

ONQ Lab Test T-002

Mass, Direction and Historical Polity Breakdown

Preregistration v1.0

Document title: **T-002 Preregistration v1.0**

1. Confirmatory claim and scope

Among historical independent states with measured material capability, the preregistered **Institutional Direction Proxy (IDP)** should contain prospective information about structural breakdown over the subsequent 50 years beyond that contained in the **Material Mass Proxy (MMP)** alone.

The primary comparison is:

- **Baseline:** MMP + calendar time.
- **PrF model:** MMP + IDP + MMP × IDP + calendar time.

PrF Mass and **PrF Direction** are ontological concepts. MMP and IDP are empirical proxies. MMP measures only the material-national-capability component of PrF Mass. IDP is a narrow operational proxy for one institutional manifestation of PrF Direction; it is not an empirical definition of PrF Direction.

T-002 does not test the total Probability Fields (PrF) ontology, prove Forward, establish causality, or establish that either proxy exhausts its corresponding ontological concept. A positive result cannot establish that IDP exhausts PrF Direction or that CINC exhausts PrF Mass. A negative result cannot be retrospectively repaired by redefining the proxies after outcomes are known.

This is a prospective predictive test, not an exercise in explaining known historical collapses. Famous cases, case identities, historical narratives and knowledge of subsequent outcomes shall not be used to reinterpret predictors or modify this specification.

2. Inferential population and frozen dataset versions

The primary inferential population comprises independent-state units jointly represented in COW National Material Capabilities, V-Dem and Polity5, beginning no earlier than 1816. Inference is primarily about modern historical sovereign states, not all human civilisations.

The intended dataset versions are:

- **COW National Material Capabilities v7.0:** MMP and its six-component reconstruction.
- **COW State System Membership v2024:** canonical state identifiers and membership dates.
- **V-Dem v16:** the four registered IDP components and the registered regional classification.
- **Polity5 v2018:** the registered breakdown outcome and entity-continuity metadata.

Newer releases shall not be silently substituted. Exact source versions and file hashes shall be recorded during subsequent execution before outcome exposure.

Deep-history Seshat replication remains separate. Seshat observations shall not be merged into the primary panel. A later replication requires its own preregistration: a polity-century unit and a 100-year primary horizon, no 25-year test, and no 50-year confirmatory test unless a separate data audit establishes adequate sub-century dating for its frozen predictors. Any assignment of Seshat

variables to empirical proxies must be frozen without using Seshat collapse cases to select them.

3. Unit of analysis and landmark calendar

The unit of analysis is the **state-decade landmark**. Predictors use information dated no later than landmark year t ; outcomes cover the subsequent prediction window.

State-decade landmarks are fixed globally to Gregorian calendar years divisible by ten: 1820, 1830, 1840,

Landmark anchors shall not be shifted according to a polity's entry date, dataset coverage, or observed outcomes. A polity entering observation after 1816 becomes eligible only at the first zero-ending landmark for which all preregistered predictor-history requirements are satisfied.

The same calendar anchor applies to the 25-, 50-, and 100-year analyses and to expanding-window temporal forecast origins. No alternative decennial anchor may be selected after outcome access.

No polity may appear in different primary cross-validation folds.

4. Prediction horizons

The **primary horizon is 50 years**, covering $t+1$ through $t+50$.

The **25-year horizon** is a preregistered short-horizon sensitivity analysis. The **100-year horizon** is a secondary long-horizon sensitivity analysis subject to the separate feasibility rules in Section 19.

No horizon may be selected, promoted or demoted according to observed performance. The 100-year sensitivity cannot replace, rescue, upgrade or overturn the primary 50-year confirmatory result.

5. Breakdown onset and entity matching

5.1 Canonical identifier

The canonical cross-dataset identifier is the Correlates of War state code and calendar year. COW NMC records will be joined to Polity5 using $CCode/CCODE \times year$ and to V-Dem using $COWcode \times year$.

Country names shall not be used to override, repair or select among conflicting code-based matches after outcome access. Any code-year mapping that is not unique and cannot be resolved mechanically under the rules below is treated as unresolved and excluded from the affected landmark analysis rather than manually adjudicated using knowledge of subsequent events. Every excluded or unresolved mapping shall be recorded in an audit table.

5.2 Breakdown event and onset

Polity5 SF is the authoritative primary outcome flag. Polity5 defines $SF=1$ for years in which the polity is in complete collapse of central authority/state failure; it also assigns $SF=1$ to state disintegration and to profound revolutionary transformations for which prior central authority is considered to have completely collapsed.

A new qualifying breakdown onset is the first year y after an eligible landmark in which $SF_y=1$ and the immediately preceding observable year in the same registered continuity lineage is not already in state failure.

Consecutive SF=1 years constitute one breakdown episode, not multiple events. A landmark already within an SF=1 episode is ineligible.

Polity5's interregnum code -77 represents complete collapse of central political authority and is entered from the year in which the interregnum begins until authority is regained or a new polity is established. The primary analysis will nevertheless use the SF flag rather than reconstructing failure independently from component codes.

Ordinary government turnover, leadership succession, limited regime change, temporary instability or declining scores do not by themselves qualify as breakdown.

For horizon H , $Y_{i,t,H}=1$ if a new qualifying breakdown begins during $t+1$ through $t+H$. Otherwise $Y=0$, provided the polity remains observable through the required window. Non-failure disappearance before complete follow-up is censoring rather than automatically survival or breakdown.

5.3 State disintegration: CHANGE=96

Polity5 defines disputed disintegration into successor states as collapse of central state authority and sets SF=1 in the final polity year. It therefore qualifies through the ordinary SF-onset rule; successor states do not create additional versions of the same event.

5.4 State transformation: CHANGE=97

Polity5 explicitly treats a transformed state as continuous with the former polity when authority continues despite a substantial territorial/state-code change. Where Polity5's transformation metadata establishes one unambiguous predecessor-successor lineage, follow-up therefore continues across the new COW code as the same analytical lineage. If such continuity cannot be resolved mechanically and uniquely, censor at the last unambiguous observation rather than manually choosing a successor.

5.5 State demise: CHANGE=98

Voluntary dissolution or incorporation into another state is not by itself a T-002 breakdown event. If no qualifying SF onset has occurred, the observation is censored at demise.

5.6 State creation: CHANGE=99

A newly created state begins a new analytical entity and becomes eligible only when it possesses the full preregistered predictor-history window.

5.7 Interruption: -66

Interruption or foreign occupation is not independently classified as breakdown. The registered SF flag determines whether the year qualifies as state failure. This avoids creating a separate analyst-defined outcome category.

5.8 Other exits and succession restrictions

A state that leaves COW/Polity observation before completion of its prediction window without a qualifying breakdown and without an explicitly registered CHANGE=97 continuity link is censored at its last unambiguous observable year.

No merger, split, renaming or territorial succession may be reclassified after examination of model performance.

6. Material Mass Proxy (MMP)

The primary empirical proxy for PrF Mass is the five-year mean Composite Index of National Capability:

$$\text{MMP}_t = \text{mean}(\text{CINC}[t-4:t]).$$

CINC measures national material capability using total population, urban population, energy consumption, iron/steel production, military personnel and military expenditure. It is treated in T-002 as a Material Mass Proxy, not as a complete operationalisation of PrF Mass.

PrF Mass is ontologically broader: the effective capacity of a structure to make particular outcomes real. T-002 measures only the material-national-capability component for which a sufficiently long historical series is available.

All references in the preregistration, analysis, figures and eventual paper shall distinguish PrF Mass from the empirical Material Mass Proxy. A result involving MMP cannot establish that CINC exhausts or fully measures PrF Mass.

The five-year mean is frozen to reduce one-year measurement noise. No component weights will be fitted against breakdown outcomes. Within each training fold, MMP will be centred and scaled using the training-set mean and standard deviation only. Test observations use the training transformation.

7. Institutional Direction Proxy (IDP)

IDP is a narrow proxy for the extent to which a polity's institutional capacity is oriented toward broadly sustaining internal conditions, incorporating adaptive feedback, administering capacity impartially and limiting diversion of public resources.

7.1 Orientation/Adaptation subscale

Public-goods allocation: $v2dlencmps_ord$, transformed as:

$$G = v2dlencmps_ord / 4.$$

Higher values indicate greater orientation of social/infrastructure expenditure toward public rather than particularistic beneficiaries. The measure concerns the composition of national-budget social and infrastructure expenditure.

Range of consultation: $v2dlconslt_ord$, transformed as:

$$A = v2dlconslt_ord / 5.$$

Higher values indicate wider consultation when major policy changes are considered.

The Orientation/Adaptation subscale is:

$$O = (G + A) / 2.$$

7.2 Stewardship/Execution subscale

Rigorous and impartial public administration: $v2clrspct_ord$, transformed as:

$$I = v2clrspct_ord / 4.$$

Higher values indicate public officials performing their duties rigorously and impartially rather than arbitrarily, nepotistically or discriminatorily.

Public-resource stewardship:

$$S = 1 - v2x_pubcorr.$$

The public-sector corruption index is coded from low to high corruption and captures corrupt exchanges and theft/misappropriation of state resources. Higher S therefore indicates less diversion of public resources.

The Stewardship/Execution subscale is:

$$E = (I + S) / 2.$$

The full institutional level is:

$$L = (O + E) / 2.$$

Orientation/Adaptation and Stewardship/Execution receive equal subscale weight. This also equal-weights the four components. The equality is a methodological restriction on researcher freedom, not a PrF assertion about intrinsic ontological weights.

7.3 Excluded components and limits

Transparent laws with predictable enforcement (v2cltrnslw) is not a primary IDP component. Alongside administration and corruption, it would make the index excessively similar to generic rule-of-law/governance quality.

The following may not be introduced into the confirmatory IDP after outcome inspection: military expenditure/projection, military personnel, political killings, civil conflict, opposition activity, coup variables, anti-system movement strength, government instability, war onset, state-failure precursors and regime-duration variables. No revolutionary, collapse or other outcome-proximal variables are permitted inside IDP.

Military and coercive allocation are not assigned an intrinsically positive or negative PrF Direction sign because the ontology does not establish such a monotonic relationship independent of context. IDP does not measure all possible PrF Direction, including whether foreign military projection was strategically Forward or Drag.

8. IDP level, trajectory and frozen weighting

For each component, calculate:

$$X_recent = \text{mean}(X[t-4:t])$$

$$X_prior = \text{mean}(X[t-9:t-5]).$$

Construct L_recent and L_prior from the four frozen components using the equal weighting above. Then:

$$\Delta L = L_recent - L_prior$$

$$IDP = ((2L_recent - 1) + \Delta L) / 2.$$

Higher values correspond to the preregistered proposition of more Forward-oriented institutional PrF Direction under this proxy.

Weighting interpretation. IDP deliberately represents **present institutional orientation conditioned by recent trajectory**, rather than trajectory alone. The formula assigns equal formal weight to two signed terms: centred current orientation, $2L_recent - 1$, and recent trajectory, ΔL . Because L_recent also forms part of the trajectory term, the expression can equivalently be written:

$$IDP = (3L_recent - L_prior - 1) / 2.$$

Recent institutional orientation therefore has greater total algebraic influence on the resulting score than prior orientation. This is deliberate: PrF Direction is operationalised here primarily as where institutional capacity is currently oriented, while recent change provides information about the trajectory through which that present orientation was reached.

This weighting is a methodological operationalisation, not a PrF claim that these quantities possess intrinsic ontological weights. No alternative level/change weighting will be selected, optimised or substituted after breakdown outcomes are accessed. No weights, windows or transformations may be altered after outcome inspection.

9. Missing-data rules

Primary analysis is complete-case at the constructed-predictor level. All four IDP components must have valid values throughout both five-year blocks required for their recent and prior values.

There shall be no outcome-dependent imputation, cross-component substitution, extrapolation beyond a component's observed historical coverage, or relaxation because sample size becomes inconvenient.

If the complete-case rules reduce the analysis below mandatory sample/event thresholds, T-002 is **Invalid / Uninterpretable**, rather than rescued through post-hoc relaxation.

9.1 Exhaustive Meaning of Measurement Failure

For purposes of Section 23 Step 1, a mandatory primary measurement condition fails only if at least one of the following occurs in the frozen named dataset versions:

1. a variable or identifier required to construct the primary MMP, IDP, calendar-time term, canonical entity linkage or registered breakdown outcome is absent;
2. the documented coding or measurement scale of a required primary variable is incompatible with the frozen transformation and therefore prevents mechanical construction of that registered quantity;
3. the source does not provide the temporal resolution required to construct the registered annual predictor windows or outcome timing; or
4. a required primary identifier or outcome field cannot be interpreted according to the registered source schema, preventing mechanical implementation of the primary analysis.

No additional qualitative measurement-validity judgment may activate **Invalid / Uninterpretable**.

Ordinary observation-level missingness, unresolved code-year mappings, excluded records, censoring and individual invalid/malformed values shall be handled through the registered missing-data, entity-matching, exclusion and feasibility rules. They do not independently constitute global measurement failure.

Concerns about historical data quality, retrospective coding, construct validity, proxy incompleteness or substantive interpretation shall be reported as limitations but cannot activate **Invalid / Uninterpretable** after outcome access unless they satisfy one of the explicit mechanical conditions above.

Section 23's reference to failure of a mandatory measurement condition refers exclusively to the conditions enumerated in this section.

10. Primary models

Both models use exactly the same observations. They receive identical linear centred calendar-year nuisance adjustment, denoted T . This adjustment receives no evidentiary credit for PrF.

Baseline:

$$\text{logit}(P(Y=1)) = \alpha + \beta_M \text{MMP} + \beta_T T.$$

PrF model:

$$\text{logit}(P(Y=1)) = \alpha + \beta_M \text{MMP} + \beta_D \text{IDP} + \beta_{MD}(\text{MMP} \times \text{IDP}) + \beta_T T.$$

Standardisation and preprocessing are learned within training samples only. There shall be no stepwise selection, outcome-optimised weights, post-hoc nonlinearities or addition of variables because they improve fit.

10.1 Model Estimability and Non-Convergence

All registered logistic models are unpenalized maximum-likelihood logistic regressions with the frozen covariate specification.

A required model fit is estimable only when all of the following hold:

1. the training outcome contains both outcome classes;
2. the model matrix has full column rank;
3. complete or quasi-complete separation does not prevent existence of a finite maximum-likelihood estimate;
4. the fitting procedure converges;
5. all required fitted coefficients and linear predictors are finite; and
6. all required held-out predictions and registered evaluation quantities can be calculated finitely.

Log loss shall be calculated using a numerically stable logistic-loss implementation from the fitted linear predictor. Analyst-selected probability clipping is prohibited.

No Firth correction, ridge/lasso penalty, Bayesian prior, covariate removal, alternative link function or other fallback estimator may be introduced in response to separation or non-convergence.

For the primary grouped-cross-validation analysis, if either the baseline or full-IDP model is non-estimable in any required fold, the required primary pooled quantities are non-estimable and T-002 is classified Invalid / Uninterpretable through Section 23.

For a construct-discrimination analysis, non-estimability of a quantity required by Sections 21.3–21.5 makes Supported unavailable as already registered, but does not independently invalidate an otherwise interpretable primary analysis.

For a registered robustness analysis, non-estimability of any fit required to produce its registered pooled result makes that robustness analysis unavailable. It cannot trigger a robustness veto, count as favourable evidence or independently invalidate the primary analysis.

11. Construct-discrimination landmark construction and models

The Orientation/Adaptation and Stewardship/Execution diagnostic constructs preserve the same level-plus-trajectory architecture used in the full IDP.

For each eligible landmark:

$$\Delta O = O_{\text{recent}} - O_{\text{prior}}$$

$$O^* = ((2O_{\text{recent}} - 1) + \Delta O) / 2,$$

where O_{recent} is the mean Orientation/Adaptation subscale during t-4:t and O_{prior} is its mean during t-9:t-5.

Likewise:

$$\Delta E = E_{\text{recent}} - E_{\text{prior}}$$

$$E^* = ((2E_{\text{recent}} - 1) + \Delta E) / 2,$$

where E_{recent} is the mean Stewardship/Execution subscale during t-4:t and E_{prior} is its mean during t-9:t-5.

The preregistered diagnostic models are:

- $MMP + O^* + MMP \times O^* + \text{calendar time}.$
- $MMP + E^* + MMP \times E^* + \text{calendar time}.$

The same missing-data rules, training-only standardisation, grouped-cross-validation procedure and expanding-window temporal-validation procedure used for the full IDP analysis apply to these diagnostic models.

Their purpose is construct discrimination. They differ from the full IDP in component content, not in temporal construction. They cannot replace the primary IDP specification or become alternative confirmatory models because their performance is more favourable.

12. Primary substantive contrast and pooling

At MMP equal to +1 training-set SD, increasing IDP from -1 SD to +1 SD is predicted to reduce 50-year breakdown probability. The interaction coefficient itself is descriptive and has no independently imposed sign.

For $X = \text{IDP}, O^* \text{ or } E^*$, define the held-out high-Mass probability contrast:

$$\Delta P_X = P(Y=1 \mid MMP=+1 \text{ SD}, X=+1 \text{ SD}) - P(Y=1 \mid MMP=+1 \text{ SD}, X=-1 \text{ SD}).$$

The preregistered PrF Direction prediction corresponds to $\Delta P_X < 0$.

For every held-out landmark t belonging to polity i , calculate:

$$\delta_{it}^X = P(Y=1 \mid MMP=+1 \text{ SD}, X=+1 \text{ SD}, T=t) - P(Y=1 \mid MMP=+1 \text{ SD}, X=-1 \text{ SD}, T=t).$$

MMP and X are fixed at +1/-1 on the relevant training-set-standardised scale. Calendar time is set to the held-out landmark's actual year and transformed using only the calendar-time centring learned from that training set. The relevant interaction term is recomputed from the imposed predictor values.

For grouped cross-validation, first take the arithmetic mean of δ_{it}^X across all held-out landmarks belonging to each polity. Then take the unweighted arithmetic mean of those polity-level contrasts across all held-out polities. Each polity therefore contributes equal final weight regardless of its number of eligible landmarks.

For expanding-window temporal validation, calculate the contrast at the actual forecast-origin year, average equally across eligible polities within that origin, and then average the resulting origin-level contrasts equally across all usable forecast origins.

No alternative calendar-time setting or landmark/polity/origin weighting scheme may be selected after outcome access.

13. Primary validation

The primary out-of-sample procedure is **five-fold grouped cross-validation by polity**. All observations belonging to one polity remain in one fold. Row-level random cross-validation is prohibited.

Fold assignment may be stratified by eventual event occurrence only after the entire predictor specification has been locked, solely to prevent folds containing essentially no events.

All transformations, standardisation and preprocessing are calculated from training data only. Primary performance is calculated on pooled held-out predictions, never in-sample fit.

14. Strict expanding-window temporal validation

Expanding-window forecast-origin validation is required for **Supported**.

At historical forecast origin C , training may include landmark s only if:

$$s + 50 \leq C.$$

The complete 50-year outcome attached to every training observation must therefore already have been observable at historical time C . The model is then applied to predictor observations at landmark C . Later historical outcomes shall not be used to fit a model purporting to predict an earlier period.

Transformations are fitted on the historically available training sample only. Use successive decennial origins on the registered calendar anchor wherever complete 50-year follow-up exists. No origin may be selected or removed according to model performance.

Grouped polity cross-validation remains the main performance estimate and tests unseen-polity generalisation. Temporal validation assesses historical forward transportability.

14.1 Usable Temporal Forecast Origin

Every zero-ending forecast origin satisfying the registered calendar rule and possessing complete 50-year outcome follow-up shall be attempted.

A candidate forecast origin C is usable only when:

1. the registered training sample satisfying $s + 50 \leq C$ can be constructed;
2. at least one eligible test landmark exists at C ;
3. the baseline training model is estimable under the registered estimability rule;
4. the full-IDP training model is estimable under that rule; and
5. both models produce finite predictions for every included test landmark.

A candidate origin failing any condition is logged as unavailable and excluded from pooled temporal statistics. It may not be replaced by a different or shifted forecast date.

The existing temporal feasibility requirements remain unchanged:

- at least three usable forecast origins; and
- at least 30 distinct qualifying breakdown onsets represented across the pooled temporal test predictions.

Failure of either condition produces Invalid / Uninterpretable.

14.2 Temporal classification conditions

Temporal classification conditions. Supported requires both the registered pooled temporal high-Mass contrast $\Delta P_T < 0$ and pooled temporal relative log-loss improvement $R_T > 0$.

“Directionally consistent” means $\Delta P_T < 0$. “Opposite temporal direction” or “clearly in the opposite direction” means $\Delta P_T > 0$. “Material temporal deterioration” or “materially worsening” means $R_T \leq -0.02$. “Strong temporal contradiction” means both $\Delta P_T > 0$ and $R_T \leq -0.02$.

Pooled temporal results alone govern these conditions. Individual forecast origins are reported but cannot override pooled results. No temporal uncertainty interval is required or used for classification.

Failure of either Supported condition, including equality at zero, prevents Supported. Temporal contradiction does not independently assign Contradicted. The final evidence category follows Section 23.

The existing temporal feasibility thresholds remain mandatory. Their failure produces Invalid / Uninterpretable. Required pooled temporal quantities that cannot be estimated also produce Invalid / Uninterpretable.

15. Evaluation metrics and material improvement

The primary predictive metric is **held-out log loss**. Secondary metrics are Brier score, precision-recall AUC, AUROC, calibration intercept and calibration slope. AUROC alone is insufficient.

Loss is also summarised at polity level to prevent long-observed polities from dominating purely through row count. Uncertainty uses the polity-cluster bootstrap in Section 16.

For the full IDP model and each construct-discrimination model X, define:

$$R_X = (\text{LogLoss_Baseline} - \text{LogLoss_X}) / \text{LogLoss_Baseline}.$$

The material primary improvement threshold is $R_{IDP} \geq 0.02$: at least a 2% relative held-out log-loss reduction. For Supported, the primary polity-cluster-bootstrap 95% interval for improvement must exclude zero. The 2% threshold is methodological, not ontological.

16. Polity-cluster bootstrap

Registered polity-cluster-bootstrap confidence intervals shall use **2,000 attempted full-refit bootstrap resamples**, subject to Section 16.1.

Polities shall be resampled with replacement within each of the frozen original grouped-cross-validation folds. All landmark observations belonging to a sampled polity travel together, and bootstrap multiplicity is preserved. A source polity therefore remains confined to its original cross-validation fold.

Within every bootstrap replicate, all training-dependent preprocessing and standardisation shall be recalculated and the baseline and relevant PrF model shall be refitted from scratch for each held-out fold. New held-out predictions shall then be generated and all registered bootstrap quantities, including ΔP and relative log-loss improvement R, recomputed according to the frozen aggregation rules.

Resampling fixed predictions without model refitting is prohibited. Registered 95% bootstrap intervals are percentile intervals defined by the 2.5th and 97.5th percentiles of the successful replicate estimates, subject to Section 16.1.

16.1 Bootstrap Failure Rule

The registered polity-cluster bootstrap consists of exactly 2,000 attempted resamples.

Failed resamples shall not be replaced.

For any registered bootstrap quantity, a replicate is successful only if every model fit and prediction required to calculate that quantity is estimable under the registered rules.

The registered percentile 95% interval for a quantity shall be calculated from its successful replicate estimates only when at least 1,900 of the 2,000 attempted replicates are successful.

If fewer than 1,900 successful replicate estimates exist, that bootstrap interval is non-estimable.

No failed replicate shall be assigned an artificial numerical value.

No additional resamples shall be drawn to replace failures.

Consequences of a non-estimable interval follow the existing evidence architecture:

- a required primary grouped-CV interval being non-estimable makes the primary classification Invalid / Uninterpretable;
- a required construct-discrimination interval being non-estimable makes Supported unavailable but does not independently invalidate an otherwise interpretable primary analysis;
- a robustness-analysis quantity becoming non-estimable makes that robustness analysis unavailable.

The same registered full-refit resampling architecture otherwise remains unchanged.

17. Primary feasibility thresholds

Before inspecting model performance, the primary analysis requires all of:

- ≥ 600 predictor-complete state-decade landmarks.
- ≥ 60 distinct polities.
- ≥ 80 distinct qualifying breakdown onsets.
- Qualifying breakdown onsets across ≥ 30 distinct polities.
- ≥ 10 distinct qualifying breakdown onsets in every grouped-CV test fold.

Events are counted as distinct historical breakdown onsets, not positive landmark rows. One breakdown event causing several overlapping 50-year windows to receive $Y=1$ is counted once for feasibility, not as several independent events.

The 80-event threshold is a conservative feasibility floor, not a guarantee of adequate statistical power. Failure of any mandatory threshold produces **Invalid / Uninterpretable**, without later relaxation or redesign. A favourable coefficient cannot convert an underpowered test into Compatible.

18. Temporal-validation feasibility thresholds

The expanding-window temporal assessment requires:

- At least three usable historical forecast origins.
- At least 30 distinct breakdown onsets represented across pooled temporal test predictions.

The same historical breakdown onset is counted once even if it falls inside multiple test windows. Failure makes the strict prospective component insufficiently resolved and the overall prospective claim **Invalid / Uninterpretable**.

19. 100-year sensitivity feasibility rule

The preregistered 100-year analysis is a secondary sensitivity analysis and cannot replace, rescue, upgrade or overturn the primary 50-year confirmatory result.

It is interpretable only if all of the following are satisfied:

- At least 400 predictor-complete state-decade landmarks.
- At least 50 distinct polities.
- At least 60 distinct qualifying breakdown onsets.
- Qualifying breakdown onsets across at least 25 distinct polities.
- At least 8 distinct qualifying breakdown onsets in every held-out grouped-cross-validation fold.

It uses the same five-fold grouped-cross-validation-by-polity procedure as the primary analysis.

These thresholds are deliberately lower than the primary 50-year thresholds because requiring 100 years of complete forward observation mechanically removes a substantially larger share of otherwise eligible historical landmarks. The reduced thresholds are a feasibility concession to follow-up length, not a relaxation based on observed T-002 performance.

The 60-event threshold remains an outcome-blind conservative floor for the registered model complexity and is not a formal guarantee of statistical power.

No separate expanding-window temporal-validation requirement applies to the 100-year sensitivity. The registered 50-year expanding-window analysis remains the required temporal/prospective condition for Supported.

Any later 100-year temporal analysis not specified here must be labelled exploratory and cannot alter the registered T-002 evidence category.

20. Precision safeguard

If the polity-cluster-bootstrap 95% interval for primary relative held-out log-loss improvement simultaneously includes $R_IDP \leq -0.02$ and $R_IDP \geq +0.02$, the analysis is too imprecise to distinguish a materially helpful IDP signal from a materially harmful one.

That result is **Invalid / Uninterpretable due to insufficient precision**. It may not be labelled Compatible merely because the point estimate is positive.

21. Deterministic evidence and construct-discrimination rules

Classification authority and numerical conventions. Evidence classification shall follow the ordered hierarchy in Section 23. Earlier qualitative descriptions do not create additional classification conditions. “Meaningful predictive relationship” means, solely for distinguishing Compatible from Not Supported after prior hierarchy conditions have been excluded, a primary pooled grouped-CV contrast below zero together with relative log-loss improvement above zero. “Sufficiently uncertain” introduces no additional judgment: uncertainty affects classification only through the explicitly registered interval comparisons. All comparisons use unrounded estimates. Equality does not satisfy a strict inequality.

21.1 Robustly opposite

The primary IDP relationship is robustly opposite when $\Delta P_IDP > 0$ and the polity-cluster-bootstrap 95% confidence interval for ΔP_IDP lies entirely above zero. An unexpected sign on an individual regression coefficient is insufficient.

21.2 Material and credible deterioration

Adding IDP produces material and credible deterioration when $R_IDP \leq -0.02$ and the polity-cluster-bootstrap 95% confidence interval for R_IDP lies entirely below zero. This is the symmetric harmful counterpart of the registered 2% material-improvement threshold.

21.3 Corresponding prospective signal

A diagnostic construct X provides a corresponding prospective signal only when all of the following hold:

1. Under grouped cross-validation, $\Delta P_X < 0$.
2. Under grouped cross-validation, $R_X > 0$.
3. Under expanding-window temporal validation, the pooled high-Mass probability contrast is < 0 .
4. Under expanding-window temporal validation, pooled relative log-loss improvement is > 0 .

No additional magnitude or statistical-significance threshold is imposed. Zero is the boundary because this diagnostic establishes the presence and direction of corresponding information, not an additional confirmatory significance test.

21.4 Good-governance exclusion

Define the good-governance exclusion as present when E^* satisfies all four corresponding-signal conditions in Section 21.3 and O^* does not.

When present, this exclusion prevents Supported. The final category is assigned through Section 23; the exclusion alone cannot assign Not Supported, Contradicted or Invalid / Uninterpretable.

This rule does not establish that Stewardship/Execution causally accounts for the full-IDP result. It identifies insufficient predictive distinction from the stewardship/good-governance alternative.

21.5 Orientation/Adaptation robust-opposition veto

The Orientation/Adaptation diagnostic is robustly opposite when its pooled grouped-CV high-Mass contrast satisfies $\Delta P_O^*,G > 0$ and the lower endpoint of its registered 95% full-refit polity-cluster-bootstrap interval is strictly above zero.

Use the registered 2,000 grouped-CV bootstrap replicates. Temporal opposition is not additionally required.

This condition is distinct from failure of the corresponding-signal rule. Either this veto or the Section 21.4 exclusion prevents Supported. Neither independently assigns Contradicted or Invalid / Uninterpretable.

If a required diagnostic contrast, improvement or grouped-CV contrast interval is non-estimable, Supported is unavailable; Section 23 assigns the final category.

22. Registered robustness analyses

The registered analyses are:

- 25-year outcome horizon.
- 100-year outcome horizon, subject to Section 19.
- Alternative equal-weight six-component MMP reconstruction, specified below.
- IDP level only: $2L_{\text{recent}} - 1$.
- IDP change only: ΔL .
- Leave-one-IDP-component-out: four analyses.
- Pre-1900 exclusion, reflecting weaker early historical measurement.
- Regional blocking, specified below.
- Orientation/Adaptation-only and Stewardship/Execution-only diagnostic models, specified in Section 11.

Role of robustness in classification. Verdict-relevant robustness analyses comprise the 25-year horizon, six-component MMP reconstruction, IDP level-only and change-only analyses, each of the four leave-one-IDP-component-out analyses, pre-1900 exclusion and pooled regional blocking.

For each estimable analysis a , calculate its registered pooled high-Mass contrast ΔP_a and relative log-loss improvement R_a against its corresponding baseline on the same observations.

An adverse result satisfies $\Delta P_a \geq 0$ OR $R_a \leq 0$.

A major contradiction satisfies $\Delta P_a > 0$ AND $R_a \leq -0.02$.

One major contradiction is sufficient to prevent Supported. Multiple adverse results that individually fail this conjunction do not collectively constitute a major contradiction. No additional confidence-interval condition applies to this robustness veto.

For regional blocking, only the registered pooled regional statistics can trigger the veto. Individual-region results cannot override pooled results.

An infeasible or non-estimable robustness analysis shall be reported as unavailable. It cannot trigger this veto, count as favourable evidence or independently invalidate the primary analysis. Primary invalidity rules remain independently applicable.

The robustness veto alone cannot assign Contradicted. Its effect is determined by the ordered hierarchy in Section 23.

The 100-year sensitivity cannot change the final category under any result or feasibility status. The O^* and E^* diagnostics affect classification only through Sections 21.3–21.5.

Robustness analyses cannot replace the primary specification or be promoted because they produce more favourable results.

22.1 Alternative six-component MMP reconstruction

Reconstruct material capability directly from the six COW NMC components. For component k , state i and year y :

$$\text{share}_{iky} = \text{value}_{iky} / \sum_j(\text{value}_{jky}),$$

where the denominator is the system total for that component and year across non-missing COW observations. Dataset-defined missing codes shall be converted to missing before arithmetic. All six component shares must be observed for the focal state-year.

Define:

$$\text{MMP6}_{iy} = \text{mean}(\text{the six component shares})$$

$$\text{MMP6}_t = \text{mean}(\text{MMP6}[t-4:t]).$$

All five annual MMP6 observations are required. MMP6 is standardised using training data only. In this robustness analysis it replaces MMP in both the MMP main effect and the MMP $\times$ IDP interaction. All other predictors, transformations, outcomes, validation rules and evidence rules remain unchanged.

22.2 Regional-blocking robustness

Regional blocking uses V-Dem v16 variable `e_regionpol_6C` and its published six-category mapping:

- 1: Eastern Europe and Central Asia.
- 2: Latin America and the Caribbean.
- 3: Middle East and North Africa.
- 4: Sub-Saharan Africa.
- 5: Western Europe and North America.
- 6: Asia and Pacific.

Each registered polity continuity lineage is assigned the non-missing region code present at its first eligible primary landmark. If later eligible observations in the same lineage contain a different region code, that lineage shall be excluded from regional-blocking robustness and the discrepancy logged rather than manually adjudicated.

Regional validation uses six leave-one-region-out folds. In each fold, all eligible polities in one region are held out and the model is trained on the other five regions. All training-dependent transformations are estimated exclusively from the training regions. Region-specific results shall be reported individually.

For pooled regional probability contrasts, first average across landmarks within polity, then equally across polities within region, then equally across the six held-out regions.

For pooled regional log loss, first average landmark-level log losses within polity, then equally across polities within region, then equally across the six regions.

No alternative region classification, regional regrouping or region-specific exclusion may be selected after outcome access.

22.3 Leave-One-IDP-Component-Out Construction

The four registered leave-one-IDP-component-out robustness analyses shall preserve the frozen two-subscale architecture of IDP.

When Public-goods allocation G is omitted:

$$O_{-G} = A.$$

When Range of consultation A is omitted:

$$O_{-A} = G.$$

When Rigorous and impartial public administration I is omitted:

$$E_{-I} = S.$$

When Public-resource stewardship S is omitted:

$$E_{-S} = I.$$

The unaffected subscale remains unchanged.

For omitted component j, define:

$$L_{-j} = (O_{-j} + E_{-j}) / 2.$$

Thus Orientation/Adaptation and Stewardship/Execution retain equal subscale weight even when one constituent indicator is omitted. The three remaining indicators are not reweighted to equal one-third weights.

For each omission, construct recent and prior values using the same registered windows:

$$L_{\text{recent},j} = \text{value for } t-4:t$$

$$L_{\text{prior},j} = \text{value for } t-9:t-5.$$

Then define:

$$\Delta L_{-j} = L_{\text{recent},j} - L_{\text{prior},j}$$

and:

$$IDP_{-j} = ((2L_{\text{recent},j} - 1) + \Delta L_{-j}) / 2.$$

All four leave-one-component-out analyses use the exact primary 50-year analytic landmark sample and the frozen primary polity-fold assignments. Removing an indicator may not admit observations excluded from the primary analysis because that indicator was missing.

These analyses therefore vary component content only; they do not simultaneously alter the subscale architecture, temporal transformation, or analytic sample.

22.4 Feasibility of Verdict-Relevant Robustness Analyses

Unless a robustness analysis has a separately registered feasibility rule, a verdict-relevant robustness analysis is estimable for classification purposes only when it satisfies all of the following secondary-analysis feasibility thresholds:

- at least 400 predictor-complete landmarks;
- at least 50 distinct polities;
- at least 60 distinct qualifying breakdown onsets;
- qualifying breakdown onsets across at least 25 distinct polities; and
- whenever five-fold grouped cross-validation is used, at least 8 distinct qualifying breakdown onsets in every held-out fold.

These thresholds are the same outcome-blind secondary-analysis floor already registered for the 100-year sensitivity and are used to avoid introducing different feasibility thresholds for different secondary analyses.

The 25-year horizon uses the exact primary predictor-complete landmark set, with the outcome redefined only over $t+1:t+25$, and must satisfy the secondary event thresholds above.

The IDP level-only, IDP change-only and four leave-one-component-out analyses use the exact primary analytic sample and frozen primary folds. Because they do not admit additional observations, the primary sample-size and polity gates need not be retested; their models must remain estimable.

The MMP6 robustness analysis begins from the primary analytic sample and may only lose observations because of its stricter six-component completeness rule. It cannot add observations excluded from the primary analysis. It must satisfy the secondary-analysis feasibility thresholds above.

The pre-1900-exclusion analysis is a subset of the primary analytic sample and must satisfy the secondary-analysis feasibility thresholds above.

Regional blocking begins from the primary analytic sample, subject only to the already registered region-lineage consistency exclusions. The resulting regional-blocking sample must satisfy at least 400 landmarks, 50 polities, 60 distinct qualifying breakdown onsets and breakdown onsets across at least 25 polities. All six leave-one-region-out model fits must be estimable. The five-fold ≥ 8 events per held-out fold condition does not apply because regional blocking does not use the five-fold grouped-CV architecture.

If a verdict-relevant robustness analysis fails its applicable feasibility or estimability conditions, it is reported as unavailable and has no effect on the final evidence category.

The separately registered 100-year sensitivity remains governed exclusively by Section 19 and cannot affect the final evidence category.

23. Evidence categories: exhaustive ordered hierarchy

Define ΔP_G and R_G as the registered pooled grouped-CV IDP contrast and relative log-loss improvement. Define ΔP_T and R_T as their registered pooled temporal counterparts. Let $L(Z)$ and $U(Z)$ denote the endpoints of the registered 95% full-refit grouped-CV bootstrap interval for quantity Z .

Apply the following steps in order. Stop at the first assigned category. No later condition can override an earlier assignment.

1. Invalid / Uninterpretable. Assign if any registered mandatory primary feasibility, temporal-feasibility, measurement, missing-data or leakage condition fails; if required primary or pooled temporal classification quantities cannot be estimated; or if $L(R_G) \leq -0.02$ AND $U(R_G) \geq +0.02$. Unavailability of a robustness or construct-discrimination analysis alone does not activate this category.

2. Contradicted. Otherwise assign if either $\Delta P_G > 0$ AND $L(\Delta P_G) > 0$, or $R_G \leq -0.02$ AND $U(R_G) < 0$.

3. Not Supported. Otherwise assign if $\Delta P_G \geq 0$ OR $R_G \leq 0$.

4. Supported. Otherwise assign only if all of the following hold: $R_G \geq 0.02$; $L(R_G) > 0$; $\Delta P_T < 0$; $R_T > 0$; required construct-discrimination quantities are estimable; the Section 21.5

Orientation/Adaptation veto is absent; the Section 21.4 good-governance exclusion is absent; and no Section 22 major robustness contradiction is present.

5. Compatible. Assign to every remaining result.

The primary contrast's interval need not exclude zero for Supported. Its favourable point-estimate sign is required; the primary predictive-improvement interval must satisfy the explicit condition above.

These conditions are exhaustive and mutually exclusive after precedence is applied. No qualitative assessment of significance, meaningfulness, consistency, robustness or evidentiary strength may add to or override this hierarchy.

24. Ontology-revision rules

A **Not Supported** T-002 remains publicly preserved and may not subsequently be cited as evidence that PrF Direction predicts civilisation-scale breakdown.

A **Contradicted** result triggers explicit revision status for the civilisation/polity application of PrF Direction if methodological review concludes that the IDP operationalisation was defensible.

Independent replication of the contradiction using a defensible operationalisation requires revision of the claimed scope of PrF Direction/Forward, the mapping from ontology to polity structure, or ultimately the ontology itself if the failure generalises beyond the proxy.

A failure cannot be answered by declaring afterward that the failed IDP was not the real PrF Direction and quietly substituting variables that fit historical outcomes. The proxy may be argued to be incomplete, but the registered T-002 result remains failed and cannot be rescued through reinterpretation.

25. Frozen execution brief and outcome-access boundary

Execute this specification exactly as written. Do not modify predictors, outcomes, horizons, thresholds, weighting, exclusion rules, validation procedures or evidence categories after accessing breakdown outcomes. If an implementation problem makes the registered analysis impossible, stop and document the issue rather than silently altering the design.

Before reading or analysing Polity5 SF values:

- Verify the document is labelled T-002 Preregistration v1.0.
- Preserve its exact Markdown/plain-text and PDF versions.
- Verify their substantive equivalence and record SHA-256 hashes for both.
- Record the immutable or version-preserving registration reference and timestamp.
- Record exact dataset versions and source-file hashes during execution.
- Implement predictor construction, outcome-independent landmark eligibility and the COW-code entity crosswalk without using SF.
- Produce an execution log confirming these steps were completed before outcome exposure.

Only after the preregistration, source hashes, predictor pipeline and identifier procedure have been locked may Polity5 SF be exposed. Apply continuity, onset and censoring rules mechanically. Do not inspect individual breakdown cases to alter definitions, exclusions or predictor construction.

After outcome construction and before confirmatory model fitting, test the registered feasibility thresholds. If any mandatory threshold fails, stop confirmatory interpretation and classify the test

Invalid / Uninterpretable. Do not relax thresholds.

Preserve the frozen preregistration and hashes, source provenance and hashes, complete reproducible analysis code, entity-crosswalk audit, inclusion/exclusion flow, sample and event threshold report, grouped-CV results, temporal-validation results, construct-discrimination results, registered robustness results, uncertainty estimates, final evidence category and implementation-deviation log. Explicitly report zero if there were zero deviations.

If an implementation ambiguity cannot be resolved mechanically from this specification, stop before making the choice and document it. Do not select whichever implementation gives the more favourable T-002 result.

Any additional exploratory analysis must occur after the confirmatory verdict has been frozen, must be labelled exploratory and cannot alter T-002's registered evidence category.

Freeze Checklist

Before any breakdown outcome is accessed or analysed, confirm every item below. Checklist entries are requirements to verify; their inclusion does not certify that registration or subsequent execution steps have already occurred.

Specification

- [] Document title is **T-002 Preregistration v1.0**.
- [] Material capability is consistently described as **Material Mass Proxy (MMP)**, not as PrF Mass itself.
- [] Institutional measurement is consistently described as **Institutional Direction Proxy (IDP)**, not as PrF Direction itself.
- [] MMP construction is frozen.
- [] All four IDP components are frozen.
- [] IDP transformations and component directions are frozen.
- [] IDP level and trajectory windows are frozen.
- [] IDP formula is frozen as $IDP = ((2L_{recent} - 1) + \Delta L) / 2$, where $\Delta L = L_{recent} - L_{prior}$.
- [] The preregistration explicitly states that IDP represents **present institutional orientation conditioned by recent trajectory**.
- [] No alternative level/change weighting may be selected after outcome access.
- [] Primary 50-year prediction horizon is frozen.
- [] 25-year and 100-year secondary/sensitivity rules are frozen.
- [] Breakdown definition is frozen.
- [] Polity5 SF onset rule is frozen.
- [] COW-code × year entity matching is frozen.
- [] CHANGE = 96/97/98/99, -66, succession and censoring rules are frozen.
- [] Inclusion and exclusion rules are frozen.
- [] Missing-data rules are frozen.
- [] Calendar-time adjustment is frozen.
- [] Baseline model is frozen.
- [] PrF model is frozen.
- [] High-Mass substantive Direction contrast is frozen.
- [] Grouped-CV procedure is frozen.
- [] Expanding-window temporal validation is frozen.
- [] Evaluation metrics are frozen.
- [] 2% material log-loss improvement threshold is frozen.
- [] Precision safeguard is frozen.
- [] Construct-discrimination analyses are frozen.
- [] Robustness analyses are frozen.
- [] Evidence categories are frozen.

- [] Ontology-revision rules are frozen.

Feasibility thresholds

Confirm that the preregistration requires, without later relaxation:

- [] ≥ 600 predictor-complete state-decade landmarks.
- [] ≥ 60 distinct polities.
- [] ≥ 80 distinct qualifying breakdown onsets.
- [] Breakdown onsets across ≥ 30 distinct polities.
- [] ≥ 10 distinct breakdown onsets in every grouped-CV test fold.
- [] Temporal-validation minimum thresholds exactly as registered.

Failure of mandatory thresholds must produce **Invalid / Uninterpretable**, not a redesigned analysis.

Immutable freeze

- [] Export the exact final document as **Markdown/plain text**.
- [] Export the same document as **PDF**.
- [] Verify that the Markdown/text and PDF contain the same substantive preregistration.
- [] Calculate a **SHA-256 hash** for the Markdown/text file.
- [] Calculate a **SHA-256 hash** for the PDF.
- [] Record both hashes in a separate freeze record.
- [] Timestamp/register the frozen preregistration in an immutable or version-preserving repository.
- [] Record the registration URL/reference and timestamp.
- [] Record the exact dataset versions named for execution.
- [] Preserve the frozen execution brief alongside the preregistration.

Outcome-access boundary

Before Polity5 breakdown outcomes are exposed:

- [] Frozen preregistration exists.
- [] Both file hashes have been recorded.
- [] Registration/timestamp has been recorded.
- [] Predictor definitions are locked.
- [] MMP/IDP construction code can be implemented without SF.
- [] Entity-crosswalk rules are locked.
- [] Inclusion/exclusion logic independent of breakdown outcomes is locked.

Only after all items above are satisfied may execution access the registered breakdown indicator.

Post-freeze rule

After outcome access:

No predictors, transformations, weights, horizons, thresholds, outcome definitions, exclusion rules, validation procedures, or evidence categories may be altered because of observed T-002 performance.

If the frozen specification contains an implementation ambiguity that cannot be resolved mechanically:

Stop, document the ambiguity, and preserve the preregistered test rather than selecting an implementation according to which result is more favourable to PrF.

Any analysis added after the confirmatory verdict must be labelled **exploratory** and cannot alter T-002's registered evidence category.