Response to "Black Box Warning Did Not Cause Increased Suicides"

Stephen B. Soumerai, ScD, Robert B. Penfold, PhD, Anne M. Libby, PhD, Christine Y. Lu, PhD, MSc

In their letter regarding our recent PRCP study (1), Spielmans et al. demonstrate a lack of familiarity with rigorous quasi-experimental research designs. Such designs, however, are essential in studies of health policies, which can seldom or ever be randomized, for example, one can't issue national drug safety warnings to a random sample of the population. Before responding to their specific conclusions, we would like to refer readers to an informative table of the hierarchy of strong and weak research designs. Table 1 is based on 100s of years of science (2). It shows a hierarchy of strong research designs that often yield valid results, in contrast to weak designs without baselines (cited by Spielmans et al.) that are largely untrustworthy (i.e., post-only designs without baselines cannot control for common biases, such as history bias and secular trends) (3).

Further guidance on research design hierarchy is available (4–6).

Spielmans et al. critique our strong interrupted time-series (ITS) with comparison series study (multi-age groups used as comparisons) by citing uncontrolled post-only designs—which are at the bottom of the hierarchy of research designs—alleging that our study proved no effects of black box warnings on antidepressant use. Without a baseline it is impossible to estimate the counterfactual pre-intervention trend (what would have happened in the policy's absence). The simple truth is that it is misleading to attempt to measure a change occurring after a policy is enacted in the absence of any baseline (pre-intervention) measure.

Spielmans et al. conclude that treatment of youth depression has not declined substantially since the warnings, and they defend this statement by a misleading and selective observation that "Lu et al (our previous study) found a decrease in ...treatment of less than one percentage point." This conclusion is false when one simply observes the sudden and sustained change in trend of antidepressant use after a substantial increase in trend during the 14 quarterly baseline observations before the advisory (Figure 1).

What is immediately apparent in Figure 1 is that the "less than one percentage point" reduction in treatment was from less than a 2 percent prevalence of treatment just ***before*** the warnings. (The denominator was all adolescents and the relative reduction was greater than 30 percent. Thus, by ignoring the relative reduction, Spielmans et al. understate the effect by more than 30 times.) The difference (effect) between the counterfactual baseline trend and the actual observation at the time of the second black box warning is almost 50% of all adolescents per quarter (approximately 1.1 million adolescents in the 11 US health systems). Failing to provide both absolute and relative changes in effect estimates is a major bias in both media and scientific reporting of the effects of health technologies and policies. Such failings distort findings in ways that adversely affects both science and health policies, sometimes with patient harms (7, 8).

The sudden reduction in the above ITS and comparison series study is evident to a non-scientist—the tremendous decrease in both level and slope, controlling for the rising baseline trend probably persisted for about 7 years following the initial advisory warning in the fall of 2003. But we conservatively estimated medication use and suicidality effects only for several years beyond the warnings because the confidence intervals of ITS effects gradually become wider over time.

The other studies cited by Spielmans et al. to refute our results (that the warnings reduced treatment) are weak post-only designs (Table 1) with uninterpretable findings that violate the basic research design criteria of the worldwide Cochrane Collaboration's systematic reviews (9–11). Post-only studies are excluded altogether from any rigorous Cochrane reviews (12). Studies with insufficient baseline trends (pre-post designs) also offer weak, if any, evidence of causal inference. For example, the Valluri study has only three monthly data points before the first advisory (Oct. 2003) and none from prior years. The data showing the well-known, steep rise in antidepressant use before the first advisory (see baseline in Figure 1) are missing in the Valluri paper; absent these critical observations of baseline trends, their data are insufficient. Moreover, their paper lacks any data occurring during the

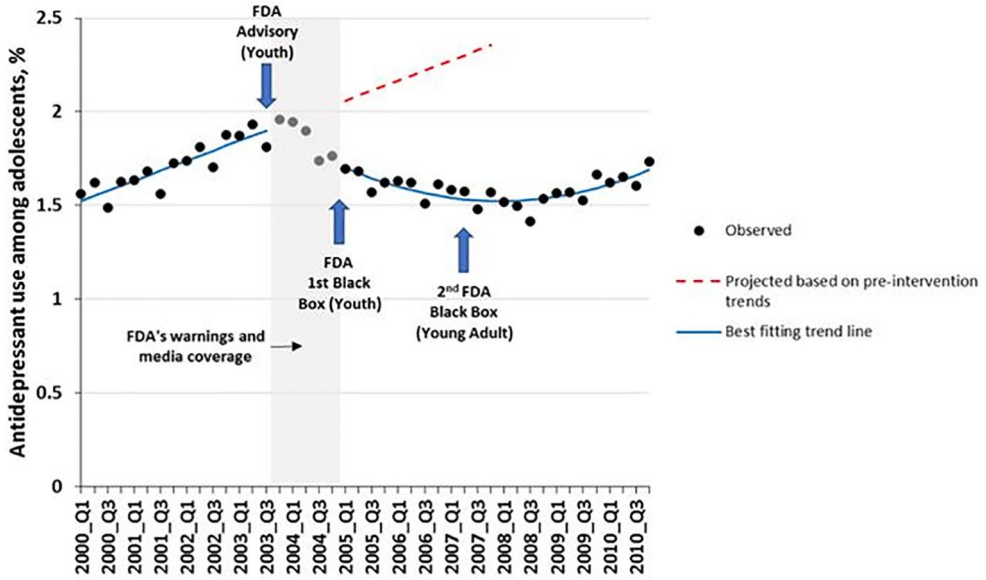
[Corrections added on March 16, 2021 after first online publication: The title has been revised.]

TABLE 1. Hierarchy of strong designs and weak designs, based on design's capacity to control for most biases

Hierarchy of designs for interventions	
Strong designs	
Multiple randomized controlled trials	The "gold standard" of evidence, incorporating many RCTs of an intervention
Randomized controlled trials	A strong design, but sometimes not feasible or generalizable, especially for health policies
Interrupted time series with a control series	Strong quasi-experimental design that controls for common biases. This design has two controls: baseline trend and control group to measure sudden discontinuities in trend soon after an intervention.
Intermediate designs	
Single interrupted time series	Measures changes in level or slope of trend controlling for baseline trend, but has no comparison group
Before and after with comparison group (single observations, sometimes called "difference in difference" design)	Pre-post changes using single observations. Comparability of baseline trend often unknown
Weak designs	No controls for common biases, excluded from literature syntheses
Uncontrolled before and after (pre-post)	Single observation before and after, no baseline trends
Cross-sectional or post-only designs	Simple correlation, no baseline, no measure of change from before intervention

Source: Soumerai SB et al. Prev Chronic Dis. 2015; 12:E101.

FIGURE 1. Rates of antidepressant use per quarter before and after the warnings among adolescents enrolled in 11 health plans in nationwide mental health research network. BMJ. 2014; 348:g3596.



second black box warning in 2007, as well as any other year after the warning. Effect estimates (change from before to after) are impossible without reliable measures of the pre- and post-policy trends. Similarly, Kafali et al cannot reliably measure pre-advisory counterfactual trends with only three points. Baseline trends can be rising, stable or falling; without adequate baseline data, attempts to measure before-to-after changes are biased, and all too frequently misleading and deceptive.

Speilmans et al. then cite post-only correlations between antidepressant use and suicide attempts only after the warnings began, representing ecological fallacies without any baseline. This claim is based on Ploderl's study using self-reported antidepressant use and self-reported suicide attempts. Correlations based on self-reported measures are often severely compromised by recall and social desirability biases.

In their previously published narrative review, Speilmans et al. (13) incorrectly depict our prior ITS study as an ecological study examining the relationship between antidepressants and suicide attempts. Narrative reviews are inadequate for informing policy-making because they do not assess the methodological quality of studies in the field before summarizing credible results. Our ITS studies and

other studies that we cited (14–21) employed rigorous quasi-experimental research methods to examine the effects of FDA warnings on antidepressant use, non-drug treatments, suicide attempts, and suicides. As shown in Table 1 and in major research design texts and the Cochrane Collaboration, research on drug safety policies require strong quasi-experimental designs (preferably ITS with comparison series to assess interruptions in trends, controlling for pre-policy levels and slopes). The ITS designs can control for many biases, such as history, maturation, and selection (4). Our study measured the multifactorial effects of risk communications. We did not, as stated by Spielmans, study "whether a drug causes suicidality." Nor did we measure the adverse effects of policy-induced reductions in medications alone. Most of the effects were related to demonstrated reductions in both drug and non-drug depression treatment rates following the warnings (14–21).

The majority of longitudinal ITS studies have demonstrated, in different large samples (including national), that the youth antidepressant warnings have almost simultaneous, unintended effects on identification of depression, psychotherapy, antidepressant treatment, suicidal behavior, and suicide deaths (15–21). They cause sudden shifts in the level and slopes of the trends. A public health policy analysis cannot ignore this number of simultaneous unintended outcomes in different datasets. The burden of proof of policy harms should be on the policymakers creating those policies, not on the outside scientists who have no or fewer conflicts of interest (22).

Together, findings from these ITS studies (including our own) suggest the boxed warnings may have contributed to the very thing FDA was trying to prevent: youth suicidal behavior and suicides. Even before COVID, more than two thirds of depressed teens did not receive any depression care (23), an issue now further exacerbated by both the pandemic and the continuing barrage of alarming suicidality warnings contained in all drug labels and TV advertisements. It is time for the FDA to err on the side of caution and reduce the severity of the continuing antidepressant warnings.

AUTHOR AND ARTICLE INFORMATION

Department of Health Services Research, Kaiser Permanente Washington Health Research Institute and University of Washington, Seattle, (Penfold); Department of Emergency Medicine, School of Medicine, University of Colorado, Anschutz Medical Campus, Denver, (Libby); Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, (Soumerai, Lu).

Send correspondence to Dr. Soumerai (ssoumerai@hms.harvard.edu).

The authors declare no conflict of interest.

This is a Response to a Letter available here: [https://doi.org/10.1176/appi.prcp.20200038]. The Article that is the subject of the Letter is available here: [10.1176/appi.prcp.20200012].

Received December 30, 2020; revised December 30, 2020; accepted January 3, 2021.

REFERENCES

1. Spielmans G, Parry P. Black box warning did not cause increased suicides. Psychiatr Res Clin Pract 2021:1. https://doi.org/10.1176/appi.prcp.20200038
2. Semmelweis I. The etiology, concept, and prophylaxis of childbed fever. Budapest and Vienna, Hartleben, 1861
3. Naci H, Soumerai SB. History bias, study design, and the unfulfilled promise of pay-for-performance policies in health care. Prev Chronic Dis 2016; 13:E82
4. Shadish W, Cook T, Campbell D. Experimental and quasi-experimental designs for generalized causal inference. Belmont, CA, Wadsworth Cengage Learning, 2002
5. Soumerai SB, Ceccarelli R, Koppel R. False dichotomies and health policy research designs: randomized trials are not always the answer. J Gen Intern Med 2017; 32:204–209
6. Wagner AK, Soumerai SB, Zhang F, et al. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther 2002; 27:299–309
7. Soumerai SB, Starr D, Majumdar SR. How do you know which health care effectiveness research you can trust? A guide to study design for the perplexed. Prev Chronic Dis 2015; 12:E101
8. Moynihan R, Bero L, Ross-Degnan D, et al. Coverage by the news media of the benefits and risks of medications. N Engl J Med 2000; 342:1645–1650
9. Valluri S, Zito JM, Safer DJ, et al. Impact of the 2004 Food and drug administration pediatric suicidality warning on antidepressant and psychotherapy treatment for new-onset depression. Med Care 2010; 48:947–954
10. Plöderl M, Hengartner MP. Antidepressant prescription rates and suicide attempt rates from 2004 to 2016 in a nationally representative sample of adolescents in the USA. Epidemiol Psychiatr Sci 2019; 28:589–591
11. Kafali N, Progovac A, Hou SS-Y, et al. Long-run trends in antidepressant use among youths after the FDA black box warning. Psychiatr Serv 2017; 69:389–395
12. What study designs can be considered for inclusion in an EPOC review and what should they be called? [Internet]. Oxford, UK, Cochrane Effect Practice and Organisation of Care (EPOC, 2017. https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/what_study_designs_should_be_included_in_an_epoc_review.pdf
13. Spielmans GI, Spence-Sing T, Parry P. Duty to warn: antidepressant black box suicidality warning is empirically justified. Front Psychiatry 2020; 11:18
14. Lu CY, Penfold RB, Wallace J, et al. Increases in suicide deaths among adolescents and young adults following US Food and Drug Administration antidepressant boxed warnings and declines in depression care. Psych Res Clin Pract 2020; 2(2):43–52. https://doi.org/10.1176/appi.prcp.20200012
15. Libby AM, Orton HD, Valuck RJ. Persisting decline in depression treatment after FDA warnings. Arch Gen Psychiatry 2009; 66:633–639

16. Libby AM, Brent DA, Morrato EH, et al. Decline in treatment of pediatric depression after FDA advisory on risk of suicidality with SSRIs. Am J Psychiatry 2007; 164:884–891
17. Suicidality in children and adolescents being treated with antidepressant medications [Internet]. Silver Spring, MD, The Food and Drug Administration, 2018. http://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/suicidality-children-and-adolescents-being-treated-antidepressant-medications
18. Lu CY, Zhang F, Lakoma MD, et al. Changes in antidepressant use by young people and suicidal behavior after FDA warnings and media coverage: quasi-experimental study. BMJ 2014; 348:g3596
19. Busch SH, Frank RG, Leslie D, et al. Antidepressants and suicide risk: how did specific information in FDA safety warnings affect treatment patterns? Psychiatr Serv 2010; 61:11–16
20. Kurian BT, Ray WA, Arbogast PG, et al. Effect of regulatory warnings on antidepressant prescribing for children and adolescents. Arch Pediatr Adolesc Med 2007; 161:690–696
21. Katz LY, Kozyrskyj AL, Prior HJ, et al. Effect of regulatory warnings on antidepressant prescription rates, use of health services and outcomes among children, adolescents and young adults. Can Med Assoc J 2008; 178:1005–1011
22. Soumerai S, Koppel R. How a nonfinancial conflict of interest can threaten public health [Internet]. STAT News. 2020. https://www.statnews.com/2020/10/07/nonfinancial-conflict-of-interest-threaten-public-health/
23. Mojtabai R, Olfson M, Han B. National trends in the prevalence and treatment of depression in adolescents and young adults. Pediatrics 2016; 138.e201618