Letter to the Editor

RE: "Defining and Identifying Local Average Treatment Effects"

Etsuji Suzuki[a, b], Eiji Yamamoto[c]

[a]Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan
[b]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, United States
[c]Okayama University of Science, Okayama, Japan

Corresponding author: Etsuji Suzuki, Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan. Telephone: +81-86-235-7174. Email: etsuji-s@cc.okayama-u.ac.jp

Conflicts of Interest: None declared.

Sources of Funding: ES is supported by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers JP19KK0418 and JP23K09740). The funding sources had no involvement in study design; the collection, analysis, and interpretation of data; the writing of the article; and the decision to submit the article for publication.

Acknowledgments: None.

A recent article in The *AJE* Classroom section by Naimi and Whitcomb (1) provides a thoughtful explanation about instrumental variables (IVs) to estimate compliance-adjusted effects in randomized controlled trials (RCTs). They used a hypothetical example of an RCT to evaluate the effect of aspirin on headache in a well-defined cohort. Following their notation, we let *X* denote a binary treatment assignment (1 = assigned to aspirin, 0 = assigned to placebo) and *A* denote a binary treatment (1 = aspirin, 0 = placebo). We also let *Y* denote an outcome variable used to assess headache severity, with higher values indicating worse severity. In the counterfactual framework, we let A^x and Y^x denote the potential outcomes of *A* and *Y*, respectively, if, possibly contrary to fact, there had been interventions to set *X* to *x*. Similarly, we let Y^a denote the potential outcomes of *Y* if, possibly contrary to fact, there had been interventions to set *A* to *a*.

The estimand of interest in their article is $\mathrm{E}[Y^{a=1} - Y^{a=0} | A^{x=1} > A^{x=0}]$, which is referred to as the local average treatment effect (LATE). Then, they emphasized the importance of "monotonicity" condition of treatment assignment *X* on treatment *A*, which was described as not $A^{x=1} \geq A^{x=0}$ but $A^{x=1} > A^{x=0}$ for all individuals. Note that the latter is stronger than the former. In the setting considered here, the inequality $A^{x=1} > A^{x=0}$ holds if and only if $A^{x=1} = 1$ and $A^{x=0} = 0$, and this type of individual is called a "complier". After introducing the Assumptions 1 to 3 for IVs, they additionally used the (stronger) "monotonicity" condition to provide a proof that the IV estimator (i.e., $\{\mathrm{E}[Y|X=1] - \mathrm{E}[Y|X=0]\}/\{\mathrm{E}[A|X=1] - \mathrm{E}[A|X=0]\}$) becomes equivalent to the LATE. Their conclusion is correct. However, it is important to note that under the (stronger) "monotonicity" condition, there are no "always-takers", "never-takers", or "defiers" in the population. In this scenario, only "compliers" exist. This means that *X* = *A* for all individuals, and the common cause *C* in their Figure 1 does not exist. In other words, the LATE becomes trivially equivalent to the average treatment effect (ATE) in the total population (i.e., $\mathrm{E}[Y^{a=1} - Y^{a=0}]$). Thus, in this specific scenario, we may simply calculate $\mathrm{E}[Y|A=1] - \mathrm{E}[Y|A=0]$ to obtain the ATE as well as the LATE. Note also that $A^{x=1} > A^{x=0}$ is stronger than Assumption 1 in their article (i.e., $A^{x=1} - A^{x=0} \neq 0$), under which there are no "always-takers" or "never-takers" in the population.

It is worth emphasizing that, even if we use a (weaker) "monotonicity" condition (i.e., $A^{x=1} \geq A^{x=0}$), the IV estimator becomes equivalent to the LATE. As illustrated by them, under SUTVA, exclusion restriction (Assumption 2), and exchangeability (Assumption 3), the numerator of the IV estimator becomes $\mathrm{E}[(Y^{a=1} - Y^{a=0})(A^{x=1} - A^{x=0})]$. Unlike their explanation, however, this can be decomposed without using Assumption 1 (i.e., $A^{x=1} - A^{x=0} \neq 0$) as follows:

$$
\begin{aligned}
&\mathrm{E}[(Y^{a=1} - Y^{a=0})(A^{x=1} - A^{x=0})] \\
&= \mathrm{E}[(Y^{a=1} - Y^{a=0})(A^{x=1} - A^{x=0}) | A^{x=1} - A^{x=0} = 1] \Pr(A^{x=1} - A^{x=0} = 1) \\
&\qquad + \mathrm{E}[(Y^{a=1} - Y^{a=0})(A^{x=1} - A^{x=0}) | A^{x=1} - A^{x=0} = 0] \Pr(A^{x=1} - A^{x=0} = 0) \\
&\qquad + \mathrm{E}[(Y^{a=1} - Y^{a=0})(A^{x=1} - A^{x=0}) | A^{x=1} - A^{x=0} = -1] \Pr(A^{x=1} - A^{x=0} = -1)
\end{aligned}
$$

$$= \mathrm{E}[Y^{a=1} - Y^{a=0} | A^{x=1} - A^{x=0} = 1] \Pr(A^{x=1} - A^{x=0} = 1) \\ + \mathrm{E}[-(Y^{a=1} - Y^{a=0}) | A^{x=1} - A^{x=0} = -1] \Pr(A^{x=1} - A^{x=0} = -1). \text{ [Eq. 1]}$$

Thus, $\mathrm{E}[(Y^{a=1} - Y^{a=0})(A^{x=1} - A^{x=0})]$ is a weighted sum of the ATE among the "compliers" and the *negative* ATE among the "defiers". Note that, without Assumption 1, $\Pr(A^{x=1} - A^{x=0} = 1) + \Pr(A^{x=1} - A^{x=0} = -1) = 1$ does not generally hold. Importantly, both $\mathrm{E}[Y^{a=1} - Y^{a=0} | A^{x=1} - A^{x=0} = 1] = \mathrm{E}[Y^{x=1} - Y^{x=0} | A^{x=1} - A^{x=0} = 1]$ and $\mathrm{E}[-(Y^{a=1} - Y^{a=0}) | A^{x=1} - A^{x=0} = -1] = \mathrm{E}[Y^{x=1} - Y^{x=0} | A^{x=1} - A^{x=0} = -1]$ hold, and $\mathrm{E}[(Y^{a=1} - Y^{a=0})(A^{x=1} - A^{x=0})]$ is equivalent to the intention-to-treat (ITT) effect (i.e., $\mathrm{E}[Y^{x=1} - Y^{x=0}]$). Furthermore, the ATE in the total population is larger than the ITT effect by $\mathrm{E}[Y^{a=1} - Y^{a=0} | A^{x=1} - A^{x=0} = 0] \Pr(A^{x=1} - A^{x=0} = 0) + 2 \times \mathrm{E}[Y^{a=1} - Y^{a=0} | A^{x=1} - A^{x=0} = -1] \Pr(A^{x=1} - A^{x=0} = -1)$. See Appendix S1 for details.

Finally, under the (weaker) monotonicity condition (i.e., $A^{x=1} \geq A^{x=0}$), there are no "defiers" in the population, and $\Pr(A^{x=1} - A^{x=0} = -1) = 0$ holds in equation 1. Thus, the numerator of the IV estimator becomes $\mathrm{E}[Y^{a=1} - Y^{a=0} | A^{x=1} - A^{x=0} = 1] \Pr(A^{x=1} - A^{x=0} = 1)$. Furthermore, unlike their explanation about Assumption 3, $\mathrm{E}[A^x | X = x] = \mathrm{E}[A^x]$ $(x = 0,1)$ cannot be implied by no confounding of the effect of *X* on *Y*; rather it holds in their example because *X* is randomized, and the denominator of the IV estimator becomes $\mathrm{E}[A^{x=1} - A^{x=0}] = \Pr(A^{x=1} - A^{x=0} = 1)$. Dividing the numerator by the denominator, we obtain $\mathrm{E}[Y^{a=1} - Y^{a=0} | A^{x=1} - A^{x=0} = 1]$. This is LATE.

Reference

1. Naimi AI, Whitcomb BW. Defining and identifying local average treatment effects [published online ahead of print February 29, 2024]. *Am J Epidemiol.* (doi:10.1093/aje/kwae009).