

Reassessing the Trevor Project's Suicide-Attempt Findings

New analysis raises serious questions about claims that restrictive laws increased youth suicide attempts.

June 16, 2026

Caution: discussion of suicide

This spotlight discusses content related to suicide, which may be distressing for some readers. Please exercise caution and [seek support](#) if needed. If you or someone you know is struggling, please reach out to a mental health professional or a trusted individual.

In September 2024, researchers at The Trevor Project, a U.S.-based LGBTQ+ youth suicide-prevention charity, published [a paper in *Nature Human Behaviour*](#) based on survey responses from more than 61,000 transgender and nonbinary youth in the United States. Using an online convenience sample, the study examined self-reports of past-year suicide attempts before and after the enactment of “anti-transgender” laws (defined as any restrictive laws regulating medical transition, sports, birth certificates, etc.), and concluded that such laws caused a 7–72% increase in past-year suicide attempts among youth ages 13–17.

This finding was amplified by the Trevor Project's [press release](#), titled “Anti-Transgender Laws Cause up to 72% Increase in Suicide Attempts Among Transgender and Nonbinary Youth, Study Shows.”

Media coverage [quickly embraced](#) the press release's upper-end estimate of a [72%](#) increase in suicidality, along with its [causal language](#). The paper's findings made their way into [policy briefs](#), [government guidance](#), [scholarly writing](#), and [continuing medical education \(CME\)](#), including a recent event provided by the American Psychiatric Association in which the paper was cited three times as evidence that “laws increased past-year suicide attempts” by up to 72%. The paper has been uncritically cited [over 130 times](#) in the scholarly literature. In some reports, [the qualifier has been dropped entirely](#), with the finding [reported as a flat “72% increase,”](#) including in two [academic papers](#).

However, a [peer-reviewed critique](#) by Cohn et al., published in *Nature Human Behaviour* in late May 2026, identified serious flaws in the Trevor Project's analysis, casting doubt on the study's validity. Cohn et al. reveal that the headline finding of a 72% increase in suicide attempts **came from a single state, Idaho**—which contributed only a small sample of respondents (approximately 60 per year) and **where, critically, no relevant law was in force during the relevant time period**. The critique also identified several other study limitations that undermine the credibility of its conclusions.

The Trevor Project authors responded to Cohn et al.'s critique. They acknowledged Idaho's “prominent role” in the study's analysis and conclusions. Remarkably, however, the authors did not regard it as problematic that the Idaho laws in question had been blocked or enjoined more than a year before the estimated increase in suicide attempts appeared, asserting, “the concerns raised by Cohn et al. do not alter the interpretation of our findings.”

Because youth suicide is central to the national and international debates over how best to care for gender-dysphoric youths, and because the original analysis was technical—relying on a “difference-in-differences” technique to infer a causal relationship between legislative restrictions and suicide attempts—this Spotlight clarifies the issues for readers without statistical expertise. We have provided a [Technical Appendix](#) for those wishing to examine the issues in greater detail, and those with more in-depth interest should read the Trevor Project's original paper, the critique, and the Trevor Project authors' response.

The Trevor Project Study

The Trevor Project authors hypothesized that “anti-transgender” laws cause suicide attempts among “TGNB” (transgender and nonbinary) youth. The authors provided a broad definition of “anti-transgender” laws, describing them as laws that encompass “a range of issues, from limiting access to gender-affirming healthcare (for example, puberty blockers, hormone

therapy and gender-affirming surgeries) or bathrooms to prohibiting TGNB young people from participating in sports or school activities that align with their gender identity.”

The proposed mechanism of suicide-attempt causation was twofold: the laws exacerbate the baseline mental-health vulnerabilities of affected youth and also introduce new stressors by “signal[ing] a broader societal rejection of their identities, communicating that their identities and bodies are neither valid nor worthy of protection.”

The data for the analysis came from the Trevor Project’s annual anonymous online survey, administered each year between 2018 and 2022 to young people (ages 13–24) across 48 US states and territories, which included questions about past-year suicide ideation and past-year suicide attempts. Acknowledging that, despite previously reported associations, “no research has specifically identified a causal link between anti-transgender laws and increased suicide risk among TGNB young people,” the authors set out to conduct a pioneering study that would, for the first time, demonstrate causation. To do that, they relied on difference-in-differences methodology, which is described in more detail below.

Following this methodology, the authors reported a striking finding for the subset of 13–17-year-olds. According to their analysis, after a small (7%) increase in self-reported suicide attempts in the first year after enactment of the “anti-transgender” legislation, self-reported attempts increased by 72% in the second year, and by a smaller but still highly statistically significant increase (52%) in the third year. The lowest and highest of these estimates gave the paper its headline finding that the laws caused a 7–72% increase in past-year suicide attempts among this group.

However, as a critique published in *Nature Human Behavior*—the same journal that published the original Trevor Project study—demonstrates, the study suffers from profound limitations that render its conclusions untrustworthy.

Critique of the Trevor Project Study

A critique by Cohn et al. (2026) highlighted the fact that the headline finding of a large increase in suicide attempts, reported by the Trevor Project study, rested entirely on data from the state of Idaho, where, remarkably, no law that could be described as “anti-transgender” was in force at the time. This discovery gravely undermines the quoted “72% increase in suicide attempts caused by state legislative restrictions.” Even the lower end of the estimate—7%—is in question due to numerous methodological limitations.

Below we highlight some of the study’s key limitations, drawing largely from Cohn et al. Where the Trevor Project authors responded to these critiques, we also describe their response, and provide our own commentary—so that readers can judge whether Cohn et al.’s critique survives the authors’ explanation.

- **The headline 72% increase in youth suicide attempts came from a single state—Idaho—where no relevant laws were in effect.** As Cohn et al. observe, Idaho was the only state contributing data to the year-2 and year-3 estimates (see Figure 1), which supplied the evidence for the claimed causal effect: an “up to 72%” increase in suicide attempts among youth ages 13–17 (see Table 1). Not only did Idaho supply a small sample (around 100 respondents per time period, with approximately 60 adolescents ages 13–17)—but, critically, no laws described as “anti-transgender” were in force in Idaho for at least a year before the Trevor Project authors claim to detect effects of the laws on suicidality. Idaho’s only two relevant laws—HB 500 (school sports) and HB 509 (birth certificates), both enacted in 2020—had been blocked or enjoined by fall 2020, a year before the 72% increase appeared in the data. Notably, neither law concerned restrictions on pediatric gender transition, the issue most associated with this debate.
 - ***The Trevor Project study authors’ response:*** The authors conceded “Idaho’s prominent role in the post-law estimates.” However, the authors defended Idaho as the sole basis for the 72% estimate, even though no law was in force during the analysis period. They offered two justifications: that some Idaho school districts and sports associations began implementing HB 500 despite the judicial orders, and that public discourse around the laws, which they say should be considered part of the suicide-causal pathway, continued after the laws were blocked or enjoined.
 - ***SEGM’s take:*** The first explanation would imply that the entire 72% increase in suicide attempts rests on a few instances of inconsistent, short-lived enforcement of the sports law—which strains credulity. The second—that

public discussion after the laws' passage drives suicidality much as the laws themselves do—sits uneasily with the Trevor Project's own finding that discussion of the laws before their passage had no effect. It is our assessment that the upper end of the study's estimate—the 72% increase in suicide attempts among adolescents—cannot reasonably be attributed to “anti-transgender legislation” because no such legislation was in force when the increase appeared. However, to whatever extent debates about the laws may increase suicide risk, the lesson is not that debate should cease, but that it should be conducted carefully, without the alarmist framing that has become common in this area.

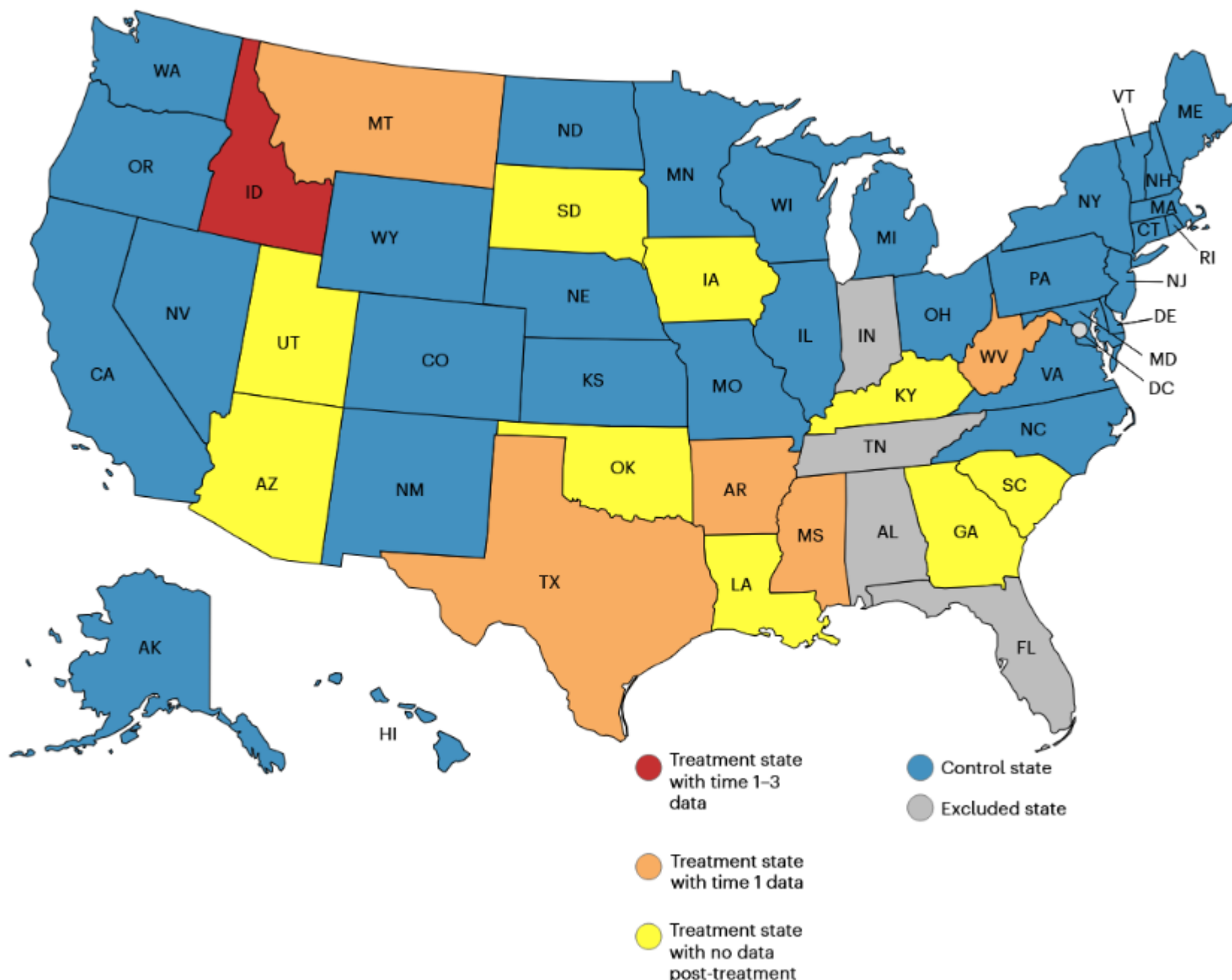


Figure 1: Mapping of treatment and non-treatment ("control") states. Only Idaho supplied data for Times 2 and 3.

- **The lower end of the estimate—a 7% increase in suicide attempts—is also highly questionable.** Even the low end of the Trevor Project estimate of the increase in suicide attempts—7%—appears questionable. Although this estimate came from five other states plus Idaho (AR, MS, MT, TX, WV), the finding barely reached statistical significance ($P = 0.049$), with the lower bound of the 95% confidence interval hovering just slightly above zero (0.001, 0.079). Critically, as Cohn et al. note, no correction for multiple testing was described by the study authors, although they conducted at least 30 comparisons looking for effects of the laws, including comparing two different age groups (13–17 and 13–24) on three past-year suicidality measures (seriously considering suicide, at least one attempt, and total number of attempts). Applying a Bonferroni adjustment for these multiple tests—which would require a p-value below roughly 0.002 ($0.05/30$) for significance—would render the already barely significant result not statistically significant.
 - **The Trevor Project study authors' response:** The authors did not concede this point to Cohn et al., insisting that a p-value below 0.05 settled the matter:

“With regard to the significant time 1 estimated effect being marginally significant, we adhered strictly to established scientific conventions and Nature’s editorial guidelines for statistical significance (with a P-value threshold of 0.05).”
 - **SEGM's take:** The Trevor Project's response reflects a widely criticized overreliance on the p-value threshold of 0.05—a “bright line” that the American Statistical Association has [formally cautioned against](#), warning that a result does not become true on one side of the cutoff and false on the other. The appeal to the journal's own conventions is especially strained given that *Nature Human Behaviour* has itself [published calls to move beyond the 0.05 standard](#). More importantly, the response sidesteps Cohn et al.'s actual point: that the authors ran these multiple comparisons

without any correction for the resulting inflation of false-positive risk. When many tests are run, some may cross the 0.05 line by chance alone; a marginal, uncorrected result of $P = 0.049$ cannot bear the weight the headline places on it.

- **Placebo tests undermine conclusions that laws affected suicidality.** As Cohn et al. point out, the authors' own internal test flagged a problem with their conclusions that the laws affected suicidality. The authors had selected two measures the laws were unlikely to affect—homelessness and full-time employment—and checked whether their model detected any spurious effect on them. In Idaho, it did: both came back positive in the same Time 2 period that produced the 72% figure—which should not have happened according to the authors' own assumptions. That is a critical warning sign that any change in that period—including the purported 72% increase in suicidality—may be driven by outside factors rather than by the laws in question.

Any claim that the laws increased suicidality must rule out the possibility that some other factor, other than the laws, accounts for the observed trends. Cohn et al. note that there were well-known outside factors that could plausibly have influenced suicidality in Idaho in 2021, the Time 2 period that produced the “72%” suicide-attempt figure. At that time, Idaho became one of only two states in the U.S. to begin rationing health care under COVID-19 “crisis standards of care.” A disruption of that magnitude is a plausible alternative cause of increased distress, entirely separate from any law. This is just one example; other potential factors in Idaho would need to be examined before any claims of association or causation could be sustained.

- **The Trevor Project study authors' response:** Commenting on their internal checks, the Trevor Project study authors argue the significant homelessness and employment results appeared only at Time 2, not at Time 1 or Time 3, and that isolated significant results are expected by chance in large datasets. Regarding the COVID-19 argument, they note that their models included a measure of COVID-19—state-level pandemic death counts—which they say showed little evidence that the pandemic explained the trend.
 - **SEGM's take:** Neither response is reassuring. A single state-level COVID-19 death count is a crude proxy for the effects of the COVID-19 pandemic, and it cannot capture the effects of something as specific as Idaho's rationing of health care in the precise period that generated the 72% figure. As for the claim that the unexpected signal of confounding from their internal checks “only” occurred at Time 2, this explanation is not satisfactory since Time 2 in Idaho is precisely the period that supplied the 72% figure.
- **The headline finding is based on a small subset of the data.** Part of what gave the paper its initial credibility was the study's impressive 61,240-person sample. But the operative numbers are far smaller. The lower end of the estimate—a 7% increase in suicide attempts—rested on a single survey wave of 13–17-year-olds from just six states, a small fraction of the 24,361 13–17 year-olds who they analyzed. As already discussed, the upper end—the 72% increase—came from an estimated 60 adolescents per period, all in Idaho: roughly 0.25% of the adolescents analyzed, and about 0.1% of the total sample size of 61,240.
 - **The central claim of causation rests on an assumption of “parallel trends” that cannot be demonstrated.** The Trevor Project authors present the value of their study as showing, for the first time, that state restrictions are not merely correlated with increased suicidality but directly cause it. That causal claim depends entirely on their use of a difference-in-differences analytic approach. The method is described in more detail in the [Technical Appendix](#), but a working understanding of what it estimates is needed to see why the causal claim cannot be sustained.
- In essence, difference-in-differences compares how an outcome of suicidality changes over time in states that enacted the restrictions (the “treatment” group) with how it changes in states that did not (the “control” group), with the restrictive laws serving as the “treatment.” Under the right assumptions, this comparison allows the authors to argue that the estimated effect reflects causation, rather than mere association.
- Critically, however, the difference-in-differences approach supports a causal interpretation only if the so-called “parallel-trends assumption” holds—that is, if the suicidality outcomes in the treatment and control groups would have continued to move in parallel had the laws not been enacted.

As Figure 2a below illustrates, it does not matter whether the two groups start at different baseline levels: even if the treatment group begins with a higher baseline rate of suicidality, an effect can still be identified and attributed as the effect of the laws, provided the two groups moved in parallel before the laws and diverged afterward.

The plausibility of parallel trends can be supported in two ways: by choosing control states that resemble the treatment states culturally, politically, and demographically—and by confirming that the two groups did in fact move in parallel before the laws were enacted. Cohn et al. identify weaknesses in the Trevor Project study on both counts.

On the first point, the control states (those without qualifying laws) included Northeast and West Coast states that differ sharply from the treatment states—and from Idaho in particular—in geography, culture, and politics. On the second point, while the Trevor Project authors did run statistical tests to reassure themselves that there was no evidence of non-parallel trends, Idaho contributed less than 4% of the data for each of those tests. The implication is that the very strong assumption that trends in suicidality among Idaho youth would move in parallel with those in states that were geographically, culturally, and politically distant, must be taken as a leap of faith.

- ***The Trevor Project study authors' response:*** The authors repeated their assertion that their pre-law-enactment estimates showed no evidence of non-parallel trends, and noted that excluding the largest control states left the results unchanged.
- ***SEGM's take:*** This explanation is not reassuring. The pre-enactment checks the authors cite come almost entirely from states other than Idaho and say little about Idaho specifically. Moreover, rather than presenting empirical data on actual suicide attempts during the pre-enactment periods in Idaho and the comparison states, the authors took a purely statistical route: they assumed the trends were parallel and relied on a statistical test that failed to disprove that assumption. But failing to disprove parallelism is not the same as establishing it. Given how much rests on this assumption—it is central to the difference-in-differences model and to the entire claim of causation—the burden is on the authors to demonstrate its plausibility, not to assume it and rest on a test's failure to reject it. The authors present no raw data or other empirical evidence that Idaho moved in parallel with the control states, and the fact that Idaho supplied less than 4% of the data to those tests makes the explanation harder still to accept.

Figure 2a: The concept of a difference-in-differences (DD) analysis. Time runs horizontally, and the vertical black line marks the point at which “treatment”—for example, enactment of a new law—is introduced. The solid teal line represents the observed data in the treatment group before and after treatment. The solid orange line represents the observed data in the control group, which does not receive the

treatment. The dotted teal line represents the DD model's projection of what the treatment group's data would have been if treatment had not been introduced.

- **The overall findings were mixed, and the public framing emphasized the most dramatic part of the analysis.** The dramatic 7–72% range comes from one specific model, for one specific outcome (self-reported number of suicide attempts), in one specific age group (13–17-year-olds). However, the Trevor Project study authors performed many analyses, evaluating three different suicide-related outcomes using two different difference-in-differences models, and two age groups (13–24 and 13–17). Altogether, there were 30 opportunities to find a statistically significant law effect. The two-way fixed-effects analyses were null, and “seriously considering suicide” moved in the opposite direction: by the paper’s own modeling logic, the Time 3 result would suggest that the laws reduced serious suicidal thoughts.
- **The way the data were presented further inflated the perceived magnitude of the effects, and the calculation used was inappropriate.** The study presented its finding of a 0.39 increase in suicide attempts per person over the sample mean of 0.54 as a “72% increase” ($0.39 \div 0.54 \approx 72\%$). Presenting risk data in relative terms is known to [inflate perceived effect sizes](#) and is not recommended.

In addition, there are serious concerns about how the authors calculated the increase attributable to the laws—specifically, that they combined “apples and oranges” in the calculation—the apples being the projected suicide increase of .39 attempts per person, and the oranges being the overall sample mean of .54 attempts per person. Instead, both sets of numbers should have been derived from the same source—namely, the model itself—and, critically, the denominator should have used the model’s projection of the expected rate of suicide attempts without the laws, rather than the overall sample mean derived from all states and all time periods. This issue is explained in more detail in the [Technical Appendix](#).

- **The study treated very different kinds of laws as if they were a single exposure.** Lee et al. combined 48 statutes under one “anti-transgender” category, and by their own account “did not differentiate the laws based on type or scope.” Yet the 48 laws addressed quite different subjects: 30 concerned eligibility rules in competitive sports, 7 concerned medical gender transition—including measures that restricted care and one resolution [requesting a study](#)—4 governed updating legal identification documents, 3 addressed participation in school activities, 3 concerned bathroom access, and 1 dealt with religion-based exemptions from anti-discrimination protections. The single “anti-transgender” label thus allows a finding generated by sports and ID laws to be presented as evidence about the category as a whole, including the medical-transition restrictions most prominent in public debate.
- **The authors’ conflicts of interest were unmanaged.** Alongside the critique by Cohn et al., we note that the original paper includes a conflict of interest (COI) statement disclosing that all authors were current or former employees of The Trevor Project, which has consistently taken an active role opposing state restrictions through public advocacy and court submissions. It has appeared as an amicus in litigation regarding [bathroom access](#), [medical gender transitions for minors](#), and [laws regulating therapy with sexual and gender minorities](#)—often citing its own research papers and survey data in support. It has [already invoked the “up to 72%” statistic](#) from this very study in *West Virginia v. B.P.J.* and *Little v. Hecox*, two Supreme Court cases concerning state limits on transgender participation in school sports.

Managing a conflict of this magnitude would call for safeguards such as pre-registration of study protocol in advance, and providing access to data to allow others to reproduce outcomes and conduct sensitivity analyses—none of which were done.

SEGM Take-away

When the Trevor Project press release promoted the study, it quoted the study’s senior author, who was the Trevor Project’s Vice President of Research, stating that the study “critically confirms—for the first time—a causal relationship between anti-transgender laws and heightened suicide risk among transgender and nonbinary young people.” However, as the discussion above demonstrates, neither the press release’s headline of an “up to 72% increase in suicide attempts” nor any conclusions about any increases in suicidality following restrictive laws are credibly supported by the data.

Media coverage [quickly embraced](#) the press release's upper-end estimate of a [72%](#) increase in suicidality, along with its [causal language](#). The paper's findings made their way into [policy briefs](#), [government guidance](#), [scholarly writing](#), and [continuing medical education \(CME\)](#), including a recent event provided by the American Psychiatric Association in which the paper was cited three times as evidence that "laws increased past-year suicide attempts" by up to 72%. The paper has been uncritically cited [over 130 times](#) in the scholarly literature. In some reports, [the qualifier has been dropped entirely](#), with the finding [reported as a flat "72% increase,"](#) including in two [academic papers](#).

Such widespread use of claims about suicidality carries special ethical and public-health risks, particularly when they concern vulnerable adolescents. As Dr. Alison Clayton has noted, "an excessive focus on an exaggerated suicide risk narrative ... may create a damaging nocebo effect ... whereby suicidality in these vulnerable youths may be further exacerbated." The Trevor Project's press-release framing also runs counter to the [Trevor Project's own suicide-reporting guidance](#), which warns that suicide is complex, rarely has a single cause, and should not be reduced to one precipitating event.

Image from the Tennessee Equality Project's 2025 policy brief on House Bill 64/Senate Bill 472, showing the suicide-attempt statistic presented as an unqualified "72%" increase.

The coverage of the Trevor Project paper follows a troubling trend of invoking suicide risk whenever gender transition for minors is questioned—from Johanna Olson-Kennedy's ["dead son or live daughter"](#) framing, to claims from activist organizations in the UK that the country's puberty-blocker ban caused a surge in youth suicides—a claim found to be [unsubstantiated](#) by the [independent Appleby report](#).

It is important to underscore that to date, no credible evidence has emerged that delaying pediatric transition increases the risk of suicide—or that providing pediatric transitions reduces it. A 2021 [systematic review](#) found that the evidence for whether medical intervention affects suicide risk is "insufficient." A comprehensive [Finnish register study found](#) that suicide deaths in this population are rare and could not establish that gender reassignment reduced suicide risk.

Our point is not that suicide risk should be minimized or ignored. It is that suicide claims demand exceptional evidentiary care, especially when they concern vulnerable adolescents and are used to influence parents, clinicians, courts, or policymakers. The Trevor Project paper by Lee et al. and the manner in which it has been presented by the Trevor Project struggle to meet that standard. We therefore welcome the new analysis, which brings to light important limitations that were not made clear in the original paper and demonstrates that its claims are much weaker than originally stated.

We commend *Nature Human Behaviour* for hosting the exchange between the original authors and Cohn et al. and we credit the Trevor Project authors for engaging with the critique. In our assessment, however, their response did not address the substance of Cohn et al.'s concerns.

Open scientific debate is valuable, but it is not sufficient here. The headline finding—that state restrictions caused a 7–72% increase in past-year suicide attempts—simply cannot be sustained by the data presented. When a widely repeated claim with policy and legal advocacy significance rests on such a weak foundation, debate in the correspondence pages is not an adequate remedy. A formal correction to the original paper is therefore warranted. That correction should make clear that the study's limitations materially undermine its causal interpretation, and that neither the analyses nor the underlying data credibly support the "up to 72% increase" claim about youth suicide attempts.

J. Cohn, one of the authors of the Trevor Project study's critique, is affiliated with SEGM.

Readers interested in the wider evidence on suicidality in youth gender medicine may also wish to view SEGM's 2023 conference presentations by [Dr. Alison Clayton](#), which examine suicide mortality, psychiatric comorbidity, minority-stress claims, and the evidence for whether gender-affirming interventions reduce suicide risk.

Technical Appendix

This appendix provides additional detail behind the concerns summarized in the Spotlight.

1. Study data and research methods

Lee et al. used online survey data collected in five waves between March 2018 and December 2022. Respondents were recruited through advertisements on Facebook, Instagram, and Snapchat targeted at LGBTQ+ youth between the ages of 13 and 24. Across the five waves, there were 162,903 respondents, of whom 61,240 self-identified as transgender or non-binary.

The authors did not analyze the full respondent pool. Their main analyses used two smaller analytic subsets. Analytic subset 1 included 43,228 individuals aged 13–24, after excluding four states and removing respondents who did not answer key survey questions or did not meet other validity checks. Analytic subset 2 was further restricted to respondents ages 13–17 and contained 24,361 individuals.

The primary analyses relied on two of approximately 145 survey questions: one asking how many times respondents had attempted suicide in the past year, and another asking whether they had seriously considered suicide in the past year. Additional data from survey questions on homelessness and full-time employment were used for some secondary analyses.

2. Difference-in-differences analysis

Lee et al. used difference-in-differences (DD) analysis, a statistical technique for analyzing observational data to examine the effects of introducing a new policy or law.

As illustrated in Figure 2b below, in a simplified DD design the policy or law is not in effect for anyone at the start of the study. In the terminology of DD analysis, individuals in all groups are initially “untreated.” During the observation period, a policy or law is introduced that affects some individuals, who form the treatment group, but not others, who form the control group. If the treatment and control groups follow the same trend before the treatment is introduced, but the trends diverge afterward, that divergence is interpreted as evidence that the law or policy introduced in the treatment group caused the observed difference.

Figure 2b: The concept of a difference-in-differences (DD) analysis. Time runs horizontally, and the vertical black line marks the point at which “treatment”—for example, enactment of a new law—is introduced. The solid teal line represents the observed data in the treatment group before and after treatment. The solid orange line represents the observed data in the control group, which does not receive the treatment. The dotted teal line represents the DD model’s projection of what the treatment group’s data would have been if treatment had not been introduced. This projected line is hypothetical and depends on the parallel trends assumption, described below. The treatment effect is estimated as the difference between the observed post-treatment data in the treatment group (point **a** in the figure) and its projected value (point **b**) under the hypothetical no-treatment scenario.

In Figure 2b above, the difference in the baseline levels (the teal line is higher than the orange line) is incidental; what matters is that the two move in parallel before treatment. The critical assumption of a DD analysis is therefore the **parallel trends assumption**: but for the treatment, the data in all groups would have followed identical trends over time. This assumption allows the model to project what the treatment group’s data would have looked like under the hypothetical scenario of no treatment. The DD estimate is the difference between the observed post-treatment data in the treatment group and the projected treatment-group outcomes under that hypothetical no-treatment scenario.

Conclusions from a DD analysis depend on the parallel trends assumption being true. This assumption can never be fully verified, because it is impossible to know what the treatment-group data would have been if treatment had never occurred. A DD analysis also assumes that nothing else influencing the outcome changed differently between the treatment and control groups over time, apart from the introduction of the treatment.

3. How Lee et al. structured “waves,” “Times,” and treatment

In Lee et al.’s article, “treatment” refers to the enactment of a law that the authors treated as anti-transgender for purposes of the analysis. Their data were more complex than the simple DD design shown in Figure 2b because the “treated” states introduced laws at different times. This required an analysis involving two time variables.

Using Lee et al.’s terminology, the study “waves” refer to the calendar timing of data collection. The “Time” variables in the analytic models refer to time relative to when a law, or “treatment,” was introduced.

Lee et al. structured their analysis so that **Time 0** is the first wave in which treatment occurred. This could be the wave during which the law was enacted or, if the law was enacted between waves, the first wave after enactment. **Time 1**, **Time 2**, and **Time 3** are the first, second, and third periods after Time 0. Negative values, from Time –4 to Time –1, mark pre-treatment periods.

The key quantities of interest in the analysis were the estimated effects for Time 1, Time 2, and Time 3. Under the DD assumptions, these estimates were interpreted as capturing the impact of enacting the laws.

4. Outcomes, models, and reported results

The investigators analyzed self-reported past-year suicide attempts in two ways. First, they treated the outcome as a quantitative variable: “the number of past-year suicide attempts.” Second, they dichotomized the outcome as one or more attempts versus no attempts: “at least one past-year suicide attempt.” A third analysis examined a binary variable indicating whether respondents reported that they had seriously considered suicide in the past year.

For each of these three outcomes, Lee et al. performed two types of DD analysis: an “event-study model” and a “two-way-fixed effects model,” with two variants of the latter. The event-study model estimated separate time effects for Time 1, Time 2, and Time 3 after enactment of a law. The two-way-fixed effects model estimated a single effect across all post-treatment times.

The authors also performed each analysis twice: once for respondents aged 13–24 and once restricted to youth aged 13–17. Taken together, these analyses yielded 30 opportunities to find a statistically significant result for an effect of laws.

Most of the analyses reported in the article did not support the claim that the laws increased suicidality. All of the two-way-fixed effects analyses produced null results. The analysis of whether respondents reported seriously considering suicide in the past year produced results in the opposite direction, suggesting that enacting laws led to less suicidality.

The article's main findings came from the event-study model, which estimated separate time effects for Time 1, Time 2, and Time 3. In this model, Lee et al. reported an increase of 0.04 in reported past-year suicide attempts among TGNB young people aged 13–17 at Time 1, the first period after legislation was enacted ($P = 0.049$, 95% CI 0.001 to 0.079). As Lee et al. noted, this estimated effect was very small. It also only narrowly met the conventional threshold for statistical significance, and no adjustment was reported for the multiplicity of tests.

The more striking results came from the DD estimates for Time 2 and Time 3. The largest result was an estimated Time 2 increase in past-year suicide attempts among TGNB young people aged 13–17 of 0.39 ($P < 0.001$, 95% CI 0.35 to 0.42).

5. A caution about interpreting the "72% increase"

As explained above, the target quantity in a DD analysis is the excess or deficit of an outcome that can, under certain critical assumptions, be attributed to the introduction of a new law or policy. The DD analysis produces estimated effects on the scale of the outcome variable being analyzed.

For example, in the analysis of "number of past-year suicide attempts" among participants aged 13–17, the Time 2 DD estimate was 0.39. This means that, under the model assumptions, the average number of reported past-year suicide attempts was estimated to be 0.39 higher at Time 2 than it would have been without the enactment of the laws.

The article's abstract, and some reports on the article, described this Time 2 estimate of 0.39 as a "72% increase." As Lee et al. noted within the article, this 72% figure was obtained by comparing 0.39 with the overall sample mean of 0.54. It is true that 0.39 is approximately 72% of 0.54, but this percentage cannot be interpreted as the percentage change in the outcome attributable to treatment.

To obtain a percentage change with that interpretation, the treatment effect—in this case, 0.39—should instead be compared with the projected value under the hypothetical no-treatment scenario. In the simple DD framework in Figure 2b, if the observed post-treatment outcome in the treatment group is **a**, and the projected no-treatment outcome is **b**, then the treatment effect is **a – b**. The relevant comparison for a treatment-effect percentage would be **a – b** relative to **b**, not **a – b** relative to the overall sample mean. The original paper does not report the projected no-treatment value needed to calculate that percentage.

Another possible misunderstanding is to interpret "72% increase" as summarizing the outcome after treatment versus before treatment. That is also incorrect. Although Lee et al. documented that the percentages were calculated relative to overall sample means, it is not clear from the article's abstract that these percentages do not retain the same interpretation as the DD model estimates. This ambiguity matters because the "up to 72%" figure became the public shorthand for the study.

6. The Matters Arising commentary and the role of Idaho

The Matters Arising commentary by Cohn et al. identifies important limitations in the research. Except for the one small and marginally significant Time 1 effect, the study's claims about the effects of state laws came from Time 2 and Time 3.

Because data collection ended in December 2022, a state could contribute data for estimating the Time 2 and Time 3 treatment effects only if it had enacted a relevant law in 2020 or earlier. The original article did not provide details about the 15 states that passed laws or how the timing of their laws corresponded to the model's time periods. Lee et al. later provided those details to the authors of the Matters Arising commentary.

The commentary shows (see Figure 3 below) that although 15 treatment states were included in the analysis, only one of them, Idaho, contributed data as a post-treatment state to Time 2 and Time 3. Nine of the 15 treatment states contributed no data beyond Time 0.

	Pre-treatment				Law enacted	Post-treatment		
Treatment state	Time -4	Time -3	Time -2	Time -1	Time 0	Time 1	Time 2	Time 3
AZ, GA, IA, KY, LA, OK, SC, SD, UT	wave 1 (2018)	wave 2 (2019- 2020)	wave 3 (2020)	wave 4 (2021)	wave 5 (2022)			
AR, MS, MT, TX, WV		wave 1 (2018)	wave 2 (2019- 2020)	wave 3 (2020)	wave 4 (2021)	wave 5 (2022)		
ID a N=412 b				wave 1 (2018) N=56	wave 2 (2019- 2020) N=67	wave 3 (2020) N=88	wave 4 (2021) N=109	wave 5 (2022) N=92

Figure 3: (Data from Table A in the Matters Arising commentary) This shows how waves of data collection corresponded to “Times” in the event-study model for the 15 treatment states included in the Lee et al. analysis. Nine of the 15 treatment states contributed no data beyond Time 0, and only one treatment state, Idaho, contributed data for Time 2 and Time 3. The commentary also estimated sample sizes for Idaho, which are shown in the table. Lee et al.’s main findings hinge on a few hundred participants in a single state.

^a A typographical error in the published article states that “Time 1” is 2021 for Idaho rather than 2020.
^b The authors of the Matters Arising paper estimated 412 Idaho respondents in Lee et al.’s Table 3 dataset of 43,228 individuals. The actual numbers may differ somewhat from the estimates, but the issues they raise do not depend on the precise numbers.

Across all five waves of data collection, Idaho provided 518 of the 61,240 total TGNB survey respondents. The Matters Arising paper estimated that Idaho had 109 respondents at Time 2 and 92 respondents at Time 3, or 201 responses in total. When the analysis was restricted to ages 13–17, the number was approximately 60 respondents in each period.

Thus, the Matters Arising commentary provides important context: Lee et al.’s headline results were based on data from a few hundred respondents in a single state, rather than on a broad national post-law pattern across many treatment states.

7. Is the parallel trends assumption plausible?

As discussed above, the validity of a DD analysis depends on the parallel trends assumption: that, in the absence of treatment, the outcome in the treatment states would have changed in parallel with the outcome in the control states.

Because the parallel trends assumption is unverifiable in DD analysis, control states should be chosen that are similar to the treatment states to help make this assumption as plausible as possible (see Figure 1, reproduced below). In Lee et al., however, all states without legislation were used as controls, without considering geographic, cultural, political, or other differences among states. For example, all Northeastern and West Coast states were included as control states, even though none of the treatment states came from those regions.

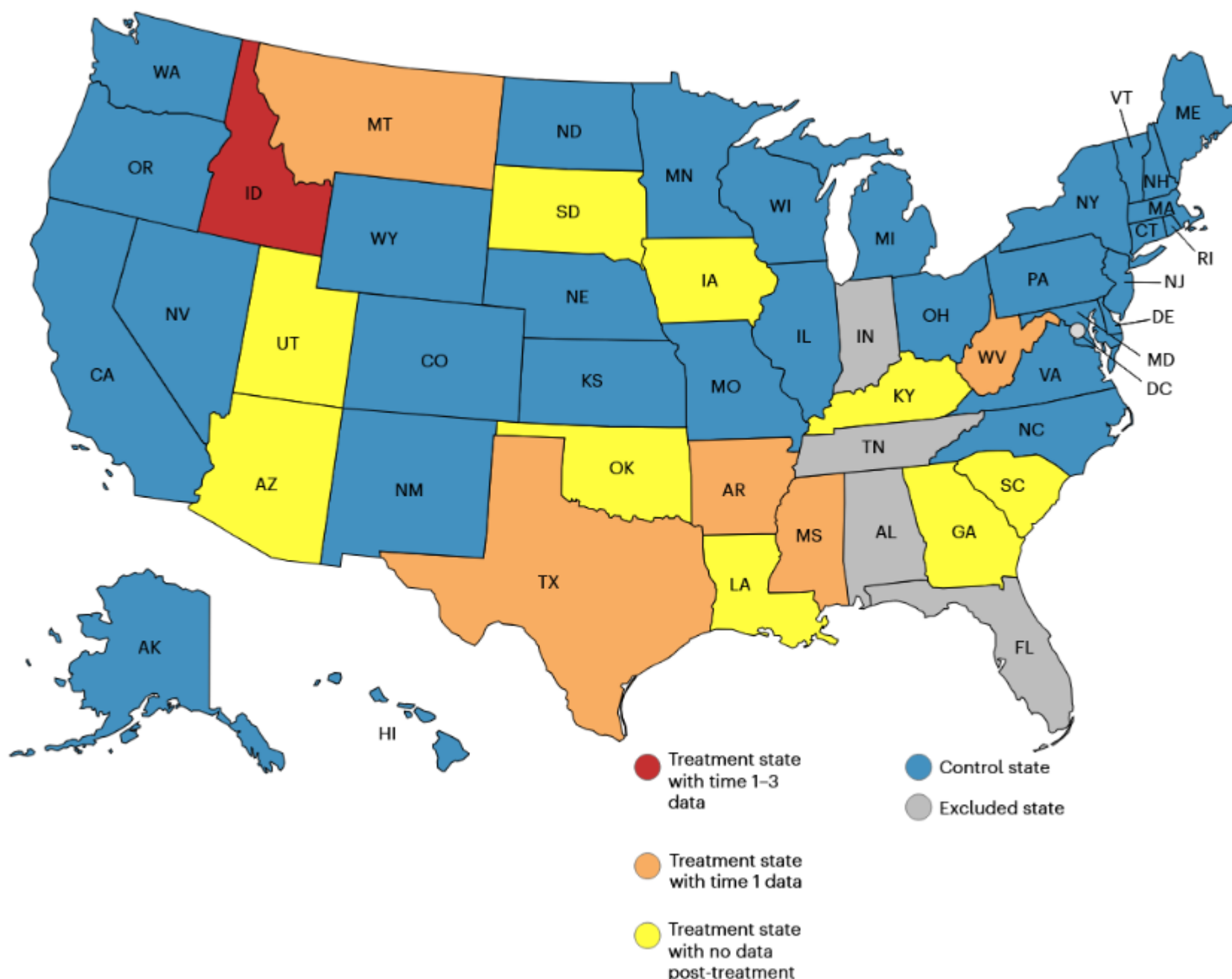


Figure 1: Mapping of treatment and non-treatment ("control") states.

The plausibility of the parallel trends assumption can sometimes be assessed empirically by showing that outcomes in the treatment and control states rose and fell in parallel during the pre-treatment period. The original publication, however, did not provide descriptive data showing such pre-treatment trends. In any case, Idaho had only one time period before Time 0, so any such exploration would necessarily rely on data from states other than Idaho.

A DD analysis also assumes that there were no other relevant changes during the period under consideration that affected the treatment and control states differently, apart from the treatment being studied. But data collection for Lee et al. overlapped with the COVID-19 pandemic, during which states had widely differing policy responses that could have affected mental health and suicidality.

For example, in fall 2021, Idaho was one of only two states that activated crisis standards of care and began rationing health care. This occurred during wave 4 of data collection, which provided the Idaho data used to estimate Time 2 effects in the event-study model.

Thus, there are reasons to doubt that changes in suicidality a year or more after the introduction of the laws in question—laws that were blocked soon after enactment—can be attributed to those laws rather than to other factors.