Still No Evidence: A Reply to Khan et al.

In the Spring 2026 issue of *Regulation*, Arif Khan and coauthors argued that the "boxed warning" on antidepressants discouraged providers from prescribing them to youth, thereby increasing suicides (Khan et al. 2026a). We disputed this in the Summer issue, arguing that use of the drugs themselves correlates with higher suicides, and concluding that the boxed warning is appropriate (Stone et al. 2026). In the same issue, Khan and coauthors responded to our arguments (Khan et al. 2026b). We now write in reply to those arguments.

Their response confirms rather than resolves the concerns we raised. It fails to address most of the empirical arguments we presented, relies heavily on a systematic review that is itself methodologically compromised, introduces new claims that are either self-refuting or unsupported by evidence and misrepresents the basis for our analysis.

Methodology is not a technicality / Khan et al. characterize our critique as focused on "methodological objections" while ignoring "real-world detrimental effects." This framing treats methodology as an obstacle to be set aside when the stakes are sufficiently high. It is not. Methodology is the mechanism by which we distinguish true findings from artifacts of analytical error (Hammad & Rochon 2026). A study that uses invalid methods does not demonstrate real-world effects; it demonstrates what invalid methods produce.

The "real-world detrimental effects" Khan et al. invoke have not been demonstrated independently of the causal assumption that antidepressant prescribing rates determine population-level suicide incidence. Remove that assumption—which the evidence requires—and there are no demonstrated real-world effects. What Khan et al. present as established fact is the unproven conclusion of a causal argument. They are using their conclusion to dismiss the critique that shows the conclusion is unproven. That is circular reasoning, not evidence.

Causal claim remains unsupported / Khan et al.'s principal response to our timing argument is that the warning's effect was delayed: that it took years to propagate through the healthcare system before becoming visible in population-level outcomes. This argument fails on multiple grounds.

For one, their causal claim contradicts the prescribing timeline. Antidepressant prescribing to young people declined by approximately 20 percent between 2005 and 2007 but, they claim, this had nothing to do with the boxed warning. The "suppression" did not take effect until 2008 when prescribing rates began a sustained increase. They provide no basis for their attribution of specific numbers of excess deaths—2,365 children and 3,593 young adults—to the warning. Suicide rates fell during the period of maximum prescribing decline and did not begin rising until 2007–2009 as prescribing was already recovering. The delayed effect argument requires us to believe that the decline in prescribing from 2004 through 2007, whatever the cause, resulted in falling suicide rates while prescribing recovery produced rising ones—the opposite of what their theory predicts. Between 2016 and 2022, antidepressant prescriptions for adolescents aged 12–17 increased by 43 percent and among those aged 18–25 by more than 47 percent (Chua et al. 2024). Current prescribing is at historic highs, yet suicide rates remain elevated. Khan et al. offer no account of this.

Another problem is that their causal claim contradicts the literature they cite. The majority of the studies Khan et al. invoke—including Gibbons et al. 2007, Kurian et al. 2007, Bridge et al. 2008, and Lu et al. 2014—describe adverse effects occurring immediately after the warning, in 2005–2007, precisely the period when youth suicide rates were falling. If the delayed effect argument is correct, these papers detected nothing real. If these papers detected real effects, the delayed effect argument fails. Khan et al. invoke both simultaneously without acknowledging the contradiction.

Several of our strongest arguments receive no response at all. The sex-specific divergence in suicide trends—rates among boys peaked around 2017–2018 and have since declined approximately 30 percent while rates among girls have continued to rise—is not mentioned. Since prescribing decisions are not made differently for boys and girls, a warning-driven suppression of prescribing cannot explain this divergence. The misrepresentation of the comparison group is similarly unaddressed: Suicide rates for adults aged 55–64, presented as an unaffected control population, increased by 25–44 percent over the same period, comparable to the increases in adolescents that Khan et al. attribute to the warning. The pharmaceutical marketing explanation for the 2005 prescribing decline—which affected every age group through 54 and coincided with an

estimated annual $800 million drop in manufacturer promotional spending—is ignored. The healthcare contact data showing that more than one third of people who die by suicide had no healthcare contact of any kind in the preceding year are not addressed. The argument that moving the warning to Section 5 would not change the clinical or legal considerations facing physicians receives no response.

Barrier claim is asserted without evidence / The central claim of their response—that the boxed warning creates meaningful barriers to care—is presented as self-evident but supported by no real evidence. No quantification is offered of how many patients who needed antidepressants failed to receive them because of the warning's status specifically as a box, as opposed to other sources of concern about the drugs causing suicidal thinking and behavior, and as distinct from the many other well-documented barriers to mental health care including access, insurance, stigma, and prescriber availability. They base this claim solely on unspecified clinical anecdotes of physicians reporting hesitancy; this appears to be no more than their subjective impression. Khan et al. provide no detail about what these anecdotes actually describe—even qualitatively—and do not describe any organized effort to collect clinical opinion. There is no transcript, no protocol, no indication of how the opinions were solicited, and no way to assess whether the questions put to physicians were neutral or leading. Even if such efforts were made, opinions gathered without methodological discipline are easily shaped—wittingly or unwittingly—by the framing of the inquiry, and what is presented as independent clinical testimony may in fact be the elicited echo of the interviewer's premises.

Khan et al.'s own rendering of these reports is itself processed through their priors: Authors who have argued for more than a decade that the warning causes harm are not neutral curators of physician sentiment; they are predisposed to motivated reasoning in the selection, summarization, and emphasis of such material. The reports themselves are also a self-selected sample: Physicians who feel constrained by the warning and want it removed have reason to come forward and be heard, while physicians who prescribe without difficulty have none. A body of testimony filtered in this way cannot establish the prevalence of the phenomenon it describes. What Khan et al. characterize as "real-

world effects" may amount to little more than the projection of their own opinions onto a sympathetic and unverifiable sample.

Anecdote sits at the bottom of the hierarchy of clinical evidence, and Khan et al. invoke it here to override administrative prescribing data. They offer no account of how a barrier-creating warning has presided over a 43 percent increase in adolescent prescriptions since 2016. Nor do they show that patients not prescribed antidepressants because of such hesitancy are those at high risk for suicide. Physicians unwilling to prescribe to high-risk patients would also be obliged to refer such patients to physicians willing to treat them properly.

Khan et al. also argue that the practical effect of a labeling change lags months to years behind the regulatory decision because dispensed inventory carries the old label until stock turns over. This claim reflects a fundamental misunderstanding of how physicians access prescribing information. Physicians decide whether to prescribe before the prescription is written and before any package insert reaches the patient. The insert that accompanies dispensed medication is a patient-facing document; it plays no role in the prescribing decision, which has already been made. Prescribers access drug labeling through electronic references, FDA databases, clinical decision support tools, promotional materials, and professional training.

Widely used drugs such as antidepressants do not require years for supplies to turn over; for most drugs and pharmacies, it is a few months. Most antidepressants have a shelf life of about two to three years from the date of manufacture. Drugs present on pharmacy shelves before the warning was issued would have been dispensed or expired well before youth suicide rates began their increase in 2008. This also directly contradicts the source on which Khan et al. primarily rely: Soumerai et al. (2024) describe the effects of the warning as "sudden, simultaneous, and sweeping," an account irreconcilable with a mechanism that operates gradually as old inventory clears.

The logic of the barrier argument also defeats itself. A physician who believes a patient is at high risk for suicide and thinks an antidepressant could reduce that risk has every reason to prescribe despite the warning because withholding treatment only increases the likelihood of the outcomes they fear, including a lawsuit should a suicide occur. The population for whom a physician could be deterred from prescribing by the

warning must therefore consist of lower-risk patients, for whom the drug-induced suicidality risk most plausibly exceeds the therapeutic benefit. The barrier Khan et al. invoke, if it exists at all, operates precisely where removing it would do the most harm.

These anecdotal reports also deserve scrutiny before being accepted as evidence of a genuine barrier. Managing emotionally troubled patients—particularly adolescents—is difficult, and many providers feel ill-equipped to do so. The invocation of the boxed warning may just be a convenient rationalization to avoid managing such patients. If so, changing the label would not address the underlying problem. The barriers Khan et al. describe are more plausibly symptoms of inadequate training and systemic gaps in mental health care than consequences of drug labeling, and those underlying problems would not be addressed by any change to the warning.

Self-referential citations / Khan et al. characterize the body of evidence supporting their position as "14 years of strong research," citing Soumerai et al. (2024) as its synthesis. Neither the systematic review nor the broader citation list survives scrutiny.

The Soumerai review is substantially self-referential. Christine Y. Lu appears as author or co-author on at least four of the 11 studies the review includes as primary evidence: Lu et al. 2014, Lu et al. 2018, Lu et al. 2020, and Soumerai et al. 2024 itself. Stephen Soumerai appears on two. Across Khan et al.'s broader citation list of 11 studies, the same two investigators account for the majority of entries. This is not an independent and converging literature; it is a self-referential citation network in which a small group of researchers repeatedly cites its own prior work as corroboration. That is not independent replication; it is iteration.

The selection criteria of the review favor a predetermined methodology. Of 1,841 studies screened, 11 were selected: overwhelmingly interrupted time series analyses and similar before-and-after comparisons, precisely the methodology Stone (2018a) documented in detail to be invalid for this purpose. Restricting a systematic review to a methodologically invalid study design does not overcome that invalidity.

Several of the 11 included studies use hospital admissions for poisoning (whether self-inflicted or not) by psychotropic agents as a proxy measure for suicide attempts. Stone (2018a) showed that this proxy was explicitly rejected as a useful proxy in the

Patrick paper that Lu and Soumerai misrepresent as validating it. In a published point–counterpoint exchange in the journal *Medical Care* (Lu et al. 2018; Stone 2018b), Stone charged that Lu et al. "knowingly and repeatedly misrepresented their proxy as peer-reviewed when the Patrick paper, in fact, rejected its use," and that "whether due to ineptitude or dishonesty, errors of this magnitude ethically require a paper's retraction." The Soumerai et al. (2024) systematic review was written with full knowledge of it and does not address it.

Lu defended the proxy by citing its 99.3 percent specificity. Stone (2018b) showed this statistic is meaningless here: Because suicide attempts are rare relative to all hospital admissions, almost any diagnosis code looks highly specific by this calculation. Appendectomy, which has no connection to suicide, scores 99.1 percent by the same method. The proxy also moves in the wrong direction: between 2004 and 2009, hospital admissions for psychotropic drug misuse were strongly negatively correlated (−0.85) with actual suicide attempt rates in national data.

The proxy measure has additional problems that Lu's response does not address. It does not require that the poisoning be a suicide attempt; many are accidental overdoses. But even if psychotropic drug poisonings were a generally satisfactory proxy for attempted suicide, it makes little sense to use it as a proxy when the concern is that FDA actions discouraged the prescription of antidepressants and other psychiatric care. Antidepressants and antipsychotics are psychotropic drugs. People who present with poisoning from antipsychotics or antidepressants cannot reasonably be considered to have been discouraged by FDA actions from receiving these drugs. According to the Drug Abuse Warning Network, nearly half of the psychotropic drugs involved in suicide attempts were antidepressants (29 percent) or antipsychotics (17 percent) (Stone 2018a).

As our *Regulation* article documents in detail, suicide rates fell for several years after the warning was issued, and prescribing has risen well above pre-warning levels throughout the period the review covers. The claim that consistency of observed harms across 14 years "indicates this is not a coincidence" is circular: Consistency across a literature built on shared methodological errors reflects consistency of method, not convergence of truth.

Khan et al.'s continued citation of Gibbons et al. 2007 is particularly difficult to justify. Our *Regulation* article documented—as Jureidini (2007) had pointed out in correspondence to the journal at the time of publication—that this paper presented two graphs not aligned to the same period: The suicide uptick it presents as evidence of harm occurred in 2004, before the warning existed, while the prescription downtick occurred in 2005, after it. An effect that preceded its alleged cause by a year is not evidence; it is a logical impossibility. The paper was never retracted despite this documented error and continues to be cited as though it had never been identified. Both Khan and Soumerai are well aware of this. For them, it appears, evidence is "strong" when it confirms their priors—logic, methodology, and documented error notwithstanding.

Khan's data transparency charge / Khan et al. state that Stone "obtained data from unknown pharmaceutical companies" and that the data "were confidential," framing this as a transparency concern. Both characterizations are incorrect. Stone et al. (2009) explicitly identify every product included in the analysis, and the sponsors of these products are publicly known: They are marketed pharmaceutical products whose manufacturers are a matter of public record. The data are not merely confidential; they are proprietary clinical trial data submitted to the FDA under legal obligation as part of the regulatory review process. Releasing such data to third parties without the companies' permission would violate federal law governing trade secrets and confidential commercial information. The analysis was examined and reviewed by both senior FDA officials and an advisory committee that had full access to the data but found no reason to doubt its conclusions.

The appropriate vehicle for transparency is peer-reviewed publication, which is precisely what occurred. The *BMJ*, which published Stone et al. 2009, maintains among the most rigorous peer review standards in medicine. Khan et al.'s demand for independent access to the underlying data, while ignoring the published analysis those data produced, is not a principled call for transparency; it is a rhetorical device for casting doubt on findings they have not otherwise attempted to refute.

There is also an irony in their demand. Authors who rest their central barrier claim on undocumented, unsourced, and unverifiable physician anecdote—with no transcript, no

protocol, and no accounting of who said what to whom—are not well placed to insist on independent access to data already submitted under legal obligation, published in peer-reviewed form, and reviewed by the FDA and its advisory committee. The standard of disclosure they demand of others is one they conspicuously decline to meet themselves.

The irony extends further: Having dismissed methodological critique as a technicality, they nonetheless invoke a methodological argument—the non-significance of individual drug analyses—to cast doubt on the pooled finding. However, all risk ratios shown in their table are greater than one, which means that in each case there were more suicidal events with the active drug than placebo; when combined, the excess was statistically significant. Objecting to the result of the combined analysis because the findings for the drugs considered separately are not significant is a misapplication of statistical reasoning. Non-significance in individual drugs evaluated in two to five trials each is precisely the problem that meta-analysis exists to address. When events are rare and sample sizes are small, individual studies lack the statistical power to detect real associations. Pooling data across trials and drugs is not a methodological convenience; it is the only valid way to evaluate a signal under these conditions. Khan et al.'s argument reflects a misunderstanding of why meta-analysis is conducted in the first place. Researchers who do not grasp this are not in a position to conduct the independent analysis they are requesting.

Self-refuting European evidence / Khan et al. cite Bachmann et al. (2016) in support of their European argument. Bachmann's data do the opposite. From 2005 to 2012, antidepressant (ATD) use increased markedly in both the United States and Europe, with usage rates consistently higher in the United States: "In 2012, ATD prevalence was 1.6% (US), 1.1% (UK), 1.0% (DK), 0.6% (NL) and 0.5% (DE). Increase was greatest in 10−14-year-olds (NL, UK) and 15−19 year olds (DK, DE, US)." If prescribing was rising on both sides of the Atlantic, and was highest in the United States, the boxed warning cannot explain divergent suicide trends. Once again, Khan et al. cite as support evidence that directly contradicts their own argument.

Khan et al. also argue that the European approach preserves clinical discretion while the FDA warning constrains it. This is exactly backwards. The FDA warning does not

advise against prescribing; it instructs clinicians to weigh documented risk against clinical need. European labeling, by contrast, states that antidepressants should not be used to treat depression in children and adolescents under 18 except under limited circumstances. That is a direct restriction on prescribing, and one the FDA warning does not impose.

The difference is equally clear in approved indications. In Europe, only one antidepressant—fluoxetine—has a pediatric indication, and that indication is narrow: moderate to severe major depression, only after psychological therapy alone has failed after four to six sessions, and only in combination with ongoing psychotherapy. In the United States, by contrast, *seven* antidepressants have FDA approval for pediatric use across a range of indications (Table 1). If youth self-harm mortality declined in the European Union under this more restrictive framework, that fact cuts directly against Khan et al.'s claim that the FDA warning uniquely created harmful barriers to care.

Risk side is ignored / A proposal to remove a drug safety warning cannot be taken seriously unless it shows that the benefits of doing so outweigh the risks. Khan et al. do not come close. The confirmed risk of drug-induced suicidal behavior in young patients—an odds ratio of 2.30 in adults under 25, statistically significant across multiple methods, according to the most comprehensive analysis of the question ever conducted (Stone et al. 2009)—receives no substantive response. Khan et al. question the analysis on procedural grounds addressed elsewhere in this reply, but they do not argue the finding is wrong, and they cite no evidence that it has been superseded.

They do not apply the same scrutiny to the evidence for benefit, which is considerably weaker. Clinical trial data show that approximately 85 percent of patients with major depressive disorder derive no drug-specific benefit beyond placebo, with even weaker efficacy in children and adolescents (Stone et al. 2022). Expanding prescribing through warning removal therefore exposes more patients to the risk of treatment-emergent suicidality without a proportionate increase in therapeutic benefit. For patients with less severe depression or non-depression indications such as generalized anxiety disorder—an increasingly common indication for adolescent prescribing—the risk of drug-induced suicidal thinking or behavior may exceed the risk posed by the disorder

itself. A proposal that interrogates the risk evidence while taking the benefit evidence on faith is not a balanced policy argument.

Attribution and disclosure / Khan et al. (2026b) acknowledge that our pre-submission review was quoted without attribution and that the wording incorrectly implied current FDA affiliation. We appreciate the partial acknowledgment, but the problem was not a failure of courtesy; it was a false attribution. Quoting a passage from a private review and linking it to "the FDA" as an institution misrepresents both the source and the standing of the critique.

Khan et al. repeat the meritless claim, originating with Lu and Soumerai (2018), that Stone failed to disclose his FDA affiliation in his publications. Not only did his affiliation appear with the published pieces, the articles were written at the invitation of the journals, and the editors were fully aware of his position at FDA. In direct contradiction of this accusation, Khan et al. also suggest that those publications represented Stone speaking for the FDA. They did not; each explicitly stated that the views expressed were Stone's own and did not represent the FDA's position. As for their claim that they were unaware of the authors' current affiliations, two of the three of us were contacted by Khan et al. directly through our LinkedIn profiles and personal email addresses, not FDA addresses, channels that display institutional affiliations explicitly. Making a claim when in possession of information that directly contradicts it is consistent with the pattern their responses exemplify throughout.

Finally, the characterization of the Stone et al. 2009 meta-analysis as "not designed for detecting suicidality" is incorrect. The FDA specifically requested that manufacturers submit trial data for the purpose of examining suicidality risk across the age spectrum. Every event was adjudicated using a validated, blinded assessment methodology designed precisely for that purpose. The trials provided the data; the meta-analysis provided the purpose-built analytical framework. Furthermore, Khan et al. fail to appreciate that any inaccuracy in the identification of suicidality would diminish the possibility of a positive finding, not promote a spurious one; ironically, their objection actually supports the conclusions of the study.

Conclusion / Khan et al. (2026a) began by asserting that the FDA's antidepressant boxed warning caused the deaths of thousands of American children. When examined against the evidence—including evidence they themselves cited—that assertion failed at every link in the causal chain. Their response (2026b) does not repair those failures. It introduces a delayed-effect argument incompatible with the literature they invoke, relies on a systematic review built largely from the review authors' own prior work, asserts barriers to care without evidence, and ignores most of the specific arguments we made. It asks to weaken a prominent drug safety warning without confronting the risk such a change would create for the very patients the warning is meant to protect. What remains, after the rhetoric is stripped away, is not a substantiated case for policy change but unchecked confirmation bias: questionable evidence that confirms their conviction is accepted uncritically, while unequivocal evidence that contradicts it—the falling suicide rates during the prescribing decline, the mistiming of their own cited studies, the documented and unretracted errors—is dismissed without cogent explanation or simply ignored. The aphorism they deploy against us—"I wouldn't have seen it if I hadn't believed it"—describes their own approach with precision: For these authors, belief does not follow the evidence, it dictates which evidence is permitted to count. In the end, their position is less an argument than an act of denial running in two directions at once: The confirmed risk that the drugs themselves cause suicidality does not register, while the harms they ascribe to the warning—refuted at every step in our original response—remain to them as real as ever. They do not want to believe it, and so they do not see it.

—Marc Stone, MD; Andrew Mosholder, MD, MPH; and Tarek Hammad, MD, PhD, MSc, MS, FISPE

Readings

Bachmann, Christian J., Lise Aagaard, Mehmet Burcu, et al., 2016, "Trends and Patterns of Antidepressant Use in Children and Adolescents from Five Western Countries, 2005–2012," *European Neuropsychopharmacology* 26(3): 411–419.

Chua, Kao-Ping, Anna Volerman, Jason Zhang, et al., 2024, "Antidepressant Dispensing to US Adolescents and Young Adults: 2016–2022," *Pediatrics* 153(3): e2023064245.

Hammad, Tarek A., and Justine Rochon, 2026, "Analytic Misjudgment of Drug Safety Evidence and Causality: From the Prosecutor's Fallacy and Simpson's Paradox to Artificial Intelligence," *Drug Safety*, in press.

Gibbons, Robert D., C. Hendricks Brown, Kwan Hur, et al., 2007, "Early Evidence on the Effects of Regulators' Suicidality Warnings on SSRI Prescriptions and Suicide in Children and Adolescents," *American Journal of Psychiatry* 164(9): 1356–1363.

Jureidini, Jon, 2007, "The Black Box Warning: Decreased Prescriptions and Increased Youth Suicide?" *American Journal of Psychiatry* 164(12): 1907.

Khan, Arif, Anshu Arora, and Aishwarya Prasad, 2026a, "A Tragic Unintended Consequence," *Regulation* 49(1): 26–29.

Khan, Arif, Anshu Arora, Aishwarya Prasad, et al., 2026b, "Barriers to Care Remain," *Regulation* 49(2): 44–47.

Lu, Christine Y., Fang Zhang, Matthew D. Lakoma, et al., 2014, "Changes in Antidepressant Use by Young People and Suicidal Behavior after FDA Warnings and Media Coverage: Quasi-Experimental Study," *BMJ* 348: g3596.

Lu, Christine Y., Gregory Simon, and Stephen B. Soumerai, 2018, "Staying Honest When Policy Changes Backfire," *Medical Care* 56(5): 384–390.

Patrick, Amanda R., Matthew Miller, Catherine W. Barber, et al., 2010, "Identification of Hospitalizations for Intentional Self-Harm When E-Codes Are Incompletely Recorded," *Pharmacoepidemiology and Drug Safety* 19: 1263–1275.

Soumerai, Stephen B., Ross Koppel, Huseyin Naci, et al., 2024, "Intended and Unintended Outcomes after FDA Pediatric Antidepressant Warnings: A Systematic Review," *Health Affairs* 43(10): 1360–1369.

Stone, Marc B., 2014, "The FDA Warning on Antidepressants and Suicidality—Why the Controversy?" *New England Journal of Medicine* 371: 1668–1671.

Stone, Marc B., 2018a, "In Search of a Pony: Sources, Methods, Outcomes, and Motivated Reasoning," *Medical Care* 56(5): 375–381.

Stone, Marc B., 2018b, "Honest Opinion versus Dishonest Analysis," *Medical Care* 56(5): 391–393.

Stone, Marc, Thomas Laughren, M. Lisa Jones, et al., 2009, "Risk of Suicidality in Clinical Trials of Antidepressants in Adults: Analysis of Proprietary Data Submitted to US Food and Drug Administration," *BMJ* 339: b2880.

Stone, Marc B., Zimri S. Yaseen, Brian J. Miller, et al., 2022, "Response to Acute Monotherapy for Major Depressive Disorder in Randomized, Placebo Controlled Trials Submitted to the US Food and Drug Administration: Individual Participant Data Analysis," *BMJ* 378: e067606.

Stone, Marc, Andrew Mosholder, and Tarek Hammad, 2026, “In Defense of the Boxed Warning,” *Regulation* 49(2): 38–44.

Table 1

FDA-Approved Pediatric Indications for Antidepressants

Drug	Indication	FDA-Approved Age Range
Fluoxetine	Major Depressive Disorder	8–18 years
	Obsessive-Compulsive Disorder	7–17 years
Escitalopram	Major Depressive Disorder	12–17 years
	Generalized Anxiety Disorder	7–17 years
Sertraline	Obsessive-Compulsive Disorder	6–17 years
Fluvoxamine	Obsessive-Compulsive Disorder	8–17 years
Clomipramine	Obsessive-Compulsive Disorder	10–17 years
Duloxetine	Generalized Anxiety Disorder	7–17 years
Imipramine	Childhood Enuresis	6–17 years