NBER WORKING PAPER SERIES

# PARTIAL IDENTIFICATION OF TREATMENT-EFFECT DISTRIBUTIONS WITH COUNT-VALUED OUTCOMES

John Mullahy

Working Paper 31005
http://www.nber.org/papers/w31005

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2023

No funding sources or material and relevant financial relationships to declare. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Partial Identification of Treatment-Effect Distributions with Count-Valued Outcomes
John Mullahy
NBER Working Paper No. 31005
March 2023
JEL No. C10,C25,I1


**ABSTRACT**

With count-valued outcomes y in {0,1,...,M} identification and estimation of average treatment effects raise no special considerations beyond those involved in the continuous-outcome case. If partial identification of the distribution of treatment effects is of interest, however, count-valued outcomes present some subtle yet important considerations beyond those involved in continuous-outcome contexts. This paper derives appropriate bounds on the distribution of treatment effects for count-valued outcomes.



John Mullahy
University of Wisconsin-Madison
Dept. of Population Health Sciences
787 WARF, 610 N. Walnut Street
Madison, WI 53726
and NBER
jmullahy@facstaff.wisc.edu

## 1. Introduction and Definitions

With count-valued outcomes $y \in \{0,1,\ldots,M\}$ identification of *average* treatment effects (ATEs) presents no special considerations beyond those involved with continuous outcomes. If features of the *distribution* of treatment effects (TEs) are of interest, however, count-valued outcomes raise subtle yet important considerations beyond those involved with continuous outcomes. Since such outcomes are encountered often in empirical health research an assessment of how their count-valued nature affects (partial) identification of treatment-effect distributions seems warranted.

Fan and Park, 2010, (FP) provide key results for partial identification (PI) of TE distributions with continuous outcomes. This paper shows that a subtlety that can be ignored in the continuous case must be addressed with count-valued outcomes. Specifically, distinguishing strict from weak inequalities is essential for defining best bounds on count-valued outcomes' TE distributions. Key is the simple recognition that $\Pr(Z \le a) = 1 - \Pr(Z \ge a)$ holds for continuous but not for count-valued Z.[1]

The count-valued potential outcomes are $y_0$ and $y_1$ with $y_j \in V = \{0,1,\ldots,M\}$. M is assumed to be finite, an assumption made mainly for computational and expositional reasons since most results go through if V is unbounded. The joint distribution $\Pr(y_0, y_1) = \Pr(y_0 = j, y_1 = k) = p_{jk}$ has support $V \times V$.

The unobservable TE is $\Delta = y_1 - y_0$. It merits noting that the difference that defines TE is a meaningful one if the $y_j$ have true ratio-scale measures but not if the $y_j$ are arbitrarily-labeled ordinal-scale outcomes (see Stevens, 1946), the latter encountered often in empirical health research. The support of the TE distribution $F_\Delta(\delta)$ is $T = \{-M, -(M-1), \ldots, M\}$. $F_\Delta(\delta) = \Pr(\Delta \le \delta)$ is defined as is typical based on a weak inequality with respect to its support points in T. To illustrate, for M=3 figure 1 depicts the sample space, TEs, and joint probability distribution. The sum of the blue dots' $p_{jk}$ ($0 \le k \le j \le M$) equals $\Pr(\Delta \le \delta)$ for $\delta = 0$.

The goal is to identify the TE distribution $\Pr(\Delta \le \delta)$ for all $\delta \in T$. $\Pr(\Delta \le \delta)$ cannot generally be point identified—though $\Pr(\Delta \le M) = 1$ is trivially point identified—but informative PI may be possible. Let D denote the binary treatment that gives rise to $y_0$ and $y_1$. As is standard the observed outcomes y are given by $y = Dy_1 + (1-D)y_0$. Like FP assume that the marginal distributions $\Pr(y_j)$ are point

[1] Machado and Santos Silva, 2005, and Manski, 2007 (page 40), raise related considerations in the context of quantile TEs.

identified from treatments being randomly assigned. PI of $\Pr(\Delta \le \delta)$ then follows from careful derivation of its Boole-Fréchet-Hoeffding (BFH) lower and upper bounds (LB, UB) that are based exclusively on the identified marginal distributions $\Pr(y_j)$.

Sections 2 and 3 present the results for $LB(\Pr(\Delta \le \delta))$ and $UB(\Pr(\Delta \le \delta))$, respectively. Section 4 presents an illustrative empirical example. Section 5 concludes. To streamline the discussion proofs and additional discussion appear in the Appendix.

## 2. Best Lower Bounds on $\Pr(\boldsymbol{\Delta} \le \delta)$

Let the best LB on $\Pr(\Delta \le \delta)$ be denoted $LB^*(\Pr(\Delta \le \delta))$.

Proposition 1:

a. For each $t \in V$ and $\delta \in T$ an identifiable LB on $\Pr(\Delta \le \delta)$ is

$$\max\left\{0, \Pr(y_1 \le t) - \Pr(y_0 + \delta < t)\right\} . \tag{2.a}$$

b. For each $\delta \in T$ the identifiable best LB on $\Pr(\Delta \le \delta)$ is

$$LB^*(\Pr(\Delta \le \delta)) = \max_{t \in V} \max\left\{0, \Pr(y_1 \le t) - \Pr(y_0 + \delta < t)\right\} . \tag{2.b}$$

Proof: See Appendix section A.1.

The key finding of this section arises from comparison of expressions (2.a) and (2.b) (note the strict inequalities in the max{...} arguments' subtrahends) with, respectively,

$$\max\left\{0, \Pr(y_1 \le t) - \Pr(y_0 + \delta \le t)\right\} \tag{2.c}$$

and

$$\max_{t \in V} \max\left\{0, \Pr(y_1 \le t) - \Pr(y_0 + \delta \le t)\right\} . \tag{2.d}$$

(note the weak inequalities in the max{...} arguments' subtrahends). Expression (2.d) is this paper's representation of FP's equation (2):

$$F^{L}(\delta) = \sup_{y}\left(\max\left(F_1(y) - F_0(y-\delta)\right), 0\right). \tag{FP.2}$$

With continuous outcomes the distinctions between (2.a) and (2.c) and between (2.b) and (2.d) vanish—thus rationalizing FP not drawing the distinction in their Lemma 2.1—but with count-valued outcomes they do not. Thus while (2.d) is an identifiable LB on $\Pr(\Delta \le \delta)$ that would arise from applying (FP.2), (2.b) will in general be a better (indeed best) LB.

So long as both (2.a) and (2.c) are in $(0,1)$ then (2.a) exceeds (2.c) by

$$\Pr(y_0 + \delta \le t) - \Pr(y_0 + \delta < t) = \sum_{k=0}^{M} \Pr(y_0 = t - \delta, y_1 = k) \tag{2.e}$$

for each $t \in V$. As such $LB^*(\Pr(\Delta \le \delta))$ from (2.b) must be at least as large as (2.d). Since it corresponds to a fraction $\frac{1}{M+1}$ of the joint sample space the difference (2.e) may be empirically meaningful, especially when M is not large. Note also that the difference between (2.b) and (2.d) may in some data turn out to be a difference between a positive (informative) and a zero (uninformative) best LB.

## 3. Best Upper Bounds on $\Pr(\Delta \le \delta)$

Let the best UB on $\Pr(\Delta \le \delta)$ be denoted $UB^*(\Pr(\Delta \le \delta))$.

Proposition 2:

a. For each $t \in V$ and $\delta \in T$ two identifiable UBs on $\Pr(\Delta \le \delta)$ are

$$1 - \max\left\{0, \Pr(y_0 \le t) - \Pr(y_1 \le t + \delta)\right\} \quad \text{and} \tag{3.a}$$

$$1 - \max\left\{0, \Pr(y_0 < t) - \Pr(y_1 < t + \delta)\right\}. \tag{3.b}$$

b. For each $\delta \in T$ the identifiable best UB on $\Pr(\Delta \le \delta)$ is

$$UB^*(\Pr(\Delta \le \delta)) = 1 - \max_{t \in V} \max\left\{0, \Pr(y_0 \le t) - \Pr(y_1 \le t + \delta)\right\} \tag{3.c}$$

$$= 1 - \max_{t \in V} \max\left\{0, \Pr(y_0 < t) - \Pr(y_1 < t + \delta)\right\}. \tag{3.d}$$

Proof: See Appendix section A.2.

Note that $UB^*\left(\Pr\left(\Delta \le \delta\right)\right) = 1 - LB^*\left(\Pr\left(\Delta > \delta\right)\right)$. One feature of the TE distribution $\Pr\left(\Delta \le \delta\right)$ that may be of interest (Mullahy, 2018) is $\Pr\left(\Delta > 0\right) = 1 - \Pr\left(\Delta \le 0\right)$. It follows that

$$LB^*\left(\Pr\left(\Delta > 0\right)\right) = LB^*\left(1 - \Pr\left(\Delta \le 0\right)\right) = 1 - UB^*\left(\Pr\left(\Delta \le 0\right)\right). \tag{3.e}$$

Attention to which inequalities are strict and which are weak is again critical.

## 4. An Illustrative Example

The 2015 Health Survey of England (HSE) elicits the specific dates in the week preceding a subject's interview on which the subject engaged in some form of vigorous or moderate exercise (see Mullahy, 2022, for details on the HSE). Preliminary inspection of the sample suggests that weekly totals of such exercise days may tend to be greater in spring and summer than in fall and winter. For this example survey interview dates may plausibly be treated as if randomly assigned. To gauge the "effect" of season on weekly exercise frequency the treatment D is defined to be one if the interview took place in the second or third quarter of 2015 (spring and summer, N=3,167) and zero otherwise (winter and fall, N=3,602). The PI task is to compute the best bounds on $\Pr\left(\Delta \le \delta\right)$ over the TE distribution support $T = \left\{-7, -6, \ldots, 7\right\}$. The resulting $UB^*\left(\Pr\left(\Delta \le \delta\right)\right)$ and $LB^*\left(\Pr\left(\Delta \le \delta\right)\right)$ are shown in columns 2 and 3 of table 1 and are depicted in figure 2.[2]

## 5. Conclusions

A subtlety that is ignorable when computing bounds on continuous outcomes' TE distributions is consequential for count-valued outcomes, particularly if M is not large. Obtaining best bounds with count-valued outcomes is straightforward using this paper's results.[3] In empirical work $LB^*\left(\Pr\left(\Delta \le \delta\right)\right)$ and $UB^*\left(\Pr\left(\Delta \le \delta\right)\right)$ can be estimated via standard analogy principle methods.


## Acknowledgements

Thanks are owed to Chris Adams, Chuck Manski, and Dan Millimet for sharing helpful insights.


---

[2] The bounds are computed using an author-written Stata program that is available here.

[3] It may sometimes be possible to obtain tighter bounds by making (unverifiable) assumptions about the joint distribution $\Pr\left(y_0, y_1\right)$. See section A.3 of the Appendix for discussion.

## References


Fan, Y. and S.S. Park. 2010. "Sharp Bounds on the Distribution of Treatment Effects and Their Statistical Inference." *Econometric Theory* 26: 931-951.

Machado, J.A.F. and J.M.C. Santos Silva. 2005. "Quantiles for Counts." *JASA* 100: 1226-1237.

Manski, C.F. 2007. *Identification for Prediction and Decision*. Cambridge, MA: Harvard University Press.

Mullahy, J. 2018. "Individual Results May Vary: Inequality-Probability Bounds for Some Health-Outcome Treatment Effects." *Journal of Health Economics* 61: 151-162.

Mullahy, J. 2022. "Investigating Health-Related Time Use with Partially Observed Data." *Review of Economics of the Household* 20: 103-121.

Stevens, S.S. 1946. "On the Theory of Scales of Measurement." *Science*, New Series 103: 677-680.

## Appendix

### A.1: Lower Bounds

*Proof of Proposition 1*— For each $\delta \in T$:

$$\Pr(\Delta \le \delta) = \Pr(y_1 - y_0 \le \delta) \tag{A1.1}$$

$$= \Pr(y_1 \le y_0 + \delta) \tag{A1.2}$$

$$\ge \Pr(y_1 \le t \le y_0 + \delta) \quad \text{for any } t \in V \tag{A1.3}$$

$$= \Pr(y_1 \le t \;\wedge\; t \le y_0 + \delta) \tag{A1.4}$$

$$\ge \max\{0, \Pr(y_1 \le t) + \Pr(y_0 + \delta \ge t) - 1\} \tag{A1.5}$$

$$= \max\{0, \Pr(y_1 \le t) + (1 - \Pr(y_0 + \delta < t)) - 1\} \tag{A1.6}$$

$$= \max\{0, \Pr(y_1 \le t) - \Pr(y_0 + \delta < t)\}, \tag{A1.7}$$

where (A1.5) is a BFH LB on (A1.4) for any $t \in V$, which is a valid LB since $\Pr(\Delta \le \delta) \ge \max\{0, \Pr(y_1 \le t) - \Pr(y_0 + \delta < t)\}$ and which is identifiable since it relies only the identified marginal distributions $\Pr(y_j)$. This proves part (a) of Proposition 1.

Since (A1.7) is the largest valid bound at each $t \in V$ (see *Intuition* below) then the best LB is the maximum of (A1.7) over $t \in V$, i.e.

$$LB^*(\Pr(\Delta \le \delta)) = \max_{t \in V} \max\{0, \Pr(y_1 \le t) - \Pr(y_0 + \delta < t)\}. \tag{A1.8}$$

This proves part (b) of Proposition 1. ■

*Intuition*— Defining an identifiable best lower bound requires using only information contained in the marginal distributions $\Pr(y_j)$ that are identified by assumption. The basic structure of any such bounds will be based on a difference in marginal probabilities of the sort shown in (A1.8). This can be visualized in figure A2 for the case of M=3, $\delta = 0$, and $t = 2$. At $t = 2$ the top panel depicts $\Pr(y_1 \le t) - \Pr(y_0 < t)$

in expression (2.a) as the sum of the blue dots' $p_{jk}$ minus the sum of the orange squares' $p_{jk}$. The bottom panel depicts $\Pr(y_1 \le t) - \Pr(y_0 \le t)$ in expression (2.c) analogously. It is immediately evident that at $t = 2$—and, by similar reasoning, for any $t \in V$—(2.a) cannot be smaller and will in general be greater than (2.c).

Any bound claiming to be a best bound will thus turn out to be the difference in the overall joint probability of a subset $R_{(+)}$ of the points $\{(y_0, y_1) | \Delta \le \delta\}$ minus the overall joint probability of a subset $R_{(-)}$ of the points $\{(y_0, y_1) | \Delta > \delta\}$, i.e.

$$\sum_{(y_0,y_1)\in R_{(+)}} \Pr(y_0, y_1) - \sum_{(y_0,y_1)\in R_{(-)}} \Pr(y_0, y_1). \qquad \text{(A1.9)}$$

Such a difference will necessarily be no larger than the overall joint probability of the points in $R_{(+)}$ which in turn is no larger than the overall joint probability of all the points satisfying $\{(y_0, y_1) | \Delta \le \delta\}$. It is thus a valid lower bound.

To appreciate that (A1.8) is the best such lower bound suppose the possible existence of an even better (greater) lower bound,

$$LB^{**}(\Pr(\Delta \le \delta)) = \max\{0, \Pr(A_1) - \Pr(A_0)\} \ge LB^{*}(\Pr(\Delta \le \delta)) \qquad \text{(A1.10)}$$

for events $A_j$ in the marginal distributions $\Pr(y_j)$. For LB** to exceed LB* it must hold either that (a) $\Pr(A_1) > \Pr(y_1 \le t)$ (e.g. $\Pr(A_1) = \Pr(y_1 \le t + k)$, $k \in \{1, 2, \ldots\}$ ), or that (b) $\Pr(A_0) < \Pr(y_0 + \delta < t)$ (e.g. $\Pr(A_0) < \Pr(y_0 + \delta < t - k)$, $k \in \{1, 2, \ldots\}$ ) so that for some index sets $S_{(+)}$ and $S_{(-)}$ defined implicitly by $\Pr(A_0)$ and $\Pr(A_1)$,

$$\sum_{(y_0,y_1)\in S_{(+)}} \Pr(y_0, y_1) - \sum_{(y_0,y_1)\in S_{(-)}} \Pr(y_0, y_1) \ \ge\ \sum_{(y_0,y_1)\in R_{(+)}} \Pr(y_0, y_1) - \sum_{(y_0,y_1)\in R_{(-)}} \Pr(y_0, y_1). \qquad \text{(A1.11)}$$

It is straightforward to show, however, that for either (a) or (b) the difference $\Pr(A_1) - \Pr(A_0)$ results in at least one point in $\{(y_0, y_1) | \Delta > \delta\}$ being included in the index set $S_{(+)}$ in (A1.11) so that LB** cannot be a valid lower bound since it may exceed $\Pr(\Delta \le \delta)$.

**A.2: Upper Bounds**

*Proof of Proposition 2*— For each $\delta \in T$:

$$\Pr(\Delta \le \delta) = \Pr(y_1 - y_0 \le \delta) \tag{A2.1}$$

$$= 1 - \Pr(y_1 - y_0 > \delta) \tag{A2.2}$$

$$= 1 - \Pr(y_1 - \delta > y_0) \tag{A2.3}$$

$$\le 1 - \Pr(y_1 - \delta \ge^a t \ge^b y_0) \quad \text{for any } t \in V. \tag{A2.4}$$

If one and only one of the inequalities $\ge^a$ or $\ge^b$ in (A2.4) is strict then (A2.4) necessarily follows from (A2.3) since then $\Pr(y_1 - \delta > t \ge y_0) \le \Pr(y_1 - \delta > y_0)$ or $\Pr(y_1 - \delta \ge t > y_0) \le \Pr(y_1 - \delta > y_0)$. If neither of the inequalities $\ge^a$ or $\ge^b$ in (A2.4) is strict then (2.4) will not necessarily follow from (A2.3). If both inequalities $\ge^a$ and $\ge^b$ in (A2.4) are strict then (A2.4) will be too large to serve as the basis of a best UB relative to other identifiable alternatives. The relevant cases for determining a best UB thus involve one and only one of $\ge^a$ or $\ge^b$ being strict. It is shown below that $UB^*(\Pr(\Delta \le \delta))$ is the same regardless of which one is strict.

Case 1: $\ge^a$ is strict, $\ge^b$ is weak. Then continuing from (A2.4):

$$1 - \Pr(y_1 - \delta > t \ge y_0) = 1 - \Pr(y_1 - \delta > t \ \wedge \ t \ge y_0) \tag{A2.5}$$

$$\le 1 - \max\{0, \Pr(y_1 > t + \delta) + \Pr(t \ge y_0) - 1\} \tag{A2.6}$$

$$= 1 - \max\{0, (1 - \Pr(y_1 \le t + \delta)) + \Pr(y_0 \le t) - 1\} \tag{A2.7}$$

$$= 1 - \max\{0, \Pr(y_0 \le t) - \Pr(y_1 \le t + \delta)\}, \tag{A2.8}$$

where (A2.6) is one minus a BFH LB on the subtrahend in (A2.5). This proves expression (3.a) in part (a) of Proposition 2.

Then:

$$UB^*\left(\Pr\left(\Delta \le \delta\right)\right) = \min_{t\in V}\left\{1 - \max\left\{0, \Pr\left(y_0 \le t\right) - \Pr\left(y_1 \le t+\delta\right)\right\}\right\} \quad \text{(A2.9)}$$

$$= 1 - \max_{t\in V} \max\left\{0, \Pr\left(y_0 \le t\right) - \Pr\left(y_1 \le t+\delta\right)\right\}. \quad \text{(A2.10)}$$

This proves expression (3.c) in part (b) of Proposition 2.

Case 2: $\ge^b$ is strict, $\ge^a$ is weak. Then continuing from (A2.4):

$$1-\Pr\left(y_1 - \delta \ge t > y_0\right) = 1 - \Pr\left(y_1 - \delta \ge t \ \wedge \ t > y_0\right) \quad \text{(A2.11)}$$

$$\le 1 - \max\left\{0, \Pr\left(y_1 \ge t+\delta\right) + \Pr\left(t > y_0\right) - 1\right\} \quad \text{(A2.12)}$$

$$= 1 - \max\left\{0, \left(1 - \Pr\left(y_1 < t+\delta\right)\right) + \Pr\left(y_0 < t\right) - 1\right\} \quad \text{(A2.13)}$$

$$= 1 - \max\left\{0, \Pr\left(y_0 < t\right) - \Pr\left(y_1 < t+\delta\right)\right\}, \quad \text{(A2.14)}$$

where (A2.12) is one minus a BFH LB on the subtrahend in (A2.11). This proves expression (3.b) in part (a) of Proposition 2.

Then:

$$UB^*\left(\Pr\left(\Delta \le \delta\right)\right) = \min_{t\in V}\left\{1 - \max\left\{0, \Pr\left(y_0 < t\right) - \Pr\left(y_1 < t+\delta\right)\right\}\right\} \quad \text{(A2.15)}$$

$$= 1 - \max_{t\in V} \max\left\{0, \Pr\left(y_0 < t\right) - \Pr\left(y_1 < t+\delta\right)\right\}. \quad \text{(A2.16)}$$

This proves expression (3.d) in part (b) of Proposition 2.

The equivalence of (3.c) and (3.d) claimed in part (b) of Proposition 2 can be shown by noting that $\Pr\left(y_j < 0\right) = 0$ and $\Pr\left(y_j \le M\right) = 1$ so that the $\max_{t\in V}$ operation is essentially running over the same arguments but offset by one unit. To see this concretely consider M=3 and $\delta = 0$. Then from (A2.10),

$$1-\max_{t\in V}\max\left\{0,\Pr\left(y_0\le t\right)-\Pr\left(y_1\le t\right)\right\}=1-\max\left\{\begin{matrix}\max\left\{0,\Pr\left(y_0\le 0\right)-\Pr\left(y_1\le 0\right)\right\},\\ \max\left\{0,\Pr\left(y_0\le 1\right)-\Pr\left(y_1\le 1\right)\right\},\\ \max\left\{0,\Pr\left(y_0\le 2\right)-\Pr\left(y_1\le 2\right)\right\},\\ 0\end{matrix}\right\} \quad \text{(A2.17)}$$

since $\max\left\{0,\Pr\left(y_0\le 2\right)-\Pr\left(y_1\le 2\right)\right\}$, whereas from (A2.16),

$$\begin{aligned}1-\max_{t\in V}\max\left\{0,\Pr\left(y_0< t\right)-\Pr\left(y_1< t\right)\right\}&=1-\max\left\{\begin{matrix}0,\\ \max\left\{0,\Pr\left(y_0< 1\right)-\Pr\left(y_1< 1\right)\right\},\\ \max\left\{0,\Pr\left(y_0< 2\right)-\Pr\left(y_1< 2\right)\right\},\\ \max\left\{0,\Pr\left(y_0< 3\right)-\Pr\left(y_1< 3\right)\right\}\end{matrix}\right\}\\ &=1-\max\left\{\begin{matrix}0,\\ \max\left\{0,\Pr\left(y_0\le 0\right)-\Pr\left(y_1\le 0\right)\right\},\\ \max\left\{0,\Pr\left(y_0\le 1\right)-\Pr\left(y_1\le 1\right)\right\},\\ \max\left\{0,\Pr\left(y_0\le 2\right)-\Pr\left(y_1\le 2\right)\right\}\end{matrix}\right\}\end{aligned} \quad \text{(A2.18)}$$

since $\max\left\{0,\Pr\left(y_0<0\right)-\Pr\left(y_1<0\right)\right\}=\max\left\{0,0-0\right\}=0$. ■

**A.3: Tightening Bounds by Invoking Additional Assumptions**

As with continuous outcomes it is sometimes possible to obtain tighter bounds if one is willing to invoke nonverifiable assumptions about features of the joint distribution $\Pr\left(y_0,y_1\right)$. As an example of how such assumptions may come into play with count-valued outcomes suppose one considers as a candidate LB

$$\Pr\left(y_0=M\right)+\Pr\left(y_1=0\right). \quad \text{(A3.1)}$$

The implied overall probability of this sum is depicted in figure A2 where the overall probability (A3.1) is the sum of the blue dots' $p_{jk}$ plus twice the magenta dot's $p_{jk}$. Obviously this cannot be a valid LB because of the double counting of $p_{M0}$. Were $p_{M0}$ known then it could be subtracted from (A3.1) to give a valid LB,

$$\max\left\{0,\Pr\left(y_0=M\right)+\Pr\left(y_1=0\right)-p_{M0}\right\}. \qquad \text{(A3.2)}$$

Absent knowledge of $p_{M0}$ a valid LB could be obtained as

$$\max\left\{0,\Pr\left(y_0=M\right)+\Pr\left(y_1=0\right)-UB\left(p_{M0}\right)\right\}, \qquad \text{(A3.3)}$$

but all that is known without additional assumptions is that $p_{M0}\in\left[0,1\right]$, implying $UB\left(p_{M0}\right)=1$, which is uninformative.

For illustrative purposes suppose one is willing to assume that $p_{jk}$ is non-increasing in $j-k$ for $j\geq k$. That is, larger-magnitude differences $\left|\Delta\right|$ are less likely than smaller-magnitude ones for nonpositive $\Delta$. Call this assumption *lower probability of large differences*, or LPLD. Suppose that in addition to the bounds described in the paper's section 2 one considers the possibility that the best LB on $\Pr\left(\Delta\leq\delta\right)$ might be improved by invoking LPLD. Specifically note that under LPLD $\frac{2}{(M+1)(M+2)}$ is the largest possible value of $p_{M0}$.[4] Thus under LPLD

$$\max\left\{0,\Pr\left(y_0=M\right)+\Pr\left(y_1=0\right)-\frac{2}{(M+1)(M+2)}\right\} \qquad \text{(A3.4)}$$

is a valid LB. Under LPLD note that the lower bound (A3.4) will be valid only for $\delta\geq 0$. If $\delta<0$ then the events $y_0=M$ and $y_1=0$ in (A3.4) will include points $\left(y_0,y_1\right)$ corresponding to both $\Delta\leq\delta$ and $\Delta>\delta$.

The rightmost column in table 1 shows $LB^*\left(\Pr\left(\Delta\leq\delta\right)\right)$ under LPLD. Invoking LPLD results in improvements on the best LBs at $\delta=0$ and $\delta=1$. Whether such tighter bounds are worth the cost of potentially invalid assumptions will depend on particular circumstances.

---

[4] To see this set $\Pr\left(y_0=j,y_1=k\right)=0$ for all points $\left(y_0,y_1\right)$ with $k>j$. This leaves at most $\frac{(M+1)(M+2)}{2}$ points $\left(y_0,y_1\right)$ with nonzero joint probability. If all such points have equal joint probability that probability would be $\frac{2}{(M+1)(M+2)}$. Under LPLD this is the largest possible $p_{M0}$ and is thus a UB on it.

Table 1: $LB^*\left(\Pr\left(\Delta \le \delta\right)\right)$ and $UB^*\left(\Pr\left(\Delta \le \delta\right)\right)$ for Weekly Total Days of Vigorous or Moderate Exercise Given Season-of-Year Treatment ( $y_0$ is Fall/Winter; $y_1$ is Spring/Summer)—2015 Health Survey of England

| $\delta$ | $UB^*\left(\Pr\left(\Delta \le \delta\right)\right)$ | $LB^*\left(\Pr\left(\Delta \le \delta\right)\right)$ | $LB^*\left(\Pr\left(\Delta \le \delta\right)\right)$ under LPLD |
|---|---|---|---|
| -7 | .147 | 0 | 0 |
| -6 | .180 | 0 | 0 |
| -5 | .240 | 0 | 0 |
| -4 | .318 | 0 | 0 |
| -3 | .425 | 0 | 0 |
| -2 | .556 | 0 | 0 |
| -1 | .643 | 0 | 0 |
| 0 | .946 | .303 | .421 |
| 1 | 1 | .393 | .421 |
| 2 | 1 | .521 | .521 |
| 3 | 1 | .635 | .635 |
| 4 | 1 | .720 | .720 |
| 5 | 1 | .788 | .788 |
| 6 | 1 | .827 | .827 |
| 7 | 1 | 1 | 1 |

Figure 1: Sample Space, TEs, and Joint Probability Distribution for M=3—
Blue Dots Indicate $\Delta \leq 0$, Orange Squares Indicate $\Delta > 0$

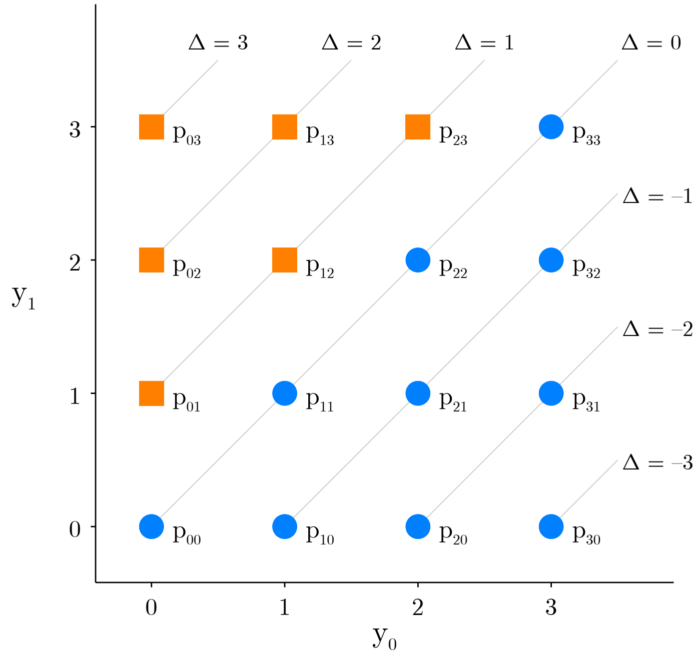

Figure 2: $LB^*(\Pr(\Delta \le \delta))$ and $UB^*(\Pr(\Delta \le \delta))$ for Weekly Total Days of Vigorous or Moderate Exercise Given Season-of-Year Treatment ( $y_0$ is Fall/Winter; $y_1$ is Spring/Summer)—
2015 Health Survey of England

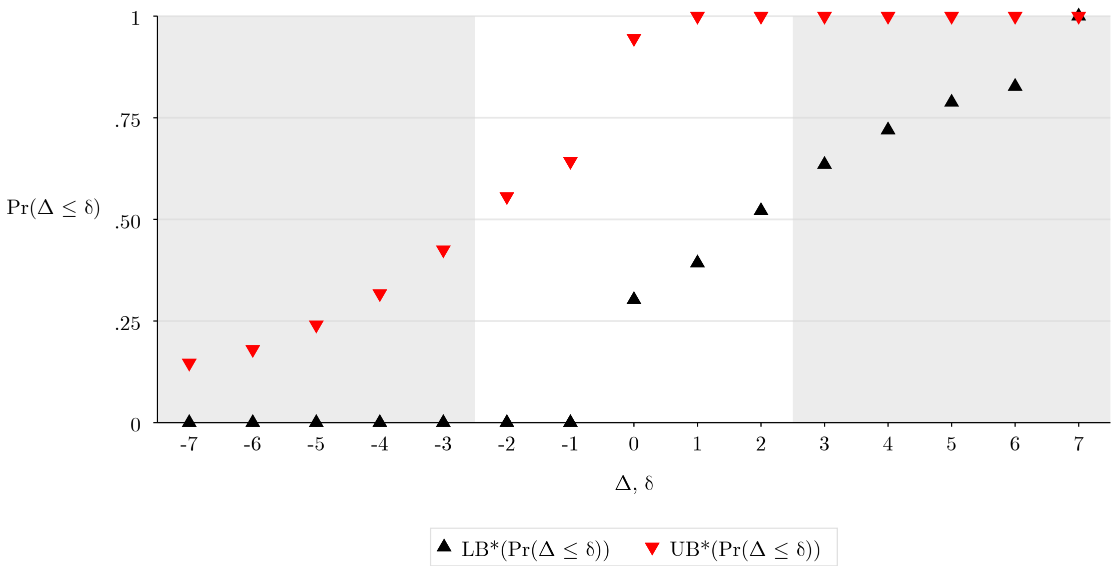

Figure A1: Contrasting Expressions (2.a) and (2.c), M=3, $\delta = 0$, and $t = 2$ —
Top and Bottom Panels Correspond to (2.a) and (2.c), Respectively

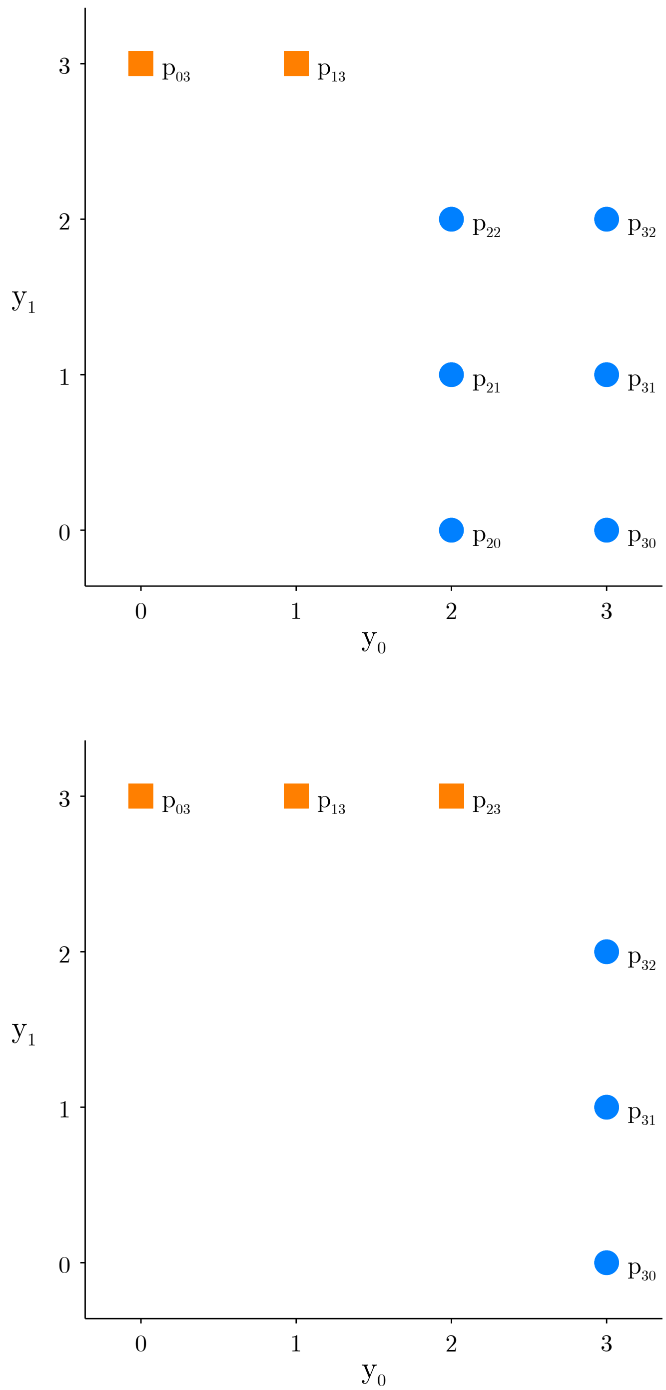

Figure A2: Possible Improvement in $LB*\left(\Pr\left(\Delta \le \delta\right)\right)$ from Assuming LPLD—
$\Pr\left(y_0 = M\right) + \Pr\left(y_1 = 0\right)$ is the Sum of the Blue Dots' $p_{jk}$ Plus Twice the Magenta Dot's $p_{jk}$