Code of Federal Regulations, 45 CFR 46.116(f), waiver or alteration of consent	Section 7.6
Regulation (EU) 2016/679, GDPR, Articles 9 and 89	Section 7.6
UK Information Commissioner's Office, research provisions guidance and special category data rules, each marked under review after the Data (Use and Access) Act 2025	Section 7.6

Source	Use in the protocol
US Government Accountability Office, *Program Evaluation: Key Terms and Concepts*, GAO-21-404SP, March 2021	Sections 1, 20
US Government Accountability Office, *Information Environment: Opportunities and Threats to DOD's National Security Mission*, GAO-22-104714, September 2022, pages 21 to 22	Section 2; Table 4
US Office of Management and Budget, M-19-23 (July 10, 2019) and M-20-12 (March 10, 2020)	Section 20
Office of the Director of National Intelligence, ICD 203, *Analytic Standards*, January 2, 2015, with later technical amendments; ICD 208, January 9, 2017	Section 2; Annex H
HM Treasury and the Evaluation Task Force, *The Magenta Book*, March 2020 edition, page updated May 15, 2026	Section 2
OECD DAC, *Better Criteria for Better Evaluation*, December 2019	Section 2

20.2 Military doctrine

Source	Use
Joint Chiefs of Staff, JP 5-0, *Joint Planning* (2020; supersedes the June 16, 2017 edition), glossary and Appendix K. Public release 18-F-1152. A 2025 edition is in force and access-controlled	Sections 1, 4
Joint Chiefs of Staff, JP 3-04, *Information in Joint Operations*, September 14, 2022; supersedes JP 3-13 (2012). Chapter VI, Assessment, and Section VII.g. Releasability LIMITED under Appendix E: no public copy; cited from a copy held by Treadstone 71	Sections 1, 2; Table 3
US Marine Corps, MCWP 8-10, *Information in Marine Corps Operations*, February 29, 2024, PCN 143 000184 00, approved for public release; Chapter 5, Assessment: Measuring Information Effectiveness and Performance	Sections 1, 2; Table 3
US Department of Defense, 2023 Strategy for Operations in the Information Environment, November 2023, page 14	Section 2; Table 3
Joint Chiefs of Staff, CJCSM 3108.01A, glossary reproducing the JP 5-0 definitions	Section 19
NATO, AJP-5, *Allied Joint Doctrine for the Planning of Operations*, Edition B, Version 1, promulgated June 17, 2026; UK national elements October 2026	Sections 1, 2
NATO, AJP-10.1, *Allied Joint Doctrine for Information Operations*, Edition A, Version 1, promulgated January 26, 2023, superseding AJP-3.10; UK edition July 2023, in the UK joint doctrine collection	Sections 2, 19
UK Ministry of Defence, JDP 2-00, *Intelligence, Counter-intelligence and Security Support to Joint Operations*, 4th edition, August 2023, Table 3.1, in the UK joint doctrine collection	Section 10

Source	Use
NATO, AJP-3.10.1, *Allied Joint Doctrine for Psychological Operations*, Edition B (September 2014), marked withdrawn; cited for lineage only	Annex I
NATO CIMIC Centre of Excellence, CIMIC handbook, 8.4.15, measures of effectiveness and measures of performance	Section 1

20.3 Reporting and professional standards

Source	Use
Hopewell et al., CONSORT 2025 statement, *BMJ* 389:e081123; Chan et al., SPIRIT 2025 statement, *BMJ* 389:e081477	Sections 2, 6
Schulz, Altman and Moher, CONSORT 2010, *BMJ* 340:c332; Campbell et al., cluster extension, *BMJ* 345:e5661 (2012); Hemming et al., stepped-wedge extension, *BMJ* 363:k1614 (2018)	Section 6
US Institute of Education Sciences, What Works Clearinghouse Procedures and Standards Handbook, Version 5.0	Section 7.5
Yarbrough, Shulha, Hopson and Caruthers, *The Program Evaluation Standards*, 3rd edition, SAGE, 2010 (Joint Committee on Standards for Educational Evaluation): utility, feasibility, propriety, accuracy, and evaluation accountability	Table 2, the nine tests
American Psychological Association, Ethics Code, Standards 8.07 and 8.08; AAPOR Code of Professional Ethics and Practices, revised June 2026; ICC/ESOMAR International Code, 2025 revision, fifth edition	Section 7.6
UK Government Communication Service, Evaluation Framework 2.0 and the Evaluation Cycle (February 2024); RESIST 2 Counter Disinformation Toolkit (2021)	Section 2

20.4 Peer-reviewed method

Source	Use
Nosek et al., The preregistration revolution, *PNAS* 115(11), 2018; Chambers, Registered Reports, *Cortex* 49(3), 2013	Sections 5, 20
Rosenbaum and Rubin, The central role of the propensity score, *Biometrika* 70(1), 1983; Stuart, Matching methods for causal inference, *Statistical Science* 25(1), 2010	Section 6
Abadie, Diamond and Hainmueller, Synthetic control methods, *JASA* 105(490), 2010; Abadie, Using synthetic controls, *Journal of Economic Literature* 59(2), 2021	Section 6
Callaway and Sant'Anna, Difference-in-differences with multiple time periods, *Journal of Econometrics* 225(2), 2021; Roth et al., What's trending in difference-in-differences?, *Journal of Econometrics* 235(2), 2023; Angrist and Pischke, *Mostly Harmless Econometrics*, Princeton, 2009	Section 6

Source	Use
Torgerson, Contamination in trials: is cluster randomisation the answer?, *BMJ* 322:355, 2001; Keogh-Brown et al., Contamination in trials of educational interventions, *Health Technology Assessment* 11(43), 2007	Section 6.4
Simmons, Nelson, and Simonsohn, False-positive psychology, *Psychological Science* 22(11), 2011	Section 5.3
Pearl and Bareinboim, External validity: from do-calculus to transportability across populations, *Statistical Science* 29(4), 2014; Stuart et al., The use of propensity scores to assess the generalizability of results from randomized trials, *Journal of the Royal Statistical Society A* 174(2), 2011	Section 11.4
Brier, Verification of forecasts expressed in terms of probability, *Monthly Weather Review* 78(1), 1950; Kahneman and Lovallo, Timid choices and bold forecasts, *Management Science* 39(1), 1993	Sections 11.2, 11.10
Lee, Training, wages, and sample selection, *Review of Economic Studies* 76(3), 2009	Section 7.5
Kent, Words of Estimative Probability, *Studies in Intelligence* 8(4), 1964	Annex H

20.5 Peer-reviewed evidence on corrections, prebunking, and research ethics

Source	Use
Roozenbeek et al., Psychological inoculation improves resilience against misinformation on social media, *Science Advances* 8(34), 2022: six randomized studies (n = 6,464) and a YouTube field study (n = 22,632)	Section 2
Chan et al., Debunking: a meta-analysis, *Psychological Science* 28(11), 2017; Walter and Murphy, How to unring the bell, *Communication Monographs* 85(3), 2018	Section 11.6
Greene and Murphy, Debriefing works, *PLOS ONE* 18(1), 2023; Murphy et al., Fool me twice, *Memory* 28(7), 2020; Greene et al., Best practices for ethical conduct of misinformation research, *European Psychologist* 28(3), 2023	Section 7.6
Marushchak, Petrov and Khoperiya, Countering AI-powered disinformation through national regulation: learning from the case of Ukraine, *Frontiers in Artificial Intelligence* 7, 2025. The third author is Deputy Head of Ukraine's Center for Countering Disinformation; claims about that center's own performance grade as interested	Table 28, Ukraine context

20.6 Research institute and international organization reports

Source	Use
Innes, Orpen and Ashford, *Estimating the Effects on the UK of State Information, Influence, and Interference Threats*, CREST, May 2026, funded through the UK Home Office	Bottom Line; Sections 2, 4, 8
Helmus and Holynska, *Ukrainian Resistance to Russian Disinformation*, RAND RR-A2771-1, 2024, sponsored by the Office of the Secretary of Defense	Bottom Line; Section 2
Paul et al., *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, RAND RR-809/1-OSD, 2015; Paul, *Worked Example*, RR-809/4-OSD, 2017	Section 2
Bateman and Jackson, *Countering Disinformation Effectively: An Evidence-Based Policy Guide*, Carnegie Endowment, January 31, 2024	Section 2
European External Action Service, 4th Report on FIMI Threats, March 2026; 1st Report, February 2023; 2nd Report, January 2024	Bottom Line; Section 2

20.7 National programs read at the issuing government

Country	Source	What it measures
Ukraine	Presidential Decree 685/2021, Information Security Strategy to 2025	Expected results stated without metrics; an annual implementation report required

Country	Source	What it measures
Ukraine	Ministry of Culture and Information Policy, media-literacy strategy to 2026	Mostly activity counts; one outcome indicator, resilience "measured through surveys"; 26,194 participants in the 2023 National Media Literacy Test
Ukraine	Filter national media-literacy project, Media Literacy Index 2025 (Detector Media)	Repeated cross-section of media-literacy levels
Ukraine	National Academy of Sciences Institute of Sociology, telephone survey, n = 1,007, October 27 to November 7, 2025	Awareness index 4.6 of 8; practice index 3.1 of 8
Ukraine	Center for Countering Disinformation, analytical reports	Adversary publications, views, interactions, inauthentic sources
Estonia	Ministry of Defence opinion survey series since 2000	Defense attitudes; random sample; mixed web and telephone mode
Latvia	National concept on strategic communication and information space security 2023 to 2027	Target of at least 75 percent of adults with minimal media and information literacy by 2027
Lithuania	Second Investigation Department and State Security Department, National Threat Assessment 2026, March 6, 2026	Threat assessment with defined probability terms; no campaign evaluation
Lithuania	Maliukevičius, *Fortifying Democracies: Lithuania's Comprehensive Approach to Counter Disinformation and Propaganda*, Vilnius University, published by the Eastern Europe Studies Centre, April 2024	Program description; 2022 Civic Attitudes Survey (n = 1,212, Norstat) viewing figures with no baseline wave printed and no comparison
Germany	Federal Ministry of the Interior, measures against disinformation	Activities only
Canada	Global Affairs Canada, Countering disinformation with facts, modified August 25, 2026	Rebuttals only
United States	Department of State, Fact vs. Fiction, January 20, 2022; Global Engagement Center, closed December 23, 2024	Rebuttals only; no published evaluation framework located

Table 28. National programs.

20.8 Registries

OSF Registries help; Center for Open Science, Registered Reports; AsPredicted; AEA RCT Registry guidelines; EGAP registry notice.

21. Use, citation, and license [Tearline 4]

Version 1.0 publishes under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license (CC BY-NC-ND 4.0), matching the license of the ASA Evidence Standard, Public Edition 1.2. The license does not permit commercial reuse or derivative works.

Treadstone 71 grants one set of permissions without fee: citation, quotation with attribution, adoption for internal analytic use by any intelligence unit, agency, funder, or enterprise, and reproduction of the tables and annexes in internal ledgers, contracts, and training material with the citation line intact. A funder is free to write the protocol into a contract without fee.

Reproduction in a commercial product, resale, incorporation into a paid service, translation or adaptation, and any instrument or software that implements the scoring rules require a written license from Treadstone 71 LLC. Ingestion into AI training corpora, embedding indexes, or retrieval pipelines requires a written license. Treadstone 71 reserves text and data mining rights under EU Directive 2019/790, Article 4.

The ASA Evidence Standard, the ASA Measurement of Effect Protocol, STEMPLES Plus, The Adversary Index, TAI, Cognitive Warfare Tradecraft Reference, CWTR, and Advanced Analytic Dominance are trademarks of Treadstone 71 LLC. The marks MOE-Registered, MOE-Measured, and MOE-Audited are used under Section 14 and no other terms.

22. Comments and requests [Tearline 1]

Comments on the draft are open from November 2 to December 2, 2026. Send them to info@treadstone71.com with the subject line “ASA-MOE comment.” Comments that include a worked case help most. Every comment receives a disposition in a register published with Version 1.0 final in January 2027, and commenters are acknowledged by name unless they ask otherwise.

The Registration Form, Indicator Register, Signpost Register, Effect Statement, Estimate Statement, data-room checklist, Audit Statement, and calibration ledger are available as fillable templates. Request them at the same address under the heading “ASA-MOE templates.” To commission an audit, use treadstone71.com/io-effectiveness-audit. Misuse of a mark goes to the same inbox, headed “ASA-MOE mark.”

Treadstone 71 LLC · Half Moon Bay, California · +1 424 234 3629 · treadstone71.com

Annexes

Annex A · Registration Form [Tearline 3]

Completed before the first exposure, saved as PDF, hashed with SHA-256, and the hash lodged in escrow. Sealed registrations publish the hash only.

Block	Field	Guidance
A.1 Identification	Registration ID	MOE-REG-YYYY-NNNN, assigned by the funder or the auditor

Block	Field	Guidance
	Parties	Funder, operator, measurer, auditor, each with a contact of record; the auditor's conflict declaration
	Timestamp and hash	Date, time, and time zone; SHA-256 of the form as PDF; escrow location and date
A.2 Hypothesis and decision	Adversary narrative	The narrative answered, with CWTR maneuver identifier and Cognitive Warfare Monitor reference where one exists
	Primary hypothesis	Directional and measurable: population, measure, minimum effect, comparison
	Secondary hypotheses	Numbered, each with its measure, each marked secondary
	Decision the result feeds	What the funder does on a positive result and on a null one
	Adversary Index mapping	Parent actor, and the indicator or indicators in its Indicators of Change set that the result bears on (Section 18, gate 3)
A.3 Populations and design	Populations	Definition, size, frame, languages, exclusions, for treatment and comparison
	Design	One of the five in Table A, with the grade forecast
	Assignment or matching	Random procedure, matching variables, or donor pool, fixed here
	Contamination detection	Exposure item, sharing check, leakage rules
	Moderators	The STEMPLES Plus and Cultural Nexus moderators expected to change the effect, each scored 0 to 100 per population
	Wave schedule	Dates for each wave, rollout step, and follow-up wave
A.4 Instruments and indicators	Instrument	Items, scale, conversion rule, validation reference or publication commitment, mode, blinding
	Indicators	One row each in Annex B
	Adversary-effort series	Monitor coverage by population; triangulation gate
	Exposure logging and data custody	Who records delivery, who holds raw data, in what form, for how long
A.5 Analysis plan	Minimum detectable effect	With the sample and power calculation
	Primary analysis	Model, covariates, estimand, code or full specification
	Missing data, attrition, multiple comparisons	Each rule fixed here; attrition rule per Section 7.5
	Rival explanations	The rivals the Analysis of Competing Hypotheses will test

Block	Field	Guidance
	Stopping rule and deviation policy	Fixed here
A.6 Estimates	Estimates requested	Each question in Table 14, with its horizon date
	Signposts and falsifiers	One row each in Annex F
A.7 Protection	Ethics approval	Body, reference, and date, or the reason none applies
	Debriefing plan	Content, exit-point timing per Section 7.6 (cross-sectional respondents at interview end; panel respondents after their last wave; withdrawals and distress at once), and withdrawal procedure
A.8 Commitments	Budget; publication; audit; signatures	Total cost; what publishes; audit stage; funder, operator, measurer, and auditor with dates
A.9 Tradecraft checks	Key Assumptions Check; Premortem; red-team read	Outputs attached

Table A-1. Registration Form.

Annex B · Indicator Register [Tearline 3]

Field	Guidance	Sample row
Indicator	Name and unit	Regional search share for the hoax term, percent of regional search volume
Family	Table 10	Behavior toward the claim
Data source	Provider and dataset	Platform search-interest data, regional, weekly
Anchor grade	Reliability and credibility	B2
Effort level	ASA level applied	Level 3, corroborates the primary measure
Attribution window	Days after exposure that count	0 to 7 days after each region’s rollout
Leakage rule	Detection and handling	Compare regions not yet reached with their own pre-period trend; report with and without regions showing campaign-content sharing above baseline
Expected direction	Up or down, and why	Down: the campaign answers the claim the term refers to
Interested party	Yes or no	No
Corroborating source	Independent source for the same behavior	Traffic to the countering source by region, B2

Field	Guidance	Sample row
C6 screen	Coordinated-behavior screen and result	Platform filtering applied; weekly pattern cross-checked against a second index; no anomaly
Status	Primary or secondary	Secondary

Table B-1. Indicator Register with the worked example's sample row.

Annex C · Effect Statement and Estimate Statement [Tearline 3]

Replace every sample value. Keep every field. Delete nothing.

Field	Sample entry
Header	Counter-narrative on a fabricated fuel-shortage claim · MOE-REG-2027-0000, hash escrowed March 2, 2027, public · Funder Ministry X · Operator Vendor Y · Measurer survey house Z · Auditor Treadstone 71 LLC
Registered hypothesis	A four-week counter-narrative in four regions reduces agreement with the claim by at least five points on a 0 to 100 scale relative to regions not yet reached, and reduces search share of the hoax term by at least 20 percent relative to comparison regions
Design	Stepped wedge across four regions, four weeks of exposure each, rollouts two weeks apart from March 9, 2027, order drawn at random March 2, 2027; one rollout slipped four days for an unrelated event, logged April 11, 2027, assignment unchanged. DS-2
Sample	Pre 4,120; post 4,088; quota-matched by region, age, and language; response rate 31 percent; contamination c = 0.04, no action; 96 removed under the registered speeder rule; attrition inside the cautious boundary, differential under five points, no bounds required. Registered minimum 392 per region per wave: met
Effect	Belief in claim: −6.4 points versus comparison, 95 percent interval −9.2 to −3.6, intention-to-treat; complier estimate −10.7, interval −15.3 to −6.0. Search share of hoax term: −27 percent versus comparison regions (B2). Information-line calls: +41 percent (C2, interested, reported without weight)
Context	Adversary effort flat across reached and not-yet-reached regions; populations matched on the two registered moderators
Verified	Belief in the claim fell 6.4 points more in reached regions than in regions not yet reached over the same weeks, and the hoax term's search share fell 27 percent more
Claimed	Vendor Y attributes the entire regional shift to the campaign; interested source, not weighted
Rival explanations	Strongest rival: a national news story on fuel prices in week three. The story reached all four regions on the same day and fails to explain the staggered drop
Causal judgment	Highly likely (85 to <95) the campaign caused a reduction of at least five points. Ceiling: DS-2; T-Score 82, Strong; corroborated by search share. Steps 3 to 6 took nothing
D-Score	30, deception Unlikely, ledger note only

Field	Sample entry
Falsifiers	Comparison regions showing a matching drop before their rollout date: not observed. Eight-week wave returning to baseline: pending. Unregistered analysis in any later report: none
Deviations	Information-line attribution window extended from seven to ten days after launch, logged; indicator reported without weight
Disclosures	Funder Ministry X; operator Vendor Y; measurer survey house Z, contracted by the funder; auditor Treadstone 71 LLC, no role in design or delivery; cost $183,000
Cost per unit of effect	$28,600 per belief-scale point; range $19,900 to $50,800 across the interval
Marks	MOE-Registered · MOE-Measured · MOE-Audited (Treadstone 71 LLC, signed June 30, 2027)
Publication	Statement, registration, instrument, and analysis code published; raw data held by survey house Z, examined by the auditor

Table C-1. Effect Statement, sample values.

Estimate field	Sample entry
Question	Persistence: does the reduction of at least five points hold to the eight-week wave?
Rests on	MOE-REG-2027-0000 Effect Statement, Highly likely, MOE-Audited
Horizon	July 13, 2027; inside the E4 horizon of August 12, 2027
Reference class (E9)	Published correction studies: belief falls after a correction and partly returns at later follow-ups; partial hold is the modal outcome. No house ledger entry yet
Most likely outcome	The reduction holds at five points or more. Likely (55 to <85)
Alternatives	Decays below the minimum detectable effect without reversing: Unlikely (15 to <45). Reverses under an adversary counter-campaign: Highly unlikely (5 to <15)
Steps under E3	One: time past the last measured wave
Signposts	Supporting: hoax-term search share stays below comparison regions four consecutive weeks after May 18, B2, latest score pending. Opposing: Monitor records a new fuel-narrative variant directed at reached regions, triangulated, latest score 0
Falsifier	The eight-week wave returns reached regions to baseline
Resolution	July 13, 2027, to the calibration ledger

Table C-2. Estimate Statement, sample values.

Annex D · Data-room checklist [Tearline 3]

Item	What the auditor needs	Why
Registration	PDF, hash, escrow record, amendments with their hashes	Proves the design was fixed and unchanged
Tradecraft outputs	Key Assumptions Check, Premortem, red-team read	Shows what the team expected before launch
Random procedure record	Method, seed or lottery record, who, when	Confirms random, fixed assignment for DS-1 and DS-2
Instruments as fielded	Every item, language, and wave; back-translations	Checked item by item
Raw survey data	Respondent-level records with frame, quotas, weights, removal flags	The effect is recomputed from here
Response and attrition records	Response rates, removals, attrition by population and wave, bounds if triggered	Required statement fields
Indicator feeds	Raw data, provider documentation, C6 screen results	Graded as sources; checked against windows and leakage rules
Adversary-effort series	Weekly items by population with triangulation records and CWTR tags	Context and D-Score markers
Exposure log	Delivery by population and date	Separates reach from effect; detects contamination
Deviation log	Every departure with date and reason	Reviewed in full
Analysis code and outputs	Primary and secondary analyses; the Analysis of Competing Hypotheses matrix	Recomputation runs the registered plan
Debriefing records	Scripts, completion counts, withdrawals	Confirms Section 7.6
Estimates	Estimate Statements, signpost scores with sources	Checked against E1 to E9
Cost	Total, with basis	Cost-per-unit line
Operator’s report	As delivered to the funder	The claim under audit
Contacts	Survey house, data providers, feed owners	Each graded as a source on its own record

Table D-1. Data-room checklist.

Annex E · Audit Statement [Tearline 3]

Field	Sample entry
Header	Campaign under audit; registration identifier; client and operator; auditor and conflict declaration; data-room completion date; statement date
Verification results	Registration before launch: verified, hash matches. Design: stepped wedge as registered, DS-2. Instrument: verified. Inputs: survey house Z B2; platform search data B2; ministry information-line counts C2, interested. Recomputation: −6.3 against −6.4 reported, within rounding. Deviation log: one finding. Statement: conforms. Estimate: one issued, inside ceiling and horizon, two signposts and a falsifier registered
Findings	1, Minor: information-line window extended from seven to ten days after launch, logged; indicator reported without weight. 2, Minor: per-region response rates omitted; supplied in the draft-findings window
Recomputed effect	Belief −6.3, interval −9.1 to −3.5; search share −27 percent, matched; $29,000 per point at $183,000 over 6.3
Grading	Level 3: T-Score 82, Strong; D-Score 30, Unlikely; DS-2; ceilings Highly likely (design) and none (T); corroborated; lower governs
Determination	Granted. MOE-Audited earned; self-measured cap lifted. Eight-week estimate stands at Likely pending resolution on July 13, 2027
Declaration	Treadstone 71 LLC had no role in the design, delivery, advice, or measurement of the campaign. Signed, Half Moon Bay, California, June 30, 2027

Table E-1. Audit Statement, sample values.

Annex F · Signpost Register [Tearline 3]

Field	Guidance
Signpost ID	SP-[registration ID]-NN
Estimate	The estimate it bears on
Direction	Supports or opposes
Observable	What an observer sees, stated so two observers agree
Source and grade	Provider, dataset, anchor grade
Window	Dates during which the signpost is watched
Scoring	0 to 6 per wave, Table 11, with the source grade of each observation
Consequence	Opposing at 4 or above drops the estimate one band; supporting at 6 from B2 or better recovers one step lost under E3

Table F-1. Signpost Register.

Annex G · Calibration ledger [Tearline 3]

Field	Guidance
Estimate ID	Registration ID plus estimate number
Party	Measurer and auditor
Band issued	Band and midpoint from Table 17
Horizon	Resolution date
Outcome	Hit or miss against the most likely outcome; a non-resolution counts as a miss
Brier score	(midpoint − outcome)², outcome 1 for hit and 0 for miss
Running mean	By party, over the most recent 20 resolved estimates once ten exist; sets the cap on the ladder in Table 17a

Table G-1. Calibration ledger.

Annex H · Probability language [Tearline 3]

The protocol prints seven Kent-style terms with half-open intervals, the same set the ASA Evidence Standard uses. ICD 203 sets a different table, quoted below from its Section D.6.e(2)(a). The two sets agree at the ends and in the middle and differ on four bands, so any product sent to a reader trained on ICD 203 prints the protocol term with its percentage range every time.

Protocol term	Protocol range (percent)	ICD 203 terms	ICD 203 range
Almost certain	95 to 99	almost certain(ly); nearly certain	95-99%
Highly likely	85 to <95	very likely; highly probable	80-95%
Likely	55 to <85	likely; probable (probably)	55-80%
Roughly even chance	45 to <55	roughly even chance; roughly even odds	45-55%
Unlikely	15 to <45	unlikely; improbable (improbably)	20-45%
Highly unlikely	5 to <15	very unlikely; highly improbable	05-20%
Almost no chance	1 to <5	almost no chance; remote	01-05%

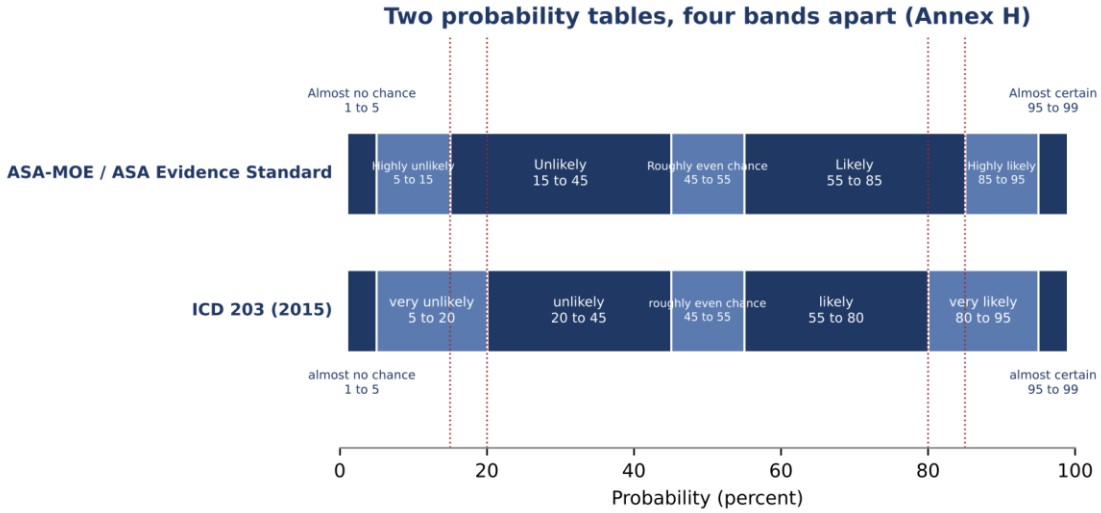


Figure 7. The two probability tables on one axis. Dotted lines mark the four boundaries where they differ: 15 against 20 at the low end, 80 against 85 at the high end. Sources: ASA Evidence Standard, Public Edition 1.2; ICD 203, Section D.6.e(2)(a); Treadstone 71.

Table H-1. Protocol bands beside ICD 203. Kent's original essay, Words of Estimative Probability (1964), is the common ancestor of the two tables.

Annex I · Change record and verification notes [Tearline 4]

Change from working draft 0.3	Reason
Estimation layer added as Component 7, with nine rules, signposts, the profile-match test, persistence arithmetic, and the calibration ledger	Measurement first, estimation second, each graded
Comparables review of 24 sources with a nine-test matrix and ranked gap findings, including JP 3-04 (2022), MCWP 8-10 (2024), the 2023 DoD information-environment strategy, and Lithuania	Shows where the protocol keeps, extends, and builds new

Change from working draft 0.3	Reason
Component 5 added: population profiles and the adversary-effort series	Context the estimation layer needs; CREST effort and exposure made graded covariates
Combination rule extended from five steps to seven: corroboration (ASA Rule 7) and attrition bounds added	Closes single-channel claims and differential-dropout bias
Analysis of Competing Hypotheses required on every causal claim	Forces rival explanations onto the page
Contamination ladder (Table 8b): none to 10 percent; attenuation correction to 30 percent; one grade off above 30; DS-5 above 50	Anchored on the Torgerson dilution formula, under which 30 percent contamination doubles the sample required
Exploratory ladder (Table 8a): Roughly even chance with a full count; Unlikely without one; no band when contradicting the registered result; the replication's ceiling when replicated	Prevents an unregistered result from passing as registered, and prices the number of analyses tried
Corroboration step restated on the Admiralty credibility axis; Almost certain requires a replication	Ties the cap to "probably true" against "confirmed"
Deception ladder for estimates (Table 16a) across all five deception bands of ASA Table 5	Replaces a single cut-off
Calibration ladder (Table 17a) with four tiers from one band to suspension	Replaces a single cut-off
Estimate rule E9, outside view first	Reference-class discipline; flagged for comment
Debriefing at each respondent's exit rather than at study end	APA 8.08: as early as feasible
Eight gates for carrying an Effect Statement into the Adversary Index (Table 24a)	States the rule the 0.3 draft only named
Respondent protection (Section 7.6) and attrition rule (Section 7.5) added	Ethics and validity gaps in 0.3
F20 mapping: DS-5 moved from 30 to 35; void added at 0	Latest mapping
D-Score floor of 25 retained	Matches ASA Public Edition 1.2
Worked example dates moved to 2027; cluster scores printed as exact means	Matches the published page; removes rounding drift (T-Score 82 either way)
Extends Public Edition 1.2 under DOI 10.5281/zenodo.23137881; license set to CC BY-NC-ND 4.0	Earlier DOI and edition withdrawn
Comment window set at 30 days from issue	Full window for reviewers

Table I-1. Change record.

Doctrine definitions as read. JP 5-0 (2020) defines the measure of effectiveness as “an indicator used to measure a current system state, with change indicated by comparing multiple observations over time” and the measure of performance as “an indicator used to measure a friendly action that is tied to measuring task accomplishment,” each sourced in its glossary to the DOD Dictionary. CJCSM 3108.01A reproduces the effectiveness definition without the word “used”; the protocol quotes JP 5-0. NATO’s agreed definitions, read in AJP-5 Edition B, AJP-10.1, and JDP 2-00: measure of effectiveness, “a criterion used to assess changes in system behaviour, capability, or operating environment, tied to measuring the attainment of an end state, achievement of an objective, or creation of an effect”; measure of performance, “a criterion that is tied to measuring task accomplishment in order to assess friendly actions.” AJP-5 Edition A, Version 2 (2019) still marked the effectiveness definition as pending NATO agreed status; Edition B marks it agreed. The withdrawn AJP-3.10.1, Edition B (2014), defined a psychological-operations measure of effectiveness as “normally expressed as an increase or decrease in a specified activity of a target audience.” The lexicon pages of AJP-5 Edition B carry a footer reading “Edition D Version 1,” a publisher error; the protocol cites them as Lex-8 of Edition B.

End of document.