

The Democracy Alarm

Research and political reasoning



Democratic backsliding, conceptual validity, aggregation and institutional implementation

Created by Yoav Duman

Start without advance reading

You can start here without reading a journal article beforehand. The two article explanations inside this assignment provide the research you need. Read them as part of the activity and return to them while answering. The original articles and further reading are available for optional follow-up.

Research words in everyday language

A claim is a conclusion you want to check. A hypothesis is a possible explanation. A mechanism is the sequence of steps through which something happens: who acts, what they do, and how that could produce a result.

A rival explanation is another plausible cause of that same result. A prediction says what you would expect to see if an explanation were right. Your job is to compare those predictions with evidence, not guess which opinion your instructor prefers.

An outcome is the result being measured. The population is the group you want to understand. The sample is the people actually studied. The unit of observation is who or what each record describes; a number's unit tells you whether it is a count, a percentage, hours, or score points.

The numerator is the number being counted. The denominator is the whole group used to calculate a percentage: 8 out of 20 means 8 divided by 20. A percentage-point difference subtracts one percentage from another.

An association means two things appear together in the data. It does not by itself tell you which caused which. Random assignment puts people into conditions by chance; a comparison then asks whether changing a feature changes the measured result.

A study's limits tell you what its evidence cannot establish. An uncertainty interval is a range describing how uncertain an estimate is under the analysis used. When an interval for a difference includes zero, it does not clearly distinguish a difference from no difference. That does not prove the true effect is exactly zero.

You do not need to memorize technical names or article statistics. Use the explanations, worked examples and records on screen. In your answer, make the reasoning visible: a specific detail, what it supports, and what remains uncertain.

The research question

These two 2024 articles belong to a published controversy in PS: Political Science & Politics. Little and Meng question strong claims of recent global decline using relatively observable indicators; Knutsen and colleagues respond directly by challenging measurement choices and testing proposed expert bias. Their disagreement offers a live problem about evidence, definitions, and the scope of a conclusion.

This is a genuine empirical and conceptual dispute. Little and Meng argue that major increases in successful entrenchment should ordinarily register in electoral performance; they regard changing expert judgment as a likely contributor to the discrepancy, while acknowledging omitted dimensions and without directly measuring the bias contribution. Knutsen and colleagues question whether those indicators cover meaningful competition and test specific implications of a shared pessimism account. Their tests find little supporting evidence. A sentence saying “both measure different things” preserves a useful distinction but leaves the inferential disagreement unresolved. Bellara cannot settle a worldwide trend.

This packet contains the complete teaching explanations for the two core articles and links to the primary texts. It keeps article findings separate from invented local cases. Read the primary texts when checking a citation, design detail or contested inference.

Article A Measuring Democratic Backsliding

Andrew T. Little and Anne Meng (2024). PS: Political Science & Politics. 57(2): 149–161 · DOI 10.1017/S104909652300063X

Dates

Published online 11 January 2024; issue April 2024. Most graphs cover 1980–2021, subject to source availability; the electoral event source extends through 2020.

Access

Open-access complete publisher text · CC BY.

Lens

If leaders are increasingly insulating themselves from defeat, why do governments still lose elections at similar rates? Compare alarming narratives with indicators that require relatively little judgment.

Read first

Start with the explanation on this screen. Then check what was compared, what was found, and what remains uncertain. The journal article link is optional.

Sections

Read What We Do; The Search for More Objective Measures; Taking Stock; Making Sense of the Data; and Conclusion.

Reading summary

The study records mostly stable aggregate electoral outcomes and argues that substantial successful entrenchment should ordinarily appear there. It favors a changing-judgment account for part of the discrepancy, without directly estimating that contribution. Coverage, performance, timing and the country-score restriction matter.

[Complete publisher text](#)

Research explanation

In everyday terms: Little and Meng ask whether recent worldwide declines in democracy also appear in relatively easy-to-check election outcomes. Entrenchment means leaders making it harder to remove them from office. Keep their worldwide question separate from judging one country.

Democratic backsliding means weakening institutions or protections associated with democracy. It need not mean an immediate switch to dictatorship. Little and Meng ask

whether the world as a whole experienced the large recent decline often described in public discussion. Their question concerns a cross-country trend, rather than whether a particular country has a serious problem.

Measuring a broad idea requires indicators: recorded signs of the thing we want to understand. An indicator such as whether the governing party lost an election is relatively easy to verify. Judging whether an entire election was free and fair requires weighing campaign access, intimidation, counting, and other conditions. The authors call the first kind more objective and the second more subjective; they argue that expert judgment offers wider coverage but can introduce bias.

Their main reasoning starts with electoral competition. Leaders weakening democracy often want to stay in office. If such efforts have become substantially more successful worldwide, we should expect governments to lose less often or winners to dominate more strongly. The authors therefore adopt a relatively narrow starting point centered on meaningful elections and acceptance of defeat. They also recognize that democracy requires rights, media freedom, and checks on executive power.

Method

The study examines records from several sources: election outcomes and competition, constitutional constraints on leaders, attempts to evade term limits, and journalists jailed or killed. The unit is often a country-year: one country observed in one year. To avoid comparing only whichever countries held an election that year, some graphs carry forward a country's most recent election if it occurred within six years.

Most main comparisons give each country equal weight. This answers a question about the average country, not the conditions experienced by the average person. The authors also examine population weighting, separate regime groups, and changes over longer periods. These are descriptive comparisons; countries and leaders were not randomly assigned to different democratic conditions.

Findings

Rates of ruling-party and leader turnover are broadly stable after the late 1990s. Winners' vote or seat shares do not generally rise, and basic multiparty competition does not show the expected global decline. Constitutional constraints are largely stable or increasing. The press evidence is mixed: journalists jailed increase while murders decrease.

The authors regard time-varying expert judgment as a likely explanation for at least part of the discrepancy with more pessimistic indices. They model changing information or standards and discuss a shift toward subtle attacks. The descriptive comparisons do not directly estimate how much of the discrepancy is bias. Their official reply restates that restriction while insisting that the stable electoral patterns still require explanation.

Limits

Stable electoral indicators support a claim with clear limits about those indicators. They do not establish that every democratic protection stayed intact or that each country is safe. The authors explicitly accept backsliding in some countries. Election performance also reflects popularity and governing performance, so turnover is not a direct measure of every democratic condition.

Their simple combined index summarizes aggregate trends; they explicitly warn against using it as a country-year democracy score. Do not convert its numeric level into a valid diagnosis of one country. A local audit can expose an indicator's limit without establishing the direction or size of a global trend.

Technical details

Sources include National Elections Across Democracy and Autocracy (NELDA), the Database of Political Institutions, expanded constitutional-constraint data, Versteeg and colleagues' term-limit data, and the Committee to Protect Journalists. Coverage varies by measure; do not claim a single balanced sample size for every graph.

The 12-component summary index normalizes available indicators to 0–1 and averages them. Journalist killings and imprisonments are excluded from that index because coverage is shorter and most country-years have zero counts. Its numeric level has no substantive country-democracy interpretation.

The term-limit source identifies 234 end-of-term instances in 2000–2018; 60 serious evasion attempts and 34 successes. These are eligible-leader opportunities, not 234 countries or elections. The study finds no clear trend in evasion rates.

Population-weighted trends are more pessimistic and sensitive to the largest countries. Formal models explore a common change in coder information or standards and a leader's choice between subtle and blatant violations. Models establish possible mechanisms, not the measured prevalence of those mechanisms.

Evidence receipts

What should appear if entrenchment grows?

The proposed reasoning predicts less frequent electoral defeat or stronger incumbent dominance. Stable turnover is relevant evidence against that particular global pattern, even when it cannot capture every democratic protection.

What was actually observed?

Most electoral indicators do not show a clear recent global decline. The press indicators point in different directions. This is a collection of findings, rather than evidence that all institutions everywhere remained unchanged.

Was coder bias proven?

No direct estimate proves how much of the discrepancy is caused by changing expert judgment. The paper models and discusses that possibility alongside subtler undemocratic strategies.

Can the summary index diagnose one country?

No. The authors say their simple combined index is for summarizing aggregate trends, not an appropriate substitute for country-year democracy indices.

Article B Conceptual and Measurement Issues in Assessing Democratic Backsliding

Carl Henrik Knutsen, Kyle L. Marquardt, Brigitte Seim, Michael Coppedge, Amanda B. Edgell, Juraj Medzihorsky, Daniel Pemstein, Jan Teorell, John Gerring, and Staffan I. Lindberg (2024). PS: Political Science & Politics. 57(2): 162–177 · DOI 10.1017/S104909652300077X

Dates

Published online 11 January 2024; issue April 2024. Includes comparisons over 1980–2020 and analysis of 2012–2022 democratic change using V-Dem version 13.

Access

Open-access complete publisher text · CC BY.

Lens

A reliable count can still miss what democracy requires. Define the democratic condition, choose signs that capture it, and investigate whether changed scores reflect changed institutions or changed coding.

Read first

Start with the explanation on this screen. Then check what was compared, what was found, and what remains uncertain. The journal article link is optional.

Sections

Read Conceptualizing Democracy; Conceptualizing Backsliding; Biases in V-Dem Measures?; Objective versus Subjective Measures; and Conclusion.

Reading summary

The response argues that meaningful contestation requires supporting freedoms which low-threshold indicators can miss. It tests several implications of shared expert pessimism and finds little support for them. This is substantive evidence against specified bias accounts, without validating every rating or excluding every bias.

[Complete publisher text](#)

Research explanation

In everyday terms: the response asks whether election outcomes alone miss important losses of rights, media freedom and checks on leaders. The disagreement concerns both what democracy means and how to measure change.

Knutsen and colleagues respond directly to Little and Meng. Their starting question is: what do we mean by democracy before we measure its change? A definition centered on the existence of competing parties captures something important. But people also need room to organize, discuss politics, obtain independent information, and challenge abuses. Losing one of these conditions may count as democratic erosion while elections continue.

A measure can be easy to reproduce yet poor at representing its intended concept. Counting whether several parties exist sets a low threshold: parties may remain legal while their ability to campaign deteriorates. Agreement about the count does not fix this missing coverage. Conversely, expert assessment may capture harder-to-observe conditions but needs clear questions, evidence, and reported uncertainty.

The authors also ask how signs should be combined. If a necessary protection disappears, should strengths elsewhere cancel it out in an average? How long a period should count as change? Several small annual losses can become a substantial decline over ten years. Giving each country equal weight and giving more weight to populous countries also answer different questions; neither choice should be concealed inside a single headline.

Method

They compare Little and Meng's measures with V-Dem's Electoral Democracy Index, which combines elected officials, voting rights, freedom of association, freedom of expression, and clean elections. V-Dem experts answer defined questions about specific conditions; they do not simply assign a general democracy mood score. A statistical measurement model combines their answers and represents uncertainty.

The authors test implications of the proposed changing-pessimism bias. They inspect whether declines are broadly synchronized, whether successive data releases revise earlier judgments downward, and whether expert disagreement follows systematic patterns. They also examine missing data, indicator thresholds, and how the combined index behaves in known historical cases.

Findings

Across 2012–2022, they identify 32 countries with clearly discernible negative changes and 11 with positive changes on the Electoral Democracy Index; more than 100 have no discernible change. “Discernible” means the estimated change's 95% uncertainty interval excludes zero, which would mean no change. The interval describes imprecision in the estimate. These are not counts of every possible small decline or improvement.

Their tests find little to no evidence for the proposed shared pessimism pattern. Declines differ across countries and components: expression and civil society often deteriorate while electoral measures are less affected. They argue that limited coverage, low thresholds, and systematic missingness weaken Little and Meng's measures as a benchmark for broader democracy.

Limits

The response supports the existence of recent backsliding on a broader, specified conception. It does not prove that every expert rating is accurate or rule out every conceivable changing bias. The authors acknowledge that their model cannot directly correct a synchronized shift in every expert's standards.

An uncertainty interval concerns an estimate, not a guarantee that a government is safe when zero is included. Definitions, periods, missingness and weighting can produce different trends. A fictional local audit is not a new V-Dem observation or a test of the worldwide conclusion.

Technical details

V-Dem version 13 underlies the response. The 2012–2022 comparison uses 95% credible intervals for estimated changes. A credible interval represents uncertainty within the statistical model; it is not a forecast of a country's political future.

The Electoral Democracy Index averages an additive and a multiplicative combination of five components. Suffrage and elected officials are fact-based components; expression, association, and clean elections aggregate several expert-coded indicators. The model adjusts for reliability and differences in how experts use the rating scales.

The response reports 43% of country-years missing a quarter or more of Little and Meng's 12 index indicators and 11% missing half or more. Their regressions show that missingness is associated with lower EDI and declining EDI, raising a selection concern rather than proving every missing value is bad.

The authors compare coder-level scores across V-Dem versions 9 (2019) and 13 (2023), plus other checks. These indirect tests address predicted patterns of bias and cannot rule out all possible common shifts. Several authors are involved in V-Dem; assess their evidence and tests rather than infer validity or invalidity from that involvement.

Evidence receipts

What can a stable party count miss?

Several parties can remain legal while access to independent information or organizing weakens. A low-threshold indicator may stay at its top value while a broader democratic condition deteriorates.

What did the ten-year comparison show?

On V-Dem's Electoral Democracy Index for 2012–2022, 32 countries show discernible negative change and 11 positive change. More than 100 show no discernible change. The finding depends on the specified measure, period, and uncertainty rule.

How was changing judgment investigated?

The researchers examined previous-score revisions, the similarity of changes across countries, and patterns of expert disagreement. The tested patterns did not support a broad shared-pessimism explanation.

Why inspect missing records?

If information is disproportionately missing from declining countries, an average calculated only from the available cases can look healthier than the full population. Missing is not the same as unchanged.

Further reading

What We Do and Do Not Know about Democratic Backsliding

Andrew T. Little and Anne Meng (2024). PS: Political Science & Politics. 57(2): 224–229 · DOI 10.1017/S1049096523001038

Open-access complete publisher text · CC BY.

The authors' official reply separates their observed electoral patterns from disputed explanations. It restates the country-score warning, acknowledges no direct test of coder bias in their original article, and asks for more evidence on rights, institutions, and resilience. It is part of the same controversy, not a correction notice.

Which disagreement concerns the recorded pattern, and which concerns whether that pattern adequately represents democratic change?

[Complete publisher text](#)

An Events-Based Approach to Understanding Democratic Erosion

Hannah Baron, Robert A. Blair, Jessica Gottlieb, and Laura Paler (2024). PS: Political Science & Politics. 57(2): 208–215 · DOI 10.1017/S1049096523001026

Open-access complete publisher text · CC BY.

The Democratic Erosion Event Dataset records political events and resistance alongside country scores. Turkey and Brazil illustrate how events can reveal erosion missed by narrow measures, or resistance obscured by an alarming summary. Event selection and classification still require judgment. Two case studies illuminate a process without settling the worldwide rate.

How would you distinguish an attack, its implementation, and the resistance that prevents it from producing a lasting loss?

[Complete publisher text](#)

Measuring Backsliding with Observables: Observable-to-Subjective Score Mapping

Daniel Weitzel, John Gerring, Daniel Pemstein, and Svend-Erik Skaaning (2024). PS: Political Science & Politics. 57(2): 216–223 · DOI 10.1017/S1049096523001075

Open-access complete publisher text · CC BY.

A model trained on earlier observable institutions predicts established democracy indices in later years. The analysis suggests some decline, with timing and size varying across indices. The authors stress that predictive consistency does not establish whether the measure captures the intended idea: a model can reproduce earlier judgments without knowing whether those judgments were right.

If a model accurately reproduces past ratings, what must still be checked before calling it a valid democracy measure?

[Complete publisher text](#)