Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

Identifying the Cumulative Causal Effect of a Non-Binary Treatment from a Binary Instrument*

Vedant Vohra Jacob Goldin

September 3, 2024

Abstract

The effect of a treatment may depend on the intensity with which it is administered. We study identification of ordered treatment effects with a binary instrument, focusing on the effect of moving from the treatment's minimum to maximum intensity. With arbitrary heterogeneity across units, standard IV assumptions (Angrist and Imbens, 1995) do not constrain this parameter, even among compliers. We consider a range of additional assumptions and show how they can deliver sharp, informative bounds. We illustrate our approach with two applications, involving the effect of (1) health insurance on emergency department usage, and (2) attendance in an after-school program on student learning. (*JEL* C01, C21, C26)

*Vohra: University of California San Diego. Goldin: University of Chicago and NBER. We are grateful to Guido Imbens and Kaspar Wuthrich for helpful comments.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

1 Introduction

In many settings, the causal effect of a treatment on an outcome depends on the intensity with which the treatment is administered. For example, the effect of a drug depends on its dosage. The effect of education on wages depends on the number of years of completed schooling. When a binary instrument is used to estimate the effect of a non-binary treatment with variable intensity, the two-stage least-squares estimator identifies the Average Causal Response (ACR), a weighted average of causal effects of a unit change in treatment intensity, where the weights depend on the fraction of compliers induced to cross the various treatment intensity levels (Angrist and Imbens, 1995).

In this paper, we study the causal effect of a change in the intensity of a non-binary ordered treatment when the researcher has access to a valid binary instrument. We focus primarily on a parameter we refer to as the cumulative complier effect (CCE), which captures the average effect of moving from a treatment's minimum intensity to its maximum intensity among those who are induced by the instrument to move to a higher treatment intensity. Frequently, this parameter is an important one for policymakers to understand, such as when a new program or policy is being rolled out to a previously unexposed group, or when a researcher investigates whether a treatment exhibits diminishing marginal returns in its intensity. In such cases, the CCE reveals the effect of providing a "full dose" of the program to individuals who would otherwise not receive it, rather than—as with the ACR—a weighted average of dose-specific causal effects based on the particulars of the intervention being evaluated.

The main challenge in identifying the CCE is understanding the dose-response relationship between the treatment and the outcome – i.e., the causal effect of an additional unit of treatment intensity at each level of treatment (referred to as unit causal effects). Under the standard instrumental variable (IV) assumptions of relevance, independence, and monotonicity, the two-stage least-squares estimator identifies only a specific weighted average of these unit causal effects (i.e., the ACR). We highlight how extrapolating from the ACR to the CCE is subject to two forms of potential biases. First, the ACR over-weights the unit causal effects among compliers whose treatment

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

intensity is more greatly affected by the instrument. Second, the ACR over-weights the unit causal effect for ranges of the dose response function through which compliers are more likely to be induced by the instrument. Thus, depending on the specific instrument being analyzed, the ACR may yield a very different picture of a treatment's effects than the CCE. In fact, under the standard set of IV assumptions, we show that the CCE is entirely unconstrained, in that the data do not permit the researcher to rule out any possible value for the CCE.

In this paper, we consider identification of the CCE under additional identifying assumptions. The key assumption we consider requires a researcher to abstract from heterogeneity in the unit causal effects within the population of compliers. This assumption is restrictive but may be a plausible approximation in research settings of interest, as we illustrate through two applications. We show how a researcher may use this assumption, along with additional structure motivated by the setting at hand, to partially identify the CCE as the solution to a constrained linear optimization problem. For example, a researcher may impose sign restrictions on the unit causal effects or on the concavity of the dose-response function, as in Goldin, Lurie and McCubbin (2021), or on the margin through which an instrument affects participation in a treatment, as in Rose and Shem-Tov (2021). Here, we develop the conditions under which this approach identifies bounds on the CCE. In some cases, the solution to the constrained optimization problem takes the form of sharp analytic bounds, which we characterize below.

We apply our approach to study two randomized evaluations, involving the free provision of health insurance and an after-school instruction program, respectively. For policymakers considering whether to expand these pilot programs, the CCE is a particularly relevant parameter; it describes the effect of fully providing the program to individuals who would not otherwise be able to participate. With respect to the health insurance expansion we study, for example, we focus on the effect of providing a full year of health insurance coverage relative to providing no coverage. As we discuss in more detail below, this parameter could differ from the ACR in practice because the experimental intervention yielded a heterogeneous first-stage effect on months of insurance taken up, and the relationship between months of insurance and medical outcomes may be non-linear.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

Moreover, some treatments are designed to be taken up for their entire duration, as with the after-school instruction program that we study. In such settings, the CCE is a natural parameter of interest because it captures the "all-or-nothing" manner in which the treatment is intended to be provided.

Although our primary focus is on the CCE, we also consider treatment intensity changes within the interior of the dose-response function. We find that the assumptions that permit partial identification of the CCE may also yield informative bounds for these related parameters. We provide details of this analysis in the Online Appendix.

In contrast to the large literature studying identification of the effects of binary treatments with binary instruments, there has been less work studying the use of binary instruments to identify the effects of non-binary treatments. Angrist and Imbens (1995) provides conditions under which two stage least squares identifies the ACR, but as discussed above and in Heckman, Urzua and Vytlacil (2006), the ACR is tied to a specific instrument rather than the treatment, and therefore may not be well-suited to assessing alternative policy interventions. A related literature focuses on extrapolating parameters beyond the LATE from IV research designs, but the methods studied in this literature typically require additional exogenous variation in the form of non-binary instruments (Heckman and Vytlacil, 2007; Imbens and Newey, 2009) or limit their focus to binary treatments (e.g., Balke and Pearl, 1997; Mogstad and Torgovitsky, 2018; Mogstad, Santos and Torgovitsky, 2018). One exception is Kamat, Norris and Pecenco (2023), which considers partial identification of a range of treatment effect parameters in the presence of multiple treatments and a discrete-valued instrument; identification is facilitated by restrictions on selection into treatment and unobserved heterogeneity. A second exception is Torgovitsky (2015), which point-identifies the dose response function for a continuous treatment under the assumption of rank invariance for both the first-stage and outcome equations. Finally, a recent working paper by Chernozhukov et al. (2024) studies treatment effect heterogeneity in an IV framework that accommodates non-binary treatments by imposing restrictions on the relationship between the treatment assignment and potential outcomes. In contrast to these methods, identification under our approach relies on restricting heterogeneity across compliant subgroups in conjunction with economically motivated restrictions on the shape

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

of the treatment's dose-response function.[1]

In addition, our work adds to a large literature that studies partial identification of treatment effect models; recent examples include Fan, Sherman and Shum (2014), Arnold, Dobbie and Hull (2022), Rambachan and Roth (2023), and Tebaldi, Torgovitsky and Yang (2023). With respect to IV models in particular, many of these papers study what can be learned when the standard IV assumptions do not hold (e.g., Manski and Pepper, 2000; De Chaisemartin, 2017); see Swanson et al. (2018) for a review. However, as we show below, even when the standard IV assumptions do hold, the data do not constrain the CCE when there is a single binary instrument and a non-binary treatment. Hence, our approach is to study how imposing additional structure beyond Angrist and Imbens (1995) allows a researcher to partially identify the CCE.

Finally, a common practice by researchers in settings with non-binary treatments is to "binarize" the treatment by collapsing it into two categories; two recent papers study the assumptions underlying the validity of this approach (Andresen and Huber, 2021; Rose and Shem-Tov, 2024). In contrast, the parameter we study is based on the dose-response relationship for the original (uncollapsed) non-binary treatment.

2 Setting and Notation

Consider a population, indexed by i, in which individuals are assigned a binary instrument, $Z_i \in \{0, 1\}$, and one level of a discrete treatment, ranging in intensity from 0 to J. Let $D_i(Z) \in \{0, 1, 2, ..., J\}$ denote i's treatment level under each value of the instrument, and let $Y_i(j)$ denote the outcome of interest that would be obtained if i were to receive treatment level j.

We assume that the following conditions are satisfied.

Assumption 1: Relevance

$$E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0] \neq 0$$

Assumption 2: Independence

[1] Lochner and Moretti (2015) impose a similar restriction on treatment effect heterogeneity to motivate a proposed test for exogeneity of a multi-valued treatment using a binary instrument.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

$$\{Y_i(0), Y_i(1), ..., Y_i(J), D_i(0), D_i(1)\} \perp\!\!\!\perp Z_i$$

Assumption 3: Monotonicity

$$\mathbb{P}\left(D_i(1) \geq D_i(0)\right) = 1$$

Angrist and Imbens (1995) show that under these assumptions, the standard two-stage least-squares estimator identifies the average causal response (ACR) of D_i on Y_i, i.e. a weighted average of the causal effect from a unit change in the treatment on the outcome, where the weights are the share of corresponding unit changes in the treatment intensity induced by the instrument. More precisely, the ACR corresponds to the right hand side of the following equation:

$$\frac{E[Y \mid Z = 1] - E[Y \mid Z = 0]}{E[D \mid Z = 1] - E[D \mid Z = 0]} = \sum_{j=1}^{J} w_j E[Y_i(j) - Y_i(j-1) \mid D_i(1) \geq j > D_i(0)] \quad (1)$$

where

$$w_j = \frac{P[D_i(1) \geq j > D_i(0)]}{\sum_{j'=1}^{J} P[D_i(1) \geq j' > D_i(0)]}.$$

Our goal is to shed light on the average cumulative effect of a treatment. We define the cumulative effect of a treatment on an outcome Y as the causal effect of a shift from $D = 0$ to $D = J$, or $Y_i(J) - Y_i(0)$. Our parameter of interest, the cumulative complier effect (CCE), is defined as the mean cumulative effect for the population of compliers:

$$CCE = E\left[Y_i(J) - Y_i(0) \mid D_i(1) > D_i(0)\right] \quad (2)$$

In the next section, we highlight challenges to identifying the CCE under Assumptions 1-3. However, these assumptions do permit us to identify the weights corresponding to each unit-response in Equation 1, $\{w_j\}_{j=1}^{J}$. In Section 4, we propose using these estimated weights, along with the estimated ACR and setting-specific assumptions about the unit causal effects, to bound the CCE.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

3 Challenges to Identifying the CCE

Although both the ACR and CCE are functions of a treatment's unit causal effects, the former cannot be directly extrapolated to identify the latter. Whereas the ACR summarizes the average causal effect associated with a specific intervention, the CCE summarizes the average causal effect of the treatment across *all* compliers and *all* treatment intensities. There are therefore two potential forms of selection that may cause the ACR to diverge from the CCE. First, those individuals who are induced by the intervention to increase their treatment intensity may also have higher per-unit treatment effects. Second, the additional levels of the treatment induced by the intervention may have higher per-unit treatment effects than other levels of the treatment (that were not induced by the intervention).[2]

Given the potential wedge between the ACR and the CCE, a natural question to ask is what can be learned about the CCE under the assumptions that identify the ACR? Unfortunately, the answer is "not much." In fact, the CCE is entirely unconstrained under Assumptions 1-3, as we show formally in Appendix C. Intuitively, since not all compliers move along the full length of the dose-response function from treatment intensity 0 to J, but still contribute to the CCE, we need to impose additional structure on the unit-causal effects to extrapolate *across* complier types.[3]

4 Identifying the Cumulative Complier Effect Under Complier Effect Homogeneity

In this section, we abstract from potential heterogeneity in the unit causal effects among compliers to facilitate identification of the CCE. We consider the following assumption:

Assumption 4: Homogeneous Incremental Effect Across Compliant Subgroups

$E[Y_i(j) - Y_i(j-1) \mid D_i(1) \geq k > D_i(0)] = \beta_j$ for all $k = 1, ..., J$

[2] In Appendix B, we formally relate the ACR and CCE in terms of these biases.
[3] When the unit causal effects are known to be bounded, the CCE is bounded as well; see Appendix C for details.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

This assumption is restrictive, but as we illustrate in Section 5, it can be plausible in real-world applications of interest (potentially after conditioning on observable characteristics along the lines of Angrist and Fernandez-Val (2010)). The value of the assumption is that it focuses attention on the uncertainty in the identification of the CCE that arises due to the unknown shape of the treatment's dose-response function. When Assumption 4 holds, the CCE corresponds to the effect of moving from minimum treatment intensity to maximum treatment intensity for any compliant subgroup. Lochner and Moretti (2015) impose a similar assumption to test exogeneity of a multi-valued treatment.

Under Assumption 4, we can express the ACR and CCE as

$$ACR = \sum_{j=1}^{J} w_j \beta_j$$

and

$$CCE = \sum_{j=1}^{J} \beta_j$$

To shed light on the range of cumulative effects of the treatment consistent with the data, we find bounds on the CCE by casting it as a linear optimization problem. Under Assumptions 1-4, the optimization problem can be written as:

$$\underset{\{\beta_j\}}{\text{Maximize/Minimize}} \quad \sum_{j=1}^{J} \beta_j \qquad \textbf{LP.1}$$

$$\text{subject to} \quad \sum_{j=1}^{J} \beta_j w_j = ACR \qquad \text{(C.1)}$$

Note that the ACR and set of weights, $\{w_j\}_{j=1}^{J}$, in **LP.1** are identified and estimable under Assumptions 1-3 (Angrist and Imbens, 1995). The ACR can be estimated using the sample moments corresponding to the expectations on the left-hand side of Equation (1). The weights can be estimated by comparing the share of units at each treatment level under $Z = 0$ versus $Z = 1$.

Denote the maximum feasible value of this objective function, $\sum_{j=1}^{J} \beta_j$, by $\overline{CCE}$ and the min-

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

imum by $\underline{CCE}$. Since β_j can be arbitrarily large or small, the CCE is generally unbounded under Assumption 1-4. More formally, for any $a \in \mathbb{R}$, non-uniform set of weights $\{w\}$, and observed ACR value, there exists a feasible solution vector $(\beta_1, \beta_2, ..., \beta_J)$ such that $CCE = a$. Hence, additional assumptions regarding the unit causal effects are needed to obtain meaningful bounds on the CCE.[4]

We now consider a range of additional assumptions that may be appropriate to impose depending on the application, in the spirit of Manski (2003). In some settings, the direction of the treatment effect will be known from theory or prior research. In such cases, without loss of generality, we can impose that the direction of each unit effects is positive:

Assumption 5: Uniform Sign of Unit Causal Effects

$$\beta_j \geq 0 \; \forall \; j$$

Under Assumptions 1-5, the CCE can be bounded based on the ACR and the empirically observable weights.

Proposition 1: Under Assumptions 1-5, the following sharp bounds hold:

$$CCE \in \left[\frac{ACR}{w_{\overline{j}}}, \frac{ACR}{w_{\underline{j}}} \right]$$

where $\underline{j} = \arg\min_{j \in \{1,...J\}} \{w_j\}$ and $\overline{j} = \arg\max_{j \in \{1,...J\}} \{w_j\}$.[5]

The proof of Proposition 1, and all subsequent results, is contained in the Online Appendix. The bounds provided in the Proposition are sharp, in the sense that all values of the CCE in the identified set are compatible with Assumptions 1-5 and the data (see Appendix G for details).

In some settings, it will be reasonable to impose additional assumptions beyond restrictions on

[4] In settings where additional exogenous variation is available in the form of multiple instruments, Angrist and Imbens (1995) show that each instrument can be used to identify an ACR and a set of compliance weights. When the number of instruments is equal to the number of treatment levels, the CCE may be point-identified. More generally, when the number of treatment levels exceeds the number of instruments, the researcher can incorporate the variation from the additional instruments to tighten the bounds on the CCE by adding them as constraints to LP.1. (see Appendix D for details).

[5] Note that if one of the weights is zero, then $w_{\underline{j}} = 0$; in this case, $CCE \in \left[\frac{ACR}{w_{\overline{j}}}, \infty \right)$.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

the sign of the treatment effect. For example, the researcher may have good reason to believe that the magnitude of the unit causal effects is non-increasing in the intensity of the treatment.[6] We refer to this assumption as concavity of the dose-response function:

Assumption 6: Concavity

$$\beta_j \geq \beta_{j+1} \; \forall \; j = 1, ..., J-1$$

Under Assumptions 1-6, bounds on the CCE can be obtained from the linear optimization problem:

$$\underset{\{\beta_j\}}{\text{Maximize/Minimize}} \quad \sum_{j=1}^{J} \beta_j \qquad \textbf{LP.2}$$

$$\text{subject to} \quad \sum_{j=1}^{J} \beta_j w_j = ACR \qquad \text{(C.1)}$$

$$\beta_j \geq 0 \qquad \forall \; j = 1, ..., J \qquad \text{(C.2)}$$

$$\beta_j \geq \beta_{j+1} \qquad \forall \; j = 1, ..., J-1 \qquad \text{(C.3)}$$

As described above, the ACR and weights in **LP.2** are identified and estimable under Assumptions 1-3. Goldin, Lurie and McCubbin (2021) solved this linear problem to estimate bounds on the cumulative effect of health insurance coverage in their setting.

Finally, in some settings the share of compliers that are induced by an instrument to cross a particular treatment intensity threshold will be non-decreasing in the threshold level:

Assumption 7: Monotonic Complier-Share Weights

$w_j \geq w_{j+1} \; \forall \; j = 1, ..., J-1$, and the inequality is strict for at least one such j.

Unlike Assumptions 5 and 6, Assumption 7 is empirically verifiable because the complier-share weights are identified under our maintained Assumptions 1-3. A sufficient condition for Assump-

[6] A different potential restriction that a researcher might impose is that the magnitude of each unit causal effect is bounded between known values. We extend Proposition 1 to this setting in Appendix H.1. A related restriction is that the outcome of interest has bounded support. In addition to implying bounds on the unit causal effects, this assumption constrains the values the unobserved potential outcomes can take on. We discuss this setting in Appendix H.2.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

tion 7 to hold is that all individuals who increase treatment intensity in response to the instrument do so on the extensive margin (Rose and Shem-Tov, 2024). In settings where these conditions hold, the bounds that emerge from LP.2 take the following simple analytic form:

Proposition 2: Under Assumptions 1-7, the following sharp bounds hold:

$$CCE \in \left[\frac{ACR}{w_1}\,,\; ACR \times J\right]$$

When Proposition 2 applies, the CCE is maximized when the dose-response function is linear and minimized when intensive margin changes in the intensity of the treatment have no effect on the outcome. The bounds provided in Proposition 2 are sharp, in the sense that all values of the CCE in the identified set are compatible with Assumptions 1-7 and the data (see Appendix G for details).

Finally, in some settings the researcher may be interested in parameters relating to the dose-response function other than the CCE, such as when a potential intervention is intended to shift behavior across a subset of treatment values. In the appendix, we study what may be identified for a broader class of parameters, relating to the effect of moving from treatment intensity j_1 to j_2 where $j_2 > j_1$ and $j_1, j_2 \in \{0, 1, ..., J\}$. We show that the sharp analytic bounds provided in Propositions 1 and 2 extend naturally to this setting; see Online Appendix I for details.

5 Applications

We illustrate the method by applying it in two empirical settings: the Oregon Health Insurance Experiment (Taubman et al., 2014) and a randomized intervention studying the effect of a computer-aided learning program for middle-school students on test scores in India (Muralidharan, Singh and Ganimian, 2019). Table 1 shows the ACR and the bounds for the CCE for both settings. Following Andrews (2000), we provide confidence intervals on the CCE bounds using a modified bootstrapping procedure that provides accurate coverage in linear optimization settings like the one we study.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

5.1 Health Insurance and Emergency Department Usage

In 2008, certain low-income adults in Oregon were selected through a lottery to enroll in Medicaid. Taubman et al. (2014) use this random variation to study the effect of Medicaid on Emergency Department (ED) usage. To shed light on the cumulative effect of obtaining a full period of Medicaid coverage, we define the treatment intensity as the number of months an individual was enrolled in Medicaid during the 19-month study period. The data imply an ACR of Medicaid on ED use of 0.53 percentage points. While this captures the average effect of the additional months of Medicaid coverage induced by the particular intervention being studied, hospitals and policymakers may also be interested in understanding the effect of providing annual coverage to previously uninsured individuals.

It is likely that there is heterogeneity in the effect of Medicaid on ED usage across compliers. However, Kowalski (2021) shows that previous ER utilization can explain the vast majority of this heterogeneity; the marginal treatment effect curve is (approximately) flat after conditioning on prior ED use. Motivated by this finding, we apply Assumption 4 conditional on prior ED use and bound the CCE separately for those with and without a history of ED use before the Medicaid lottery. We assume that the effect of Medicaid on ED use is non-negative (Assumption 5), and that the per-month effect is non-increasing in the number of months of enrollment (Assumption 6). However, it is likely that the Medicaid lottery increased months of enrollment for some individuals who would have enrolled in coverage even absent the treatment, suggesting that Assumption 7 may not hold in this context.[7] Hence, we obtain bounds on the CCE by solving the linear program in **LP.2**.

In this context, the CCE captures the causal effect on ED usage of enrolling in Medicaid for the full 19-month study period. Our results suggest this effect was an increase in the share of individuals using the ED at least once of between 6.7 and 9.8 percentage points for those *without* a prior history of ED use, and between 7.4 and 10.6 percentage points for those *with* a prior history of ED use. We can combine these into a CCE for the overall sample by taking a weighted average of the two, where the weights correspond to the observed distribution of prior ED use. The implied CCE using this

[7] That being said, Appendix Figure 1(a) and 1(b) suggests only slight deviations from monotonicity.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

approach is between 6.9 and 10.1 percentage points. Similarly, we can also derive bounds on the effect associated with providing uncovered individuals with a full year of coverage by replacing the objective function with $\sum_{j=1}^{12} \beta_j$.[8] Doing so implies an effect of a full year of Medicaid coverage of between 6.3 and 7.6 percentage points.[9]

5.2 After-School Instruction and Educational Outcomes

Muralidharan, Singh and Ganimian (2019) study the effect of a technology-aided after-school instruction program on test scores in urban India, using a lottery that provided winners with free access to the program. Using the outcome of the lottery to instrument for weeks of program attendance, the data from the paper imply an ACR for the program of a 0.045 standard deviation increase in math test scores and a 0.030 standard deviation increase for Hindi (see Table 1).[10] While the ACR identifies a specific weighted average effect of attending the Mindspark centers for a week on test scores, a policymaker considering whether to scale up the program might be particularly interested in the overall effect of providing a full course of after-school instruction to a new set of students. Indeed, since the program was designed to be taken up for its entire duration, the CCE is a natural parameter of interest.

The authors suggest that it is likely that the effects of the program are homogeneous across students using three pieces of evidence. First, the observed treatment effects are similar across the distribution of student achievement prior to the program – a likely source of heterogeneity. Second, the authors cannot reject equality of the IV and OLS value-added estimates, suggesting the ATE and LATE might be similar and heterogeneity in the compliant subgroups might be small. Finally, the authors note that the outcomes for the control group and never-takers are similar, suggesting equal-

[8] Note that under Assumption 4, this parameter reflects the effect of a full year of Medicaid coverage for those compliers who would have enrolled in partial-year coverage absent the intervention ($D_i(0) > 0$), as well as for those compliers who would not have otherwise enrolled in any coverage ($D_i(0) = 0$).

[9] In general, the width of the bounds on $\sum_{j=1}^{12} \beta_j$ is a function of the particular range of the dose-response function of interest (in this case, 0 to 12 months of coverage) and the distribution of treatment intensity changes induced by the instrument; see Online Appendix I for details.

[10] In their original analysis, the authors focused on days of attendance rather than weeks. Here, we pool the treatment intensity to the week-level to reduce statistical variability.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

ity of potential outcomes across different compliance groups. Taken together, these observations provide suggestive evidence in support of Assumption 4.

In addition, Assumptions 5 and 6 both seem likely to hold in this context: one would not expect the program to *reduce* learning and it seems plausible (though not guaranteed) that there would be non-increasing returns to scale to program attendance. Because only students who won the lottery could attend the program, we know that all compliers were affected on the extensive margin. Hence, as discussed in Section 4, the weights are guaranteed to be monotonically declining so that Assumption 7 holds. Appendix Figure 1(c) empirically verifies that this is the case.

When these conditions hold, Proposition 2 provides sharp bounds for the CCE of weeks of attendance. In particular, translating the ACR point estimate into bounds for the CCE implies that the full program course (12 weeks) would increase math scores by between 0.40 and 0.54 standard deviations and Hindi scores by between 0.27 and 0.36 standard deviations (Table 1, Column 2). Like the main IV estimate, these bounds grow substantially less precise, but continue to exclude zero, once statistical uncertainty is taken into account (Table 1, Column 3).

6 Conclusion

Researchers are often interested in evaluating the effect of non-binary treatments. In such settings, the cumulative effect of the treatment, i.e., the effect of moving from minimum to maximum treatment intensity, is a natural parameter of interest. In this paper, we have highlighted conditions that allow for partial identification of this parameter among compliers and show that, in their absence, meaningful identification is generally not feasible. While the required assumptions are strong, we illustrate with two applications that they are plausible in settings of interest to applied researchers.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

References

Andresen, Martin E, and Martin Huber. 2021. "Instrument-based estimation with binarised treatments: issues and tests for the exclusion restriction." *The Econometrics Journal*, 24(3): 536–558.

Andrews, Donald W. K. 2000. "Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space." *Econometrica*, 68(2): 399–405.

Angrist, Joshua, and Ivan Fernandez-Val. 2010. "Extrapolate-ing: External validity and overidentification in the late framework." National Bureau of Economic Research.

Angrist, Joshua D., and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association*, 90(430): 431–442.

Arnold, David, Will Dobbie, and Peter Hull. 2022. "Measuring Racial Discrimination in Bail Decisions." *American Economic Review*, 112(9).

Balke, Alexander, and Judea Pearl. 1997. "Bounds on treatment effects from studies with imperfect compliance." *Journal of the American statistical Association*, 92(439): 1171–1176.

Bickel, Peter J., and Anat Sakov. 2008. "On the choice of m in the m-out-of-n bootstrap and confidence bounds for extrema." *Statistica Sinica*, 18(3): 967–985.

Chernozhukov, Victor, Iván Fernández-Val, Sukjin Han, and Kaspar Wüthrich. 2024. "Estimating Causal Effects of Discrete and Continuous Treatments with Binary Instruments." *arXiv preprint arXiv:2403.05850*.

De Chaisemartin, Clement. 2017. "Tolerating defiance? Local average treatment effects without monotonicity." *Quantitative Economics*, 8(2): 367–396.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

Demuynck, Thomas. 2015. "Bounding average treatment effects: A linear programming approach." *Economics Letters*, 137: 75 – 77.

Fan, Yanqin, Robert Sherman, and Matthew Shum. 2014. "Identifying treatment effects under data combination." *Econometrica*, 82(2): 811–822.

Goldin, Jacob, Ithai Z Lurie, and Janet McCubbin. 2021. "Health insurance and mortality: Experimental evidence from taxpayer outreach." *The Quarterly Journal of Economics*, 136(1): 1–49.

Heckman, James J, and Edward J Vytlacil. 2007. "Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments." *Handbook of econometrics*, 6: 4875–5143.

Heckman, James J, Sergio Urzua, and Edward Vytlacil. 2006. "Understanding instrumental variables in models with essential heterogeneity." *The review of economics and statistics*, 88(3): 389–432.

Huang, Emily J., Ethan X. Fang, Daniel F. Hanley, and Michael Rosenblum. 2016. "Inequality in treatment benefits: Can we determine if a new treatment benefits the many or the few?" *Biostatistics*, 18(2): 308–324.

Imbens, Guido W, and Whitney K Newey. 2009. "Identification and estimation of triangular simultaneous equations models without additivity." *Econometrica*, 77(5): 1481–1512.

Kamat, Vishal, Samuel Norris, and Matthew Pecenco. 2023. "Identification in Multiple Treatment Models under Discrete Variation." *arXiv preprint arXiv:2307.06174*.

Kowalski, Amanda E. 2021. "Reconciling seemingly contradictory results from the Oregon health insurance experiment and the Massachusetts health reform." *Review of Economics and Statistics*, 1–45.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

Lochner, Lance, and Enrico Moretti. 2015. "Estimating and testing models with many treatment levels and limited instruments." *Review of Economics and Statistics*, 97(2): 387–397.

Manski, Charles F. 2003. *Partial identification of probability distributions.* Vol. 5, Springer.

Manski, Charles F., and John V. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica*, 68(4): 997–1010.

Mogstad, Magne, and Alexander Torgovitsky. 2018. "Identification and extrapolation of causal effects with instrumental variables." *Annual Review of Economics*, 10: 577–613.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky. 2018. "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters." *Econometrica*, 86(5): 1589–1619.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review*, 109(4): 1426–60.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A more credible approach to parallel trends." *Review of Economic Studies*, 90(5): 2555–2591.

Rose, Evan K, and Yotam Shem-Tov. 2021. "How does incarceration affect reoffending? estimating the dose-response function." *Journal of Political Economy*, 129(12): 3302–3356.

Rose, Evan K, and Yotam Shem-Tov. 2024. "On recoding ordered treatments as binary indicators." National Bureau of Economic Research.

Swanson, Sonja A, Miguel A Hernán, Matthew Miller, James M Robins, and Thomas S Richardson. 2018. "Partial identification of the average treatment effect using instrumental variables: review of methods for binary instruments, treatments, and outcomes." *Journal of the American Statistical Association*, 113(522): 933–947.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the
Massachusetts Institute of Technology

Taubman, Sarah L, Heidi L Allen, Bill J Wright, Katherine Baicker, and Amy N Finkelstein. 2014. "Medicaid increases emergency-department use: evidence from Oregon's Health Insurance Experiment." *Science*, 343(6168): 263–268.

Tebaldi, Pietro, Alexander Torgovitsky, and Hanbin Yang. 2023. "Nonparametric estimates of demand in the california health insurance exchange." *Econometrica*, 91(1): 107–146.

Torgovitsky, Alexander. 2015. "Identification of nonseparable models using instruments with small support." *Econometrica*, 83(3): 1185–1197.

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

Table 1: Bounds on the Cumulative Complier Effect (CCE)

	(1)	(2)	(3)
	Average Causal Response	CCE Bounds	CCE Bounds (with 95% CI)
Taubman et al. (2014)			
Emergency Department Usage	0.529 (0.181)	[6.918, 10.053]	[2.160, 18.064]
Muralidharan, Singh and Ganimian (2019)			
Math Test Score	0.045 (0.007)	[0.400, 0.538]	[0.175, 0.722]
Hindi Test Score	0.030 (0.007)	[0.265, 0.357]	[0.099, 0.630]

Notes: The table illustrates the application of the proposed method to estimating the effect of (1) health insurance coverage on emergency department usage and (2) after-school instruction on test scores. The Average Causal Response (Column 1) is obtained from a two-stage least-squares regression using the data reported in the specified study. The point estimates for the CCE bounds (Column 2) are calculated as described in Section 4. The 95% confidence intervals (Column 3) are obtained from an m-out-of-n bootstrapping procedure (Demuynck, 2015; Huang et al., 2016): instead of drawing samples (with replacement) of size n equal to the sample size, we draw samples of size $m \ll n$. To select the appropriate m for this procedure, we follow the method proposed in Bickel and Sakov (2008).

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01526
© 2024 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology

Downloaded from direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01526/2477302/rest_a_01526.pdf by University of Chicago user on 20 October 2026