

HaRRT Center's Response to the Washington Policy Center's "National Retraction Brief"

August 16, 2026

On August 10, 2026, the Washington Policy Center's Eric Zimmerman began distributing a document titled "National Retraction Brief: Widespread and systematic factual errors in Law Enforcement Assisted Diversion (LEAD)," which in part alleges that the published findings of the HaRRT Center's LEAD evaluation are inaccurate and that the program's evidence base does not support its adoption. This response addresses those allegations from the perspective of the evaluators who conducted the work.

The Brief's central allegation laid out in its "Summary" is that the evidence base "contains significant factual and statistical errors," which it describes as "large in both number and size." However, the Brief provides no statistical basis for many of its assertions, such as "claims typically more than double LEAD's possible effectiveness." Also of note, "effectiveness" is not a statistical quantity. Where the Brief provides numbers, they are inaccurate. For example, by apparently mistaking our reported average monthly statistics to be a raw headcount, the Brief incorrectly asserts LEAD housed 13% of participants who needed housing at baseline. That number was actually 31%. The Brief incorrectly asserts "just a 35% success rate in connecting unhoused participants with either shelter or housing." That number was actually 81% (or 70% when considering only those unsheltered at baseline). It incorrectly reports that employment gains amounted to "a 3-person difference"; it was 11 participants who gained employment and 15 who moved into vocational training, employment or retirement.

The Brief also incorrectly asserts that the program evaluation shows "little or no clear positive intervention effect." On the contrary, published and peer-reviewed between-groups analyses were statistically significant and showed meaningful reductions in the odds of arrest, felony charges, and prison incarceration, as well as statistically significant reductions in jail days, jail bookings, and system costs. We provide a summary of the accurately reported, peer-reviewed findings below.

The Brief makes additional allegations of "lost statistical control, artificial influences, and midstream methodological changes." But the terms used here are undefined and have no standard scientific meaning. They also omit the fact that, for over a decade now, we have transparently and publicly posted complete reports that discuss the limitations common to many real-world program evaluations alongside the widely used methodological and statistical steps we took to address them (i.e., propensity score weighting, group balance checks, robust standard errors, and analyses referenced to each participant's own baseline). Finally, although the Brief is titled "National Retraction Brief," it identifies no scientific basis for a retraction, and its own recommendations do not call for one.

One point in the Brief is well-taken and is being corrected. In an early summary and Aim 3 reporting, we used the word "likelihood" in place of "odds." Translating odds ratios into plain language, especially for lay materials, is a recognized difficulty. These have been corrected where we find them, and we have advised the LEAD team to correct their reporting as well. Although we were not responsible for additional reporting and were not typically consulted on how the program evaluation was discussed, we are happy to help correct the record. We are also supportive of organizations doing their best to use and discuss the evidence base in planning community-based interventions.

With that said, the statistical findings of the original HaRRT Center LEAD evaluation -- published across three peer-reviewed articles and four reports -- still stand. In the pages that follow, we restate for the record the LEAD findings and then respond to the Washington Policy Center's "National Retraction Brief" in more detail.

A decade later, the findings from the initial LEAD evaluation still stand.

Aim 1: LEAD was associated with meaningfully and statistically significantly lower odds of arrest and felony charges.

Compared with the system-as-usual control group, LEAD participants had 60% lower odds of arrest over the six-month follow-up window. Over the evaluation-wide window, they had 58% lower odds of arrest and 39% lower odds of a felony charge. Pre-to-post arrest prevalence fell from 83% to 58% among LEAD participants and rose from 77% to 80% among controls over the evaluation-wide window (see Figure 2 reproduced from Collins et al., 2017). There were no significant differences on total charges, or on felony charges in the six-month window, and we reported those null results in all documents. The peer-reviewed and published findings are available [here](#).

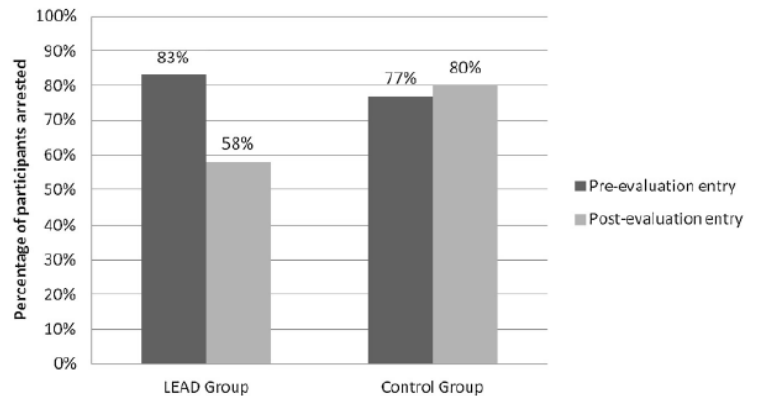
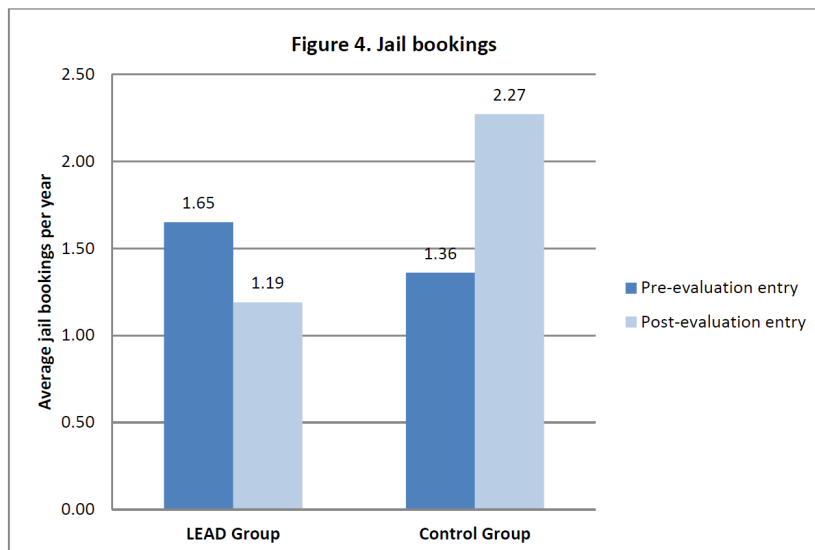


Fig. 2. Percentage of participants with ≥ 1 arrest over the entire LEAD evaluation (excluding index arrest).

Aim 2: LEAD was associated with significantly fewer jail bookings, fewer jail days, fewer felony cases, lower odds of prison incarceration, and lower system costs. Relative to controls, LEAD participants had approximately 1.4 fewer jail bookings per year, 41 fewer jail days per year, 0.42 fewer felony cases per year, and 88% lower odds of prison incarceration ($p < .001$). Misdemeanor cases showed no significant group difference. Estimated criminal justice and legal system costs fell by approximately \$2,100 per LEAD participant per year while rising by approximately \$5,961 per control participant, which was a statistically significant between-groups difference.



The figure to the left shows the raw pre- and post-referral values alongside them. We report both model estimates and raw values because each describes the data in different ways. The model estimates describe differences in changes between groups, and the raw values shown in the figure show the change within each separately. Utilization figures are counts drawn from jail, prison, prosecution and public-defense administrative records; cost figures are average-cost estimates whose every input and source is published. You can read the peer-reviewed findings [here](#).

Aim 3: Compared to their baseline status, LEAD participants had significantly higher odds of being sheltered, being housed, having legitimate income, and being on the employment continuum following referral. During the 18-month follow-up, LEAD participants' odds of being sheltered more than doubled ($OR = 2.08, p < .001$), and they had 89% higher odds of obtaining housing ($OR = 1.89, p < .001$) relative to baseline, with each case-management contact associated with incremental improvement in both. Odds of having legitimate income or benefits were 33% higher ($OR = 1.33, p = .002$), and odds of being on the employment continuum — vocational training or internship, legitimate paid employment, or retirement from legitimate employment — were 46% higher ($OR = 1.46, p = .013$). Binary employment status (employed versus unemployed) did not change significantly. Secondary analyses found that months housed and months on the employment continuum each predicted lower

odds of arrest. These analyses had no control group. We stated in both the report and the article that the findings from this aim cannot support causal conclusions, and that statement stands. These findings are published [here](#).

Aim 4: Participants described the program favorably in their own words. Thirty-two LEAD participants were interviewed in mid-2015 about their experience of case management. All but one described a positive experience; the exception reported a mixed view centered on not having obtained housing through the program. Approximately one-third said LEAD was the first social services program they had ever participated in. Participants described the case-management approach as client-centered and nonjudgmental, and identified the absence of an abstinence requirement as a reason they engaged. This was a qualitative study of a small volunteer subsample and was reported as such; it describes participant perceptions and is not a test of program effect. Although these findings were not subsequently published in peer-reviewed journals, they are available at request to the authors (harrrt@uw.edu).

Responses to specific allegations in the Washington Policy Center’s “National Retraction Brief.”

Allegation	Accuracy of allegation	Data
Summary		
"LEAD's 'benefit', as portrayed in most public sources, is not supported by causal evidence."	OVERSTATED	Aims 1 and 2 compared LEAD participants against a system-as-usual control group in a nonrandomized controlled design. That design allows for causal attribution conditional on measured confounding having been addressed — which is why we used propensity score weighting, published the balance checks including the covariate they did not resolve, used robust standard errors, and tested change accounting for participants’ baseline levels. We stated in our various articles and reports that these steps increased our confidence that the observed effects were attributable to LEAD rather than to other factors. Aim 3 had no control group at all, and we have clearly stated in published materials that causal conclusions cannot be drawn from it. Aim 4 was qualitative and describes participant perceptions rather than program effects. For aims 1 and 2, the design constrains the strength of attribution, not the existence of the differences. The reductions in arrest, felony charges, jail bookings, jail days, prison incarceration and system costs are statistically significant and drawn from administrative records held by five separate agencies.
Section I — “Scale and Significance ...”		
“LEAD program’s research demonstrates an uncertain instead of a clear reduction in recidivism—the tendency of a person to reoffend—among its participants.”	UNSUPPORTED	The reductions in arrest, felony charges, jail bookings, jail days and prison incarceration were each statistically significant, and were measured in independent administrative systems. We address specific outcomes in our Section III responses below.
“These claims typically more than double LEAD’s possible effectiveness...treat null or negative findings as successes.”	UNSUPPORTED	Effectiveness is not a statistical quantity, and no arithmetic is offered for "typically" or for "double." We reported both statistically significant <i>and</i> null findings in our publications; we did not interpret null findings as significant findings. In fact, one of the Brief's own Section III headings restates a null result we published ourselves. That said, the Brief incorrectly nullifies one positive finding (i.e., statistically significant increase in odds of being on the employment continuum from baseline to follow-up for LEAD participants). We address this in Section III below.
“Three times as many LEAD participants died” as gained jobs	FALSE	First and most important, these were 9 people who died. Using them as a multiplier against employment counts is not an analysis, and it is not an appropriate way to discuss the deaths of community members. Second, the comparison does not hold methodologically. The deaths are a cumulative person count across the 3-yr active project period, whereas the employment figure is a monthly-average prevalence among the LEAD participants who had case-

management data available, each followed 18 months from their own referral. These are different denominators, different observation periods and different quantities. (Of note, deaths in the control group could not be systematically documented, so group comparisons on mortality are unreliable, as we stated in the published article.)

Third, the Brief's math is incorrect. The percentages in our reporting do not represent a person count. The Brief does not show its derivation, but the 7.43% and 9.03% it appears to draw on are monthly-average prevalences, not counts of individuals. If we consider a person-level count among the participants we have data on: Of the 162 participants not employed at baseline, 11 became newly employed at some point during the follow-up, bringing the total employed during follow-up to 23. Of the 160 not on the employment continuum at baseline, 15 joined during the follow-up, bringing that total to 30.

Section II — "Inaccuracy of LEAD Claims"

The four quoted formulations	NOT OURS	None of the four formulations is a HaRRT Center document. On reviewing the Brief, we notified the LEAD team of the wording issue so that materials under their control can be corrected as well.
"Common incorrect versions include...Odds of being on employment continuum increased 46%"	FALSE	Although the Brief's author indicates this finding is incorrect, that is not true. This is the correct reporting. The odds of LEAD participants being on the employment continuum indeed increased 46% from baseline.
"These formulations regularly confuse concepts like odds, probability, and absolute versus relative change. They also substantially overstate the study's findings."	PARTIALLY TRUE	In an early summary and in the Aim 3 reporting, we used the word "likelihood" where we should have used the word "odds." Translating odds ratios into plain language, especially for lay materials, is a recognized difficulty. These have been corrected where we have found them, and we have advised the LEAD team to correct their reporting as well. However, this one-word change does not negate the statistically significant findings. The LEAD group showed significantly lower odds of arrest, felony charges and prison incarceration, and significantly fewer jail bookings and jail days, with lower system costs as shown on pgs 2-3.
"The original research shows little or no clear positive intervention effect in these measures."	FALSE	Our program evaluation showed consistently significant intervention effects, as presented above on pgs 2-3.
"The study is also separately impacted by what appears to be an early loss of statistical control, a high incidence of methodological artifacts, and significant design limitations which prevent it from establishing causality."	UNSUPPORTED	"Loss of statistical control" and "methodological artifacts" are not defined and no instance of these is given. The design limitations we note are common to program evaluations conducted in real-world systems and are transparently published in every document we have generated under a designated "Limitations" section. The methods we used to address them are also commonly used by professionals in the context of real-world program evaluations (e.g., propensity score weighting, robust standard errors, and reporting raw

values alongside model estimates). These are described in full in each publication.

Section III Introduction

“Though a challenging and uncertain method for small incoherent groups, this produced ‘odds ratios’ as statistical outputs.”

FALSE

It is unclear why the author finds standard statistical methods to be “challenging and uncertain” or what is being referred to as “small incoherent groups.”

However, it is false that propensity scores “produce odds ratios.” Propensity scores were entered as sampling weights for Aims 1 and 2. The odds ratios came from the generalized estimating equation models.

Section III: A) Charges filed

“A) Charges filed (No statistically significant differences)”

PARTLY FALSE

Felony charges are charges filed, and they differed significantly in favor of the LEAD group across the longer study period. A heading stating that charges filed showed no significant differences is therefore false. The text beneath the heading is narrower and accurate in part: Total charges showed no significant group differences, and we published that null result without qualification.

Section III: B) Arrests

“Arrests (unclear differences)”

FALSE

The differences are not unclear. There are statistically significant differences, they can be “reliably claimed,” and they have been published in peer-reviewed journals.

“No clear difference in arrests rates can be reliably claimed, but also cannot be altogether ruled-out.”

“Incorrect Claim: ‘LEAD reduced recidivism for all crimes by 58%’”

NOT OUR ASSERTION

This formulation is not ours. Collins et al. (2017) reports 60% lower odds of arrest in the six months subsequent to evaluation entry, and 58% lower odds of arrest and 39% lower odds of felony charge over the longer term.

The Brief implies the above is incorrect by noting that in the “Unadjusted Figure: The post-entry LEAD and control groups showed an arrest-rate difference of 22 percentage points”

INCORRECT INTERPRETATION

The 22-point post-entry difference is shown in our own published figure. It appears in the text of the March 2015 report and in the discussion of Collins et al. (2017), alongside an effect size ($d = -0.33$) and a comparison to the roughly 14-point difference reported by WSIPP for existing evidence-based approaches. But the raw difference does not negate the odds ratios; the two describe different aspects of the same finding. It is worth noting, however, that the 22-point difference in this case is conservative. It reflects only the post-entry between-groups difference and does not account for where each group started. Accounting for baseline is a hallmark of controlled designs, and that is what the model estimates do.

“The LEAD group appears to benefit from reduction in warrant arrests due to a higher rate of dropped and reduced charges, while the control group does not.”

NAMING MECHANISM, NOT A CONFOUND

Our March 2015 report presents the arrest analysis both ways. Excluding warrant arrests, the longer-term estimate indicated 34% lower odds ($p = .09$), which weakened to marginal significance. We report both analyses but retained all arrests as the primary specification in our subsequent publication for a substantive reason. Reductions in warrant arrests are partly a mechanism of the intervention, not a confound of the outcome.

Legal advocacy and case management that help a participant make a court date, or that resolve or drop a pending charge, prevent warrants from issuing in the first place. Reducing pile-up in the justice and legal systems is part of what LEAD is designed to do, acknowledging that we need both systems solutions as well as individual behavioral intervention.

"The LEAD and control groups were also subject to different recruitment practices, different patterns of law enforcement, different timelines, and different geographic areas. These differences were likely favorable to LEAD and unfavorable to the control group."

UNSUPPORTED

Recruitment differed in one respect: 57 LEAD participants entered as social contacts rather than through arrest, while all controls entered through arrest. This LEAD program feature was requested by the Seattle Police Department. Both pathways otherwise used identical criteria applied by the same officers, and the analyses were run both with and without social contacts, with the same pattern of findings. Policing was not a source of difference because the same four squads patrolled control neighborhoods and applied the same criteria to both groups. There were not two timelines. Follow-up duration did differ slightly, and we controlled for it directly. The groups were equivalent on criminal history at baseline. The direction the Brief asserts is stated without support.

Section III — C) Felony rates

"Incorrect Claim: 'LEAD reduced felony charges by 39%'"

NOT IN PUBLICATION

Our published finding is that LEAD participants had 39% lower odds of being charged with a felony over the longer term (Collins et al., 2017). That finding stands. The Brief's 11-percentage-point figure is consistent with our published raw prevalences (20% versus 31% post-entry), but again, a raw difference does not account for baseline.

"The KCPAO was therefore able to 'take into account LEAD participation and progress in the program when deciding whether and when to file felony charges.' Because these differences cannot be cleanly separated from any LEAD treatment effect, the observed differences cannot be treated as reliable evidence of a causal program effect."

OUR OWN DISCLOSURE

Two things are true at once: Coordinated prosecution is a designed component of LEAD, not an accident: legal assistance includes coordinating charging strategy, and blinding the prosecutor would mean removing the intervention. But it's also true that, because the prosecutor's discretion operates directly on the outcome being measured, felony charging is a less clean indicator of behavior change than arrest. We noted this in section 4.2 of Collins et al (2017). That said, the sweep of the data show positive effects beyond charging decisions.

Section III — D) Employment

"Incorrect Claim: 'Odds of being on employment continuum increased 46%'"

FALSE

This quoted statement is a precise and accurate statement of our findings.

Heading: "No statistically significant differences"

FALSE

The employment continuum result was significant at $p = .013$.

"With a 3-person difference, statistical significance cannot be established"

FALSE

Statistical significance is not determined by a difference of three people. The models were fitted to repeated monthly observations across the full sample. The Brief also quotes the continuum claim and then refutes it using the binary employment figures, which are two different and transparently defined measures.

The employment continuum was FALSE "improvised"	Both employment measures were transparently defined and reported, and both codings appear in the Measures sections of the March 2016 report and Clifasefi et al. (2017). Approximately 82% of LEAD participants were homeless at the time of recruitment and just over 7% were employed at baseline. Vocational training is a recognized step toward gainful employment – especially in a sample heavily impacted by unemployment and homelessness -- and documenting movement along that continuum is important. Behavioral health settings employ specialists to assist clients across these key life transitions, from vocational training to employment to retirement.
"Reference Comparison: In 2019, the Bureau of Labor Statistics (BLS) tracked employment rates for men after release from 6 months or more of incarceration. By week one, 33.9% of men were employed. By week ten, 50.3% of men were employed. LEAD's cohort is very different and plausibly includes a far a higher rate of untreated behavioral health disorders, but the magnitude difference with incarceration is striking."	NOT COMPARABLE Apples-and-oranges comparison. These are different populations, different constructs, and the BLS data were pulled five years after LEAD evaluation data collection closed. The brief calls the populations "very different" and inexplicably compares them anyway.
"With a cohort death rate three times higher than its estimated employment rate, the relative absence of a LEAD treatment effect for employment suggests insufficient depth or inadequate design for positive behavioral change outcomes." "This prototype does not demonstrate clear validity, but nevertheless appears to count an additional 3 LEAD participants who engage in activity the authors interpret as being adjacent to employment"	INCORRECT INFERENCE A cumulative mortality count across the parent evaluation cannot be set against a monthly-average employment prevalence in a smaller LEAD-group subsample over an 18 month period. This inference draws on different parts of the sample, different windows, and different outcomes with differing base rates. The employment numbers are also not a count of people. The Brief's three "job gainers" appear to be derived by applying the difference between 9.03% and 7.43% to the 176 participants. But those are monthly-average prevalences, identified as such in the table note, and the resulting figure is an average expressed in persons rather than a count of anyone. Returning to the data: Of the 162 participants not employed at baseline, 11 were employed at least one month during follow-up, bringing the total ever employed during our data window to 23. Of the 160 not on the employment continuum, 15 joined it, bringing that total to 30. Further, it is important to consider the elevated baseline mortality risk of a population experiencing homelessness, frequent incarceration, environmental stress, and medical, psychiatric and substance use disorders. We reported these deaths, and we support the Brief's recommendation that mortality be measured systematically across community-based programming serving this population.

Section III — E) Housing

"Incorrect Claim: 'Participants' odds of being housed increased 89%'"	FALSE	The statement as quoted is correct based on the odds ratios.
"13% success rate in securing permanent housing"	FALSE	The Brief does not show its derivation. Our published Table 1 reports monthly-average prevalences, identified as such in the table note. A monthly average is not a count of people. Asked as a person-level question, we can calculate that, of the 145 participants who were documented as lacking permanent housing at referral, 45 obtained it during the follow-up, which is 31% not 13%.
"35% success rate" for shelter or housing among those unhoused at baseline	FALSE	Same construction as above, and same resulting error. Of the 145 participants who were documented as unhoused at baseline, 81% had shelter or permanent housing at some point during follow-up, and 71% percent of people unsheltered at baseline were sheltered at the follow-up.
HUD single-night county sheltered rate as a comparison	NOT COMPARABLE	A point-in-time count cannot serve as a counterfactual for within-person housing transitions in a cohort over 18 months.
Section III — F) Cost savings		
"In relationship to jail bookings, days in jail, and odds of prison, it is difficult to assign these benefits to the LEAD intervention. Criminal justice exposure between the LEAD and control groups was non-comparable due to the study's limitations."	UNSUPPORTED	No limitation is identified and no mechanism is offered for how exposure differed. Baseline differences were addressed through propensity score weighting with published balance checks, and each outcome was tested as change from each group's own baseline. All were significant at $p < .001$.
Differences "mostly driven by large increases in the control group"	FALSE	The control group's movement is larger. That does not negate the finding: controlled designs measure effect against the counterfactual, not against zero, and where system involvement is rising, preventing the rise is the benefit. Regardless of that, the LEAD group still declined in absolute terms and statistically significantly. Costs fell \$2,100 per participant per year, which represents a significant decline in the LEAD group ($p = .003$). The finding holds under either standard.
"Informal and improvised measurements"	FALSE	Every unit cost and its derivation is published, with sources named: public-defense FTE fractions against a stated attorney cost, prosecution at parity, jail at City of Seattle contract rates, prison at average daily bed cost.
Cost modeling "depends on utilization of sunk costs"	FALSE	Average- versus marginal-cost accounting is simply a difference in convention.
"LEAD group's comparatively modest changes fall short of demonstrating a beneficial LEAD effect."	INCORRECT INFERENCE	Estimated criminal justice and legal system costs fell by approximately \$2,100 per LEAD participant per year while rising by approximately \$5,961 per control participant. The between-group difference was significant and sizeable at $-\$8,950$ ($B = -8.95$, robust $SE = 1.40$, $F(1, 314) = 40.83$, $p < .001$, $\beta = -.39$).
"The study does not demonstrate any cost savings to government"		Because the author is implying that the rise in the control group's costs drives the group difference, we conducted a test of the LEAD group's own decline. The LEAD group shows a

statistically significant decline of -\$2100 per year in its own right ($p = .003$).

Further, these are estimates and are labeled as such. What underlies them is not an estimate: LEAD participants had fewer jail bookings, fewer jail days, lower odds of prison incarceration, and fewer felony cases, each at $p < .001$, drawn from jail, prison, prosecution and public-defense administrative records. A jurisdiction that disagrees with our unit costs can substitute its own -- every input is published -- but the reduction in system use is a matter of record.

Section IV — “Origin and Nature”

“The 2015–2026 error pattern traces to the first press releases from the Seattle Police Department and the University of Washington.”

NOT DISPUTED

The driving-speed analogy

Although the driving-speed analogy is not directly relevant to the LEAD evaluation, it illustrates the terminology point already conceded. We did use the word "likelihood" instead of "odds" in the April 2015 summary and in Aim 3 documents. However, that same summary presented the raw arrest prevalences — 83% to 58% for LEAD participants and 77% to 80% for controls — in a figure directly beneath the statement in question and identified the design as a nonrandomized controlled design with propensity score weighting. This one-word change does not bear on whether the statistical tests are correct, which they are.

Section V — “Consequences and Reach”

Federal audit of one Seattle-area LEAD grant in 2024 identifying a “significant deficiency in internal controls over financial reporting” and a “material weakness in internal control over compliance with procurement procedures”

NOT THE EVALUATION

This concerns a grantee organization’s internal financial controls in 2024. It has no bearing on the data, methods or findings of an evaluation that closed data collection on 7/31/2014 and was funded by a private foundation.

"For grant-making entities, the difference was likely material to funding decisions in competitive RFPs where higher performance claims can correspond to a higher win rate."

UNSUPPORTED

No instance is identified. This evaluation was funded by a grant from the Laura and John Arnold Foundation, disclosed in every publication. The evaluation team did not apply for, receive or administer LEAD program funding. We published null and unfavorable findings alongside significant ones.

“Between 2015-2026, implementers of LEAD in Washington State do not appear to report individual outcomes to public funders.”

NOT THE EVALUATION

This concerns ongoing program reporting by implementing agencies, not the evaluation. That said, programs of this kind benefit from sustained infrastructure and support for ongoing outcome reporting, and we said so in our published recommendations for future study.

Section VI — “Policy Recommendations”

Report mortality, loss to follow-up, employment, income and housing outcomes

AGREED

Our published conclusions call for future studies to assess additional participant outcomes. Mortality in this population is high and warrants systematic measurement. We encourage municipalities to ensure support for ongoing evaluations.

Pause expansion; restore certainty of arrest, prosecution and jailing

INCORRECT INFERENCE

This is a policy position that does not follow from anything the authors try to assert in the rest of the Brief. And it is not

warranted based on the findings – which remain unchanged and positive—from the first LEAD program evaluation.

The title

“Retraction”

UNFOUNDED

Notably, considering the title, the Brief’s own recommendations do not ask that LEAD evaluation publications be retracted. Retraction addresses findings rendered unreliable by misconduct or major error; the Brief does not identify an error of that kind in the published articles, and the one wording error it does touch on is being corrected formally in places it still exists. The LEAD findings stand as stated above.
