Published in final edited form as:

Stat Med. 2016 January 15; 35(1): 147–160. doi:10.1002/sim.6612.

Latent class instrumental variables: A clinical and biostatistical perspective

Stuart G. Baker^a, Barnett S. Kramer^a, and Karen S. Lindeman^b

^aDivision of Cancer Prevention, National Cancer Institute, Bethesda MD, USA

^bDepartment of Anesthesiology, Johns Hopkins Medical Institutions, Baltimore MD, USA

Abstract

In some two-arm randomized trials, some participants receive the treatment assigned to the other arm as a result of technical problems, refusal of a treatment invitation, or a choice of treatment in an encouragement design. In some before-and-after studies, the availability of a new treatment changes from one time period to the next. Under assumptions that are often reasonable, the latent class instrumental variable (IV) method estimates the effect of treatment received in the aforementioned scenarios involving all-or-none compliance and all-or-none availability. Key aspects are four initial latent classes (sometimes called principal strata) based on treatment received if in each randomization group or time period, the exclusion restriction assumption (in which randomization group or time period is an instrumental variable), the monotonicity assumption (which drops an implausible latent class from the analysis), and the estimated effect of receiving treatment in one latent class (sometimes called efficacy, the local average treatment effect, or the complier average causal effect). Since its independent formulations in the biostatistics and econometrics literatures, the latent class IV method (which has no well-established name) has gained increasing popularity. We review the latent class IV method from a clinical and biostatistical perspective, focusing on underlying assumptions, methodological extensions, and applications in our fields of obstetrics and cancer research.

Keywords

all-or-none compliance; causal inference; encouragement design; observational; paired availability design; principal stratification; randomized trial

1. Introduction

Consider two randomization groups with one group assigned treatment T_0 and the other treatment T_1 . Under all-or-none compliance [1] some participants assigned T_0 may receive T_1 and some assigned T_1 may receive T_0 (Figure 1). All-or-none compliance arises under a variety of situations including the following examples.

*Correspondence to: Stuart G., Divisions of Cancer Prevention, National Cancer Institute, 9609 Medical Center Dr 5E638 Bethesda MD 20892-9789. sb16i@nih.gov.

1. *Technical difficulty.* Investigators randomized participants to $T0$ (cyroanalgesia) or $T1$ (cervical epidural injection). Due to technical problems, some participants assigned $T1$ could not receive $T1$ and received $T0$ instead [2].
2. *Treatment invitation.* Investigators randomly assigned participants to $T0$ (no mammography) or an invitation for $T1$ (mammography). Some participants offered $T1$ refused and received $T0$ by default [3].
3. *Direct encouragement design.* Investigators randomly assigned smokers to usual encouragement for $T1$ (stop smoking) where the default is $T0$ (continue smoking) versus extra encouragement for $T1$ over $T0$. In each group some participants received $T0$ and some received $T1$ [4].
4. *Indirect encouragement design.* Investigators randomly assigned patients to physicians reminded to offer $T1$ (vaccination) where the default is $T0$ (no vaccination) versus physicians not reminded to offer $T1$. In each group some participants received $T0$ and some received $T1$ [5].

Similarly, consider two time periods that can be treated like randomization groups under certain assumptions. Under all-or-none availability the standard treatment $T0$ is available in both time periods, and a new treatment $T1$ has a higher level of availability in one time period. In each time period, some participants receive $T0$ and some receive $T1$ (Figure 2). This scenario is a key component of the paired availability design [6] that we discuss in Section 5.

The latent class IV method estimates the effect of treatment received while avoiding self-selection bias with all-or-none compliance or all-or-none availability. The latent class IV method dates back to at least the mid 1990's with independent formulations by Baker and Lindeman [6] in the biostatistics literature and Imbens and Angrist [7] in the econometrics literature. There is no well-established name for this method. Our terminology of latent class IV comes from a previous review [8] that used the terms “latent class” and “instrumental variables”. Other names are “the IV estimand embedded in the Rubin Causal Model [9], “principal stratification approach to broken randomized experiments,” [10], “modern instrumental variables literature” [11], “IV assumptions and estimation for binary IV and binary treatment” [12], and “instrumental variable analysis ... in comparative effectiveness research” [13]. For our review we bring a clinical and biostatistical perspective and cover topics not covered or covered sparingly in other recent reviews [11-13]. Our emphasis is on assumptions, methodological extensions, and applications in our fields of obstetrics and cancer research.

2 Latent class IV method

The latent class IV method has the following five distinguishing characteristics.

First, there are two *randomization groups*, either by design or under assumptions.

Second, there are four *latent classes* of the form (treatment received if in group assigned $T0$, treatment received if in group assigned $T1$), namely $(T0, T0)$, $(T0, T1)$, $(T1, T0)$, and $(T1, T1)$. Angrist, Imbens, and Rubin [9] labeled these classes, “never-taker”, “complier”,

“defier” and “always-taker,” respectively. For a more general post-randomization variables than treatment received, Frangakis and Rubin [14] introduced the terminology principal strata for these four latent classes.

Third, the *exclusion restriction* assumption says that the probability of outcome does not depend on group in $(T0, T0)$ and $(T1, T1)$. Imbens and Angrist [7] introduced the terminology of exclusion restriction in this context. The exclusion restriction assumption is closely related to the concept of instrumental variable. An instrumental variable is a variable that is not directly associated with outcome but is associated with variable known to affect outcome [15]. The exclusion restriction says that randomization group is an instrumental variable for never-takers and always-takers.

Fourth, the *monotonicity* assumption says there are no persons in latent class $(T1, T0)$, namely, there are no defiers. Imbens and Angrist [7] introduced the name monotonicity. The monotonicity assumption is rooted in consistent preferences. In the direct encouragement design, monotonicity says that no person would receive $T0$ when encouraged to receive $T1$ and receive $T1$ when not encouraged to receive $T1$. In the indirect encouragement design, monotonicity says that no person would receive $T0$ when treatment providers are encouraged to offer $T1$ and receive $T1$ when the treatment providers are not encouraged to offer $T1$. As will be discussed, in the paired availability design, monotonicity is a consequence of stable preferences over time and an availability of $T1$ in one time period that subsumes availability of $T1$ in the other time period.

Fifth, based on the above assumptions, the estimated treatment effect in the complier latent class $(T0, T1)$ avoids bias from self-selection. The latent class IV treatment effect among compliers goes by various names including efficacy [16-18], method effectiveness [19], effect of receipt of treatment [6], local average treatment effect [7], and complier average causal effect [20,21].

Implicit in the latent class IV formulation is the additional assumption that a person's outcome is unaffected by the treatment received by another person [9]. We use the terminology *restricted latent class IV method* for the special case of the latent class IV method which is applied when all participants in the control group receive $T0$ and participants in the experimental group receive either $T0$ or $T1$. In this case, there are only two latent classes, $(T0, T0)$ and $(T0, T1)$, so the monotonicity assumption is not applicable. Table 1 summarizes the key assumptions for the latent class IV and the restricted latent class IV methods.

2. Historical perspective

An early impetus for the development of the restricted latent class IV method was Zelen's randomized consent design [22] involving randomization to either $T0$ or an offer of $T1$ with refusers receiving $T0$. In 1983, one of us (SGB), while a graduate student in the Harvard Department of Biostatistics chaired by Marvin Zelen, proposed a restricted latent class IV method with maximum likelihood estimates to analyze Zelen's randomized consent design. (Supplementary Appendix A). In 1984, Bloom [23] formulated restricted latent class IV method to estimate the mean difference in outcomes among compliers in a randomized trial

involving controls ($T0$) and experimental group consisting of “no-shows” ($T0$) and those receiving the intervention ($T1$). Building on earlier work by Tarawoto et al. [24], in 1991, Sommer and Zeger used a restricted latent class IV method to estimate relative risk in a randomized trial of no vitamin A supplement ($T0$) and vitamin A supplement ($T1$), where some children randomized to vitamin A supplement did not receive it because of a distribution failure. In 1991, Connor, Prorok, and Weed [17] also used the restricted latent class IV method to estimate relative risk.

In 1989, Permutt and Hebel [25] used simultaneous equations to estimate the effect of maternal smoking on birth weight in a randomized trial with a direct encouragement design. For the special case of all-or-none compliance, they specified four latent classes, the exclusion restriction, and monotonicity to compute a simultaneous equation estimate identical to the latent class IV estimate. In 1994, Imbens and Angrist [7] and Baker and Lindeman [6] independently formulated the latent class IV method from first principles. Angrist and Imbens [7] estimated a difference in continuous outcomes. Baker and Lindeman [6] computed a maximum likelihood estimate for a difference in binary outcomes.

The latent class IV method should not be confused with other methods to estimate treatment effect under all-or-none compliance that yield the latent class IV estimate under different assumptions. Newcombe [2] computed a latent class IV estimate based on a linear model for the effect of treatment on outcome. Robins [26] derived a latent class IV estimate (in his Table 1, row 13) based on assumptions that differed from those with the latent class IV method. Cuzick, Edwards, and Segnan [27] computed latent classes IV estimate using a “balance equation” without a monotonicity assumption.

Imbens and Angrist [7] framed the latent class IV method as a potential outcomes analysis. The original potential outcomes formulations [28, 29] involved a randomized trial with full compliance. A key aspect of the potential outcomes framework is a potential outcomes notation that expresses outcome as an explicit function of the either the actual randomized group to which a participant was assigned or the unrealized randomized group to which a participant was not assigned [29]. For the latent class IV method, Imbens and Angrist [7] extended the potential outcomes notation to both outcome and treatment received. Taking a different perspective, Baker and Lindeman [6] and Baker, Kramer, and Lindeman [30] framed the latent class IV method as a thought experiment in which the availability in the time periods was reversed. Consequently, they did not use potential outcomes notation. Cox [31] framed the restricted latent class IV method as a hypothetical scenario and also did not use potential outcomes notation.

3. Basic formulation

3.1 Model

Consider the binary outcomes formulation in Baker and Lindeman [6]. Let $Y=0, 1$ denote outcome, and $Z=0, 1$ denote group. Also let r denote latent class, which takes three values under the monotonicity assumption: never-takers (N) for $(T0, T0)$, compliers (C) for $(T0, T1)$, and always-takers (A) for $(T1, T1)$. Let $\pi_r = pr(r)$ and $\beta_{cz} = pr(Y=1 | C, z)$. Under the exclusion restriction $\beta_r = pr(Y=1 | r)$, for $r = N, A$. See Table 2. From the definition of the

latent classes, the following relationships hold. Participants in group $Z=0$ who receive $T0$ are a mixture of never-taker and compliers. Participants in group $Z=0$ who receive $T1$ are always-takers. Participants in group $Z=1$ who receive $T0$ are never-takers. Participants in group $Z=1$ who receive $T1$ are a mixture of compliers and always-takers. Consequently, the basic equations are

$$\begin{aligned} pr(Y=1, \text{receive } T0|Z=0) &= \pi_N \beta_N + \pi_C \beta_{C0}, \\ pr(Y=1, \text{receive } T1|Z=0) &= \pi_A \beta_A, \\ pr(Y=1, \text{receive } T0|Z=1) &= \pi_N \beta_N, \\ pr(Y=1, \text{receive } T1|Z=1) &= \pi_C \beta_{C1} + \pi_A \beta_A. \end{aligned} \quad (1)$$

It is instructive to compare various measures of treatment effect under this model. The latent class IV measure is the effect of treatment in compliers,

$$\Delta_C = pr(Y=1|C, Z=1) - pr(Y=1|C, Z=0) = \beta_{C1} - \beta_{C0}. \quad (2)$$

The intent-to-treat (*ITT*) measure is the effect of assigned treatment,

$$\Delta_{ITT} = pr(Y=1|Z=1) - pr(Y=1|Z=0) = (\pi_N \beta_N + \pi_C \beta_{C1} + \pi_A \beta_A) - (\pi_N \beta_N + \pi_C \beta_{C0} + \pi_A \beta_A) = \pi_C (\beta_{C1} - \beta_{C0}). \quad (3)$$

The per-protocol (*PP*) measure is the effect of treatment received among participants who receive the assigned treatment,

$$\Delta_{PP} = pr(Y=1|\text{receive } T1, Z=1) - pr(Y=1|\text{receive } T0, Z=0) = (\pi_C \beta_{C1} + \pi_A \beta_A) / (\pi_C + \pi_A) - (\pi_N \beta_N + \pi_C \beta_{C0}) / (\pi_N + \pi_C). \quad (4)$$

The as-treated (*AT*) measure is the effect of treatment received regardless of group; under 50%-50% randomization it equals

$$\begin{aligned} \Delta_{AT} &= \frac{pr(Y=1, \text{receive } T1|Z=1) + pr(Y=1, \text{receive } T1|Z=0)}{pr(\text{receive } T1|Z=1) + pr(\text{receive } T1|Z=0)} \\ &\quad - \frac{pr(Y=1, \text{receive } T0|Z=1) + pr(Y=1, \text{receive } T0|Z=0)}{pr(\text{receive } T0|Z=1) + pr(\text{receive } T0|Z=0)} \\ &= (\pi_C \beta_{C1} + 2\pi_A \beta_A) / (\pi_C + 2\pi_A) - (2\pi_N \beta_N + \pi_C \beta_{C0}) / (2\pi_N + \pi_C). \end{aligned} \quad (5)$$

The intent-to-treat measure is always between 0 and the latent IV measure. The per-protocol and the as-treated measures can be either larger or smaller than the latent class IV measure. See Figure 3, which graphically compares these treatment effect measures as a function of π_C , where $\beta_{C0}=0.1$, $\beta_{C1}=0.2$, $\pi_A=(1-\pi_C)(1/3)$, $\pi_N=(1-\pi_C)(2/3)$, $\beta_A=0.2$, and $\beta_N=0.15$, 0.28, 0.50.

3.2 Estimation

Let n_{zxy} denote the number of persons in group $z=0, 1$ who receive treatment Tx for $x=0, 1$ and have outcome $y=0, 1$. As formulated in Baker and Lindeman [6], the likelihood kernel is

$$L = \{\pi_N(1 - \beta_N) + \pi_C(1 - \beta_{C0})\}^{n_{000}} (\pi_N\beta_N + \pi_C\beta_{C0})^{n_{001}} \{\pi_A(1 - \beta_A)\}^{n_{010}} (\pi_A\beta_A)^{n_{011}} \{\pi_N(1 - \beta_N)\}^{n_{100}} (\pi_N\beta_N)^{n_{101}} \{(\pi_C(1 - \beta_{C1}) + \pi_A(1 - \beta_A))\}^{n_{110}} (\pi_C\beta_{C1} + \pi_A\beta_A)^{n_{111}}. \quad (6)$$

Because there are six independent cell counts and six independent parameters, a simple approach that often yields maximum likelihood estimates is to set observed counts equal to their expected values [6, 31]. One set of six equalities for observed and expected counts is

$$\begin{aligned} n_0(\pi_N\beta_N + \pi_C\beta_{C0}) &= n_{001}, & n_0(\pi_A\beta_A) &= n_{011}, \\ n_1(\pi_N\beta_N) &= n_{101}, & n_1(\pi_C\beta_{C1} + \pi_A\beta_A) &= n_{111}, \\ n_0\pi_A &= n_{01+}, & n_1\pi_N &= n_{10+}, \end{aligned} \quad (7)$$

where “+” in the subscript indicates summation over that subscript, e.g. $n_{01+} = n_{010} + n_{011}$. Let p_r , b_{Cz} , and b_r denote estimates of π_r , β_{Cz} , and β_r , respectively. Solving equation set (7) yields

$$\begin{aligned} p_A &= n_{01+}/n_{0++}, & p_N &= n_{10+}/n_{1++}, & p_C &= 1 - p_A - p_A, \\ b_A &= n_{011}/n_{01+}, & b_N &= n_{101}/n_{10+}, \\ b_{C0} &= (n_{001}/n_{0++} - n_{101}/n_{1++})/p_C, \\ b_{C1} &= (n_{111}/n_{1++} - n_{011}/n_{0++})/p_C. \end{aligned} \quad (8)$$

If $b_{C0} = 0$ and $b_{C1} = 0$, the estimates in equation set (8) are maximum likelihood. Otherwise, constrained maximization is required [32, 33] although the lack of perfect fit may call into question the exclusion restriction or monotonicity assumptions. The perfect fit estimate of the risk difference for treatment effect in compliers is the difference in treatment effect between the two groups divided by the difference in the fraction receiving $T1$ in the two groups,

$$b_{C1} - b_{C0} = (f_1 - f_0)/(q_1 - q_0), \text{ where } f_z = n_{z+1}/n_z \text{ and } q_z = n_{z++}/n_z. \quad (9)$$

Although equation (9) arises from equation set (8), readers may be interested in a graphical derivation [34, 35]. Another useful outcome measure is the perfect fit estimate of relative risk in compliers [36], namely b_{C1}/b_{C0} .

4. Well suited applications

The latent class IV method is particularly well suited to the following applications.

4.1 Cost-effectiveness analysis

A commonly used measure of cost-effectiveness is the discounted treatment cost divided by the discounted life years saved from receipt of treatment. Hence, a denominator of discounted life years saved from treatment assignment would bias the estimate of cost-effectiveness. Applying the restricted latent class IV method to a cancer screening trial with all-or-none compliance, Baker [37] estimated cost-effectiveness of cancer screening.

4.3 Non-inferiority trials

The goal of most randomized trials is to determine if $T1$ is superior to $T0$. In contrast, the goal of a non-inferiority trial is to determine if $T1$ has the same effect on outcome, within a tolerance, as $T0$; this trial is of interest when $T1$ is safer, easier to use, or less expensive than $T0$. In the presence of all-or-none compliance, an intent-to-treat analysis for a non-inferiority trial is problematic because the tolerance only applies to the assignment of treatment. In contrast, the latent class IV method more appropriately estimates the tolerance for treatment received.

4.3 Meta-analysis

When combining estimates of treatment effects in a meta-analysis of randomized trials in the presence of all-or-none compliance, intent-to-treat estimates give undue weight to trials with fewer compliers than to trials with more compliers [38]. A meta-analysis based on the latent class IV method avoids this problem.

5. The paired availability design

Baker and Lindeman [6] formulated the latent class IV method in the context of the paired availability design for historical controls. Their goal was to address a major controversy about the effect of epidural analgesia on the probability of Cesarean section. At the time many investigators thought a randomized trial involving epidural analgesia versus no epidural analgesia would be impractical due to problems of recruitment [30]. Because the paired availability design has many potential applications and is gaining prominence in the medical literature [39, 40], we discuss it in detail.

The paired availability design involves latent class IV estimates for the effect of treatment received based on two time periods when the standard treatment $T0$ is fully available in both time periods and the new treatment $T1$ has greater availability in one time period than the other. Baker and Lindeman [6] originally called the $(T1, T0)$ latent class “irrationals”, but did not explicitly list this latent class because a reviewer wrote that the discussion of “irrationals” was distracting (Supplementary Appendix B). Subsequently, Baker and Lindeman [42] called the four latent classes “never-receivers”, “consistent receivers”, “inconsistent receivers” and “always-receivers.” To reduce random bias from temporal changes, the paired availability design uses data from multiple medical centers and averages the latent class IV estimates over medical centers [6, 35, 41]. The paired availability design involves six assumptions (Table 3) discussed below.

5.1 Assumptions for treating time periods like randomization groups

The following four assumptions are needed to treat time periods like randomization groups for purposes of the latent class IV analysis [35,41].

The *stable population* assumption says that the characteristics of the eligible population related to the probability of outcome do not change over time. Investigators can increase the plausibility of this assumption two ways. First, they can apply this method to a short term endpoint. Second, they can choose medical centers with little or no in- or out- migration, such as geographically isolated medical centers or army medical centers. For example, in the study involving the effect of epidural analgesia on the probability of Cesarean section, it is unlikely that women from outside a geographic area would travel far to deliver at a study medical center because of increased availability of epidural analgesia. If investigators can collect additional data from representative medical centers with no change in availability over time, they can use the estimated background change in treatment effect over time to help compensate for violations of this assumption.

The *stable ancillary care* assumption says that patient management affecting the probability of outcome does not change over time. Investigators can increase the plausibility of this assumption by following strict protocols and minimizing staff changes. With before-and-after studies of the effect of epidural analgesia on the probability of Cesarean section, some investigators reported no changes in protocols or staff over time [35].

The *stable disease progression* assumption says that the time course of disease-related events do not change over time in the absence of treatment.

The *stable evaluation* assumption says that the eligibility criteria and definitions of outcome over time do not change over time. In cancer treatment studies, this assumption would be violated if tumor staging criteria changed over time or if increasingly sensitive diagnostic tests better identify cancer relapse.

5.2 Assumptions for the latent class IV method

The following two assumptions are directly related to the latent class IV method.

The *stable treatment effect* assumption says that the effect of treatment on the probability of outcome does not change over time among always-receivers and never-receivers. This assumption would not hold if *TI* improved over time, as might occur with a surgical technique. The stable treatment effect assumption is the analog of the exclusion restriction for randomized trials.

The *stable preference* assumption says the preference for treatment does not change over time. This assumption could be violated by a widely publicized report of harmful side effects or by direct advertising of the treatment to consumers. The implications of this assumption depend on the type of availability of the new treatment *TI*. Under fixed availability, the times of availability of *TI* in one time period subsume the times of availability of *TI* in the other time period. For example, *TI* in time period 0 is anesthesiology coverage (to provide epidural analgesia to women in labor) from 8 AM to 4

PM daily while $T1$ in time period 1 is anesthesiology coverage from 8 AM to 8 PM daily. As shown in Table 4, stable preferences with fixed availability imply no inconsistent-receivers, namely *monotonicity*. Under random availability, the times of availability of $T1$ occur haphazardly in both time periods with greater overall availability in one time period than another. For example, anesthesiology coverage in time period 0 occurs 40 hours a week at a variety of times that vary from day to day, and anesthesiology coverage in time period 1 occurs for 60 hours a week at a variety of times that vary from day to day. As shown in Table 5, stable preferences with random availability allow for both inconsistent-receivers and consistent receivers. In this case, the designation of inconsistent-receivers versus consistent receivers depends on chance availabilities, so the probability of outcome is the same for inconsistent-receivers and consistent-receivers, an assumption we call *randomicity*. Either monotonicity or randomicity, when coupled with the other assumptions, yields the usual latent class IV estimate of treatment effect in compliers [35].

5.3 Application to obstetric anesthesiology

Contrary to early expectations, various investigators conducted randomized trials on the effect of epidural analgesia on the probability of Cesarean section [30]. Baker and Lindeman [35, 42] computed a latent class IV estimate for each of these randomized trials and combined these estimates into an overall meta-analytic estimate. Both the paired availability design and the meta-analysis of randomized trials yielded similar results -- an estimated effect of epidural analgesia on the probability of Cesarean section that was near zero with narrow 95% confidence intervals. In contrast, estimates from two studies using multivariate adjustments of baseline variables in concurrent controls gave a very different result—a positive effect of epidural analgesia on the probability of Cesarean section with lower bounds of 95% confidence intervals substantially greater than zero. Baker and Lindeman [35, 42] thought the estimates based on the multivariate adjustments were likely biased because they omitted the confounder of intense pain during labor.

5.4 Application to cancer screening

Baker, Kramer, and Prorok [43] modified the paired availability design to estimate the effect of breast cancer screening on cancer incidence in six Swedish counties with increased breast cancer screening over time. They proposed a sensitivity analysis for violations of the exclusion restriction assumption in always-takers and noted possible bias from improvements in therapy over time. Baker [44] proposed a paired availability design for the preliminary evaluation of cancer screening using a short-term endpoint of the number of cancers arising within a year of screening.

5.5 Generalizing treatment effect

To generalize treatment effect in consistent-receivers to treatment effect in all persons, Baker and Lindeman [35] and Baker, Lindeman, and Kramer [45] plotted treatment effect in consistent-receivers as a function of the estimated fraction of participants who were consistent-receivers. They considered four models for extrapolating to the treatment effect if all participants were consistent-receivers. Because simulations showed no extrapolation

model was the best under all circumstances, they recommended a sensitivity analysis using the extrapolation models.

6. Missing or censored outcome data

For clinicians, extensions of the latent class IV method to missing or censored outcomes greatly increase the scope of the applications. For the biostatistician, the extensions require additional assumptions (Table 6) which Frangakis and Rubin [46] called *latent ignorability* and *compound exclusion restriction* for the case of partially missing binary outcomes. We extend their terminology to assumptions involving censored outcomes or partially missing binary outcomes with an auxiliary variable.

6.1 Survival outcomes with competing risks

For the analysis of a randomized trial involving a cancer screening invitation, Baker [37] formulated a restricted latent class IV method for yearly survival data in the presence of competing risks and censoring from end of follow-up. The discrete-time cause-specific hazard for the outcome (breast cancer mortality) is the probability of outcome at time t given that outcome and death from competing risks occur at time t or later. The *latent ignorability* assumption says that, given latent class, the discrete-time cause-specific hazard rate for death from competing risk at time t is the probability of death from competing risk at time t given outcome occurs at time $t+1$ (rather than time t) or later and death from competing risks occurs at time t or later. In other words, the cause-specific hazard for death from competing risks does not depend on an unobserved outcome given the latent class. The *compound exclusion restriction* assumption says that the cause-specific hazard rates for the outcome and death from competing risks do not depend on randomization group in never-takers and always takers. There is also a standard assumption of noninformative censoring at the end of follow-up. Baker [37] derived perfect fit maximum likelihood estimates and also computed estimates based on fitting a polynomial function of time to the hazard rates. Loeys and Goetghebeur [47] and Nie, Cheng, and Small [48] formulated restricted latent class IV methods for continuous survival data without competing risks.

6.2 Partially missing binary outcomes

Frangakis and Rubin [46] formulated the restricted latent class IV method with partially missing binary outcomes, which Baker and Kramer [49] extended to the latent class IV method. The *latent ignorability* assumption says that probability of missing in outcome does not depend on outcome given latent class. The *compound exclusion restriction* says that the probabilities of outcome and missing the outcome do not depend on randomization group among never-takers and always-takers. Mealli et al. [50] modified this framework for the following application. Investigators randomized women to $T1$ (a combined mailing and course invitation) versus $T0$ (the same mailing). Some women assigned $T1$ refused the course invitation and hence received $T0$. Some women in both groups did not return the questionnaire measuring the outcome of breast self-examination skills. Mealli et al. [50] thought that never-takers assigned $T1$ would be less likely to return the questionnaire (as they refused the course invitation) than never-takers assigned $T0$, and assumed that the

probability of missing in outcome does not depend on randomization group for compliers instead of never-takers.

6.3 Auxiliary variable and partially missing binary outcomes

An auxiliary variable is a variable that is observed after randomization and before outcome. In randomized trials with partially missing outcomes, the use of an auxiliary variable can improve the adjustment for missing outcomes. Baker [51] proposed a latent class IV method when using auxiliary variables to adjust for missing outcomes in the presence of all-or-none compliance. Investigators randomized participants to either $T1$ (finasteride) or $T0$ (placebo). The outcome was prostate cancer on biopsy occurring at the end of the study or following a positive test for prostate specific antigen. Thus, missing in the outcome was associated with the auxiliary variable of test result for prostate specific antigen. The *latent ignorability* assumption says that the probability of missing in outcome does not depend on outcome given latent class and the auxiliary variable. The *compound exclusion restriction* says that the probabilities of outcome, auxiliary variable given outcome, and missing in outcome given auxiliary variable do not depend on randomization group among never-takers and always-takers. Baker[51] derived closed-form maximum likelihood estimates for the perfect fit solution.

7. Partially missing data on treatment received

In a proposed indirect encouragement design, investigators randomize patients to physicians reminded to offer $T1$ (discussion of advanced directives) instead of the default $T0$ (no discussion of advanced directives) or physicians not reminded to offer $T1$. A major concern was the cost of interviewing patients to determine if the advanced directive discussion took place. Frangakis and Baker [52] proposed a minimal cost design based on the required precision for the estimated latent class IV treatment effect.

8. Partial compliance

For clinicians, extensions of the latent class IV method from all-or-none compliance to partial compliance would greatly increase the range of applications. However, for biostatisticians, there is a challenge of finding reasonable assumptions. For example, Goetghebeur and Molenberghs [53], Goetghebeur, Molenberghs, and Katz [54], and Hin and Rubin [55] formulated latent class IV methods for multiple compliance levels in each group; however their additional assumptions are difficult to support. Even in simple situations involving partial compliance, Baker and Kramer [49] and Baker, Frangakis, and Lindeman [56] found the assumptions to be implausible.

To obtain plausible assumptions with partial compliance (Table 7), Baker, Frangakis, and Lindeman [56] proposed using three randomization groups to study the effect of three levels of walking on the probability of Cesarean section for women in labor: no walking ($T0$), one to two hours of walking ($T1$), and at least two hours walking ($T2$). Group 0 is assignment to $T0$; group 1 is assignment to $T1$; and group 2 is assignment to $T2$. The *extended monotonicity* assumption has three parts. First, all participants assigned group 0 receive $T0$, a preference supported by previous studies. Second, a participant who receives $T1$ in group 2

would receive $T1$ in group 1, a consequence of consistent preferences. Third, a participant who receives $T2$ in group 2 would receive $T1$ in group 1, a restriction based on study design. The extended monotonicity assumption yields three latent classes: never-takers ($T0, T0, T0$), partial-takers ($T0, T1, T1$), and full-takers ($T0, T1, T2$). The *extended exclusion restriction* says the probability of outcome does not depend on randomization group for never-takers or partial-takers receiving $T1$. Under these assumptions, Baker, Frangakis, and Lindeman [56] estimated the effect of $T1$ versus $T2$ in full-takers and $T0$ versus $T1$ in a mixture of never-takers and partial-takers.

In a different setting, Cheng and Small [57] also used three randomization groups to extend the latent class IV method to partial compliance; their assumptions led to four latent classes and bounds on the estimated effect of receipt of treatment. As a sensitivity analysis, Shrier et al. [58] analyzed partial compliance by considering both full compliance and no compliance.

9. Surrogate endpoints

A surrogate endpoint is an endpoint observed before the true endpoint that is used to draw conclusions about the effect of treatment on true endpoint. In introducing principal strata, Frangakis and Rubin [14] generalized latent classes based on treatment received to latent classes based on any binary post-randomization variable, most notably surrogate endpoints. Letting $S0$ and $S1$ denote two levels of a binary surrogate endpoint, the principal strata are (surrogate endpoint if randomized to control group, surrogate endpoint if randomized to experimental group), namely $(S0, S0)$, $(S0, S1)$, $(S1, S0)$, $(S1, S1)$. Frangakis and Rubin [14] proposed measuring surrogacy by the effect of randomization group on true endpoint in $\{(S0, S0), (S1, S1)\}$ and $\{(S0, S1), (S1, S0)\}$, an approach later investigators [59-61] extended. In contrast, Baker and Kramer [62] and Baker et al. [63] evaluated surrogacy using the latent class IV method. Although lacking the strong justification of monotonicity and the exclusion restriction with all-or-none compliance, the latent class IV method can play an important role in a sensitivity analysis.

9.1 Cancer prevention trials

In cancer prevention research, investigators typically use small preliminary trials with surrogate endpoints of cancer biomarkers to help decide whether to definitively evaluate the treatment in a large expensive trial with true endpoint of cancer incidence. Investigators draw conclusions based on the following extrapolation: rejecting the null hypothesis of no treatment effect on the surrogate endpoint implies rejecting the null hypothesis of no treatment effect on the true endpoint. The standard assumption underlying this extrapolation is the well-known Prentice Criterion, namely the effect of treatment on true endpoint occurs only through the surrogate endpoint [64]. For a sensitivity analysis, Baker and Kramer [62] proposed the alternative Principal Stratification Criterion, which consists of the exclusion restriction and monotonicity. They found that small deviations from either Criterion can lead to misleading conclusions when extrapolating from small to large trials. Thus the sensitivity analysis with principal stratification reinforces the message of “no free lunch” when using surrogate endpoints in this setting.

9.2 Cancer treatment trials

Often when evaluating cancer treatments via randomized trials, clinicians would like to shorten the trial by using a surrogate endpoint observed before a true endpoint. Suppose there are data from historical randomized trials with the same surrogate and true endpoint as the trial with the new treatment, but with different treatments thought to affect outcome in a similar manner as the new treatment. Investigators would like to predict the effect of the new treatment on the true endpoint using the surrogate endpoint in the new trial and a prediction model fit to surrogate and true endpoints in the historical trials. Based on a sensitivity analysis involving three studies, Baker et al. [63] found the three best performing prediction methods were principal stratification within the latent class IV framework, a mixture model, and a simple linear model.

10. Discussion

The emphasis of this review has been on the basic formulation of the latent class IV method, extensions, assumptions, and applications in obstetrics and cancer research. For a detailed discussion of the choice of instrumental variables or the inclusions of covariates, see Imbens [2], Baiocchi, Cheng, and Small [3], and Garabedian et al. [4] and Neuman et al. [65].

A recurring topic in the field of latent class IV methods is generalizing from treatment effect in compliers to treatment effect in the entire population [35, 45, 66]. Extrapolating from the treatment effect among compliers to the treatment effect in the population is qualitatively similar to extrapolating from the treatment effect in a randomized trial to the treatment effect in the population - a well-accepted challenge [67]. With a meta-analysis of randomized trials with all-or-none compliance, the previously discussed extrapolation method for the paired availability design [35] is useful for generalizability.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health.

References

1. Baker, SG. Compliance, all-or-none. In: Kotz, S.; Read, CR.; Banks, DL., editors. The Encyclopedia of Statistical Science, Update Volume 1. New York: John Wiley and Sons, Inc; 1997. p. 134-138.
2. Newcombe RG. Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur. *Statistics in Medicine*. 1988; 7:1179-1186. [PubMed: 3201044]
3. Shapiro S. Periodic screening for breast cancer: the HIP Randomized Controlled Trial. *Health Insurance Plan. Journal of the National Cancer Institute Monographs*. 1997; 22:27-30. [PubMed: 9709271]
4. Sexton M, Hebel JR. A clinical trial of change in maternal smoking and its effect on birth weight. *Journal of the American Medical Association*. 1984; 251:911-915. [PubMed: 6363731]
5. McDonald CJ, Hui SL, Tierney WM. Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. *MD Computing*. 1992; 9:304-312. [PubMed: 1522792]

6. Baker SG, Lindeman KS. The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine*. 1994; 13:2269–2278. [PubMed: 7846425]
7. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica*. 1994; 62:467–475.
8. Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Statistical Methods in Medical Research*. 2005; 14:369–395. [PubMed: 16178138]
9. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*. 1996; 92:444–455.
10. Barnard J, Frangakis CE, Hill JL, Rubin DB. Principal stratification approach to broken randomized experiments. *Journal of the American Statistical Association*. 2003; 98:299–323.
11. Imbens GW. Instrumental variables: An econometrician's perspective. *Statistical Science*. 2014; 29:323–358.
12. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Statistics in Medicine*. 2014; 33:2297–2340. [PubMed: 24599889]
13. Garabedian LF, Chu P, Toh S, Zaslavsky AM, Soumerai SB. Potential bias of instrumental variable analyses for observational comparative effectiveness research. *Annals of Internal Medicine*. 2014; 161:131–138. [PubMed: 25023252]
14. Frangakis CE, Rubin DB. Principle stratification in causal inference. *Biometrics*. 2002; 58:21–29. [PubMed: 11890317]
15. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*. 2000; 29:722–729. [PubMed: 10922351]
16. Sommer A, Zeger SL. On estimating efficacy from clinical trials. *Statistics in Medicine*. 1991; 10:45–52. [PubMed: 2006355]
17. Connor RJ, Prorok PC, Weed DL. The case-control design and the assessment of the efficacy of cancer screening. *Journal of Clinical Epidemiology*. 1991; 44:1215–21. [PubMed: 1941016]
18. White IR. Uses and limitations of randomization-based efficacy estimators. *Statistical Methods in Medical Research*. 2005; 14:327–347. [PubMed: 16178136]
19. Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology and Therapeutics*. 1995; 57:6–15. [PubMed: 7828382]
20. Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments. *Annals of Statistics*. 1997; 25:305–327.
21. Little R, Yau L. Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods*. 1998; 3:147–159.
22. Zelen M. A new design for randomized clinical trials. *New England Journal of Medicine*. 1979; 300:1242–1245. [PubMed: 431682]
23. Bloom HS. Accounting for no-shows in experimental evaluation designs. *Evaluation Review*. 1984; 8:225–224.
24. Tarwotjo I, Sommer A, West KP Jr, Djunaedi E, Mele L, Hawkins B. Influence of participation on mortality in a randomized trial of vitamin A prophylaxis. *American Journal of Clinical Nutrition*. 1987; 5:1466–1471. [PubMed: 3591726]
25. Permutt T, Hebel R. Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. *Biometrics*. 1989; 45:619–622. [PubMed: 2669989]
26. Robins, JM. The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest, L.; Freeman, H.; Mulley, A., editors. *Health service research methodology: A focus on AIDS*. Washington, DC: U.S. Public Health Service; 1989. p. 113-159.
27. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. *Statistics in Medicine*. 1997; 16:1017–1029. [PubMed: 9160496]
28. Neyman J. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*. 1990; 5:465–472.
29. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*. 1974; 66:688–701.

30. Baker SG, Kramer BS, Lindeman KS. The paired availability design: If you can't randomize, perhaps this applies. *Chance*. 2006; 19:57–60.
31. Cox DR. Discussion. *Statistics in Medicine*. 1998; 17:387–389.
32. Cheng J. Estimation and inference for the causal effect of receiving treatment on a multinomial outcome. *Biometrics*. 2009; 65:96–103. [PubMed: 18373714]
33. Baker SG. Estimation and inference for the causal effect of receiving treatment on a multinomial outcome: An alternative approach. *Biometrics*. 2011; 67:319–25. [PubMed: 20560933]
34. Baker SG. Causal inference, probability theory, and graphical insights. *Statistics in Medicine*. 2013; 32:4319–4330. correction 2014; **33**:1890. [PubMed: 23661231]
35. Baker SG, Lindeman KS. Revisiting a discrepant result: a propensity score analysis, the paired availability design for historical controls, and a meta-analysis of randomized trials. *Journal of Causal Inference*. 2013; 1:51–82. correction 2014; **2**: 113.
36. Baker, SG. The paired availability design: an update. In: Abel, U.; Koch, A., editors. *Nonrandomized Comparative Clinical Studies*. Dusseldorf: Medinform-Verlag; 1998. p. 79-84.
37. Baker SG. Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program. *Journal of the American Statistical Association*. 1998; 93:929–934.
38. Glasziou PP. Meta-analysis adjusting for compliance: the example of screening for breast cancer. *Journal of Clinical Epidemiology*. 1992; 45:1251–1256. [PubMed: 1432006]
39. Baker SG, Kramer BS, Lindeman KS. The randomized registry trial. *New England Journal of Medicine*. 2014; 370:681–682. [PubMed: 24521130]
40. Baker SG, Lindeman KS. Instrumental variable analyses for observational comparative effectiveness research: the paired availability design. *Annals of Internal Medicine*. 2014; 161:840–841. [PubMed: 25437417]
41. Baker SG, Lindeman KS, Kramer BS. The paired availability design for historical controls. *BMC Medical Research Methodology*. 2001; 1:9. [PubMed: 11602018]
42. Baker SG, Lindeman KS. Rethinking historical controls. *Biostatistics*. 2001; 2:383–396. [PubMed: 12933631]
43. Baker SG, Kramer BS, Prorok PC. Comparing cancer mortality rates before-and-after a change in availability of screening in different regions: Extension of the paired availability design. *BMC Medical Research Methodology*. 2004; 4:12. [PubMed: 15149551]
44. Baker SG. Improving the biomarker pipeline to develop and evaluate cancer screening tests. *Journal of the National Cancer Institute*. 2009; 101:1116–1119. [PubMed: 19574417]
45. Baker SG, Lindeman KS, Kramer BS. Clarifying the role of principal stratification in the paired availability design. *International Journal of Biostatistics*. 2011; 7:1.
46. Frangakis CD, Rubin DB. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*. 1999; 86:365–379.
47. Loeys T, Goetghebuer E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics*. 2003; 59:100–105. [PubMed: 12762446]
48. Nie H, Cheng J, Small DS. Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics*. 2011; 67:1397–1405. [PubMed: 21385167]
49. Baker SG, Kramer BS. Simple maximum likelihood estimates of efficacy in randomized trials and before-and-after studies, with implications for meta-analysis. *Statistical Methods in Medical Research*. 2005; 14:349–367. correction 2005;**14**: 605. [PubMed: 16178137]
50. Mealli F, Imbens GW, Ferro S, Biggeri A. Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes. *Biostatistics*. 2004; 5:207–222. [PubMed: 15054026]
51. Baker SG. Analyzing a randomized cancer prevention trial with a missing binary outcome, an auxiliary variable, and all-or-none compliance. *Journal of the American Statistical Association*. 2000; 95:43–50.

52. Frangakis C, Baker SG. Compliance adjusted double-sampling designs for comparative research: estimation and optimal planning. *Biometrics*. 2001; 57:899–908. [PubMed: 11550943]
53. Goetghebeur E, Molenberghs G. Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *Journal of the American Statistical Association*. 1996; 91:928–934.
54. Goetghebeur E, Molenberghs G, Katz J. Estimating the causal effect of compliance on binary outcome in randomized controlled trials. *Statistics in Medicine*. 1998; 17:341–55. [PubMed: 9493258]
55. Jin H, Rubin DB. Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association*. 2008; 103:101–111.
56. Baker SG, Frangakis C, Lindeman KS. Estimating efficacy in a proposed randomized trial with initial and later noncompliance. *Journal of the Royal Statistical Society Series C*. 2007; 56:211–221.
57. Cheng J, Small DS. Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society Series B*. 2006; 68:815–836.
58. Shrier I, Steele RJ, Verhagen E, Herbert R, Riddell CA, Kaufman JS. Beyond intention to treat: what is the right question? *Clinical Trials*. 2014; 11:28–37. [PubMed: 24096636]
59. Gilbert PB, Hudgens MG. Evaluating candidate principal surrogate endpoints. *Biometrics*. 2008; 64:1146–1154. [PubMed: 18363776]
60. Li Y, Taylor JMG, Elliott MR. A Bayesian approach to surrogacy assessment using principal stratification in clinical trials. *Biometrics*. 2010; 66:523–531. [PubMed: 19673864]
61. Zigler CM, Belin TR. A Bayesian approach to improved estimation of causal effect predictiveness for a principal surrogate endpoint. *Biometrics*. 2012; 68:922–932. [PubMed: 22348277]
62. Baker SG, Kramer BS. The risky reliance on small surrogate endpoint studies when planning a large prevention trial. *Journal of the Royal Statistical Society Series A*. 2013; 176:603–608.
63. Baker SG, Sargent DJ, Buyse M, Burzykowski T. Predicting treatment effect from surrogate endpoints and historical trials: an extrapolation involving probabilities of a binary outcome or survival to a specific time. *Biometrics*. 2012; 68:248–257. [PubMed: 21838732]
64. Prentice RL. Surrogate endpoints in clinical trials: Definitions and operational criteria. *Statistics in Medicine*. 1989; 8:431–430. [PubMed: 2727467]
65. Neuman MD, Rosenbaum PR, Ludwig JM, Zubizarreta JR, Silber JH. Anesthesia technique, mortality, and length of stay after hip fracture surgery. *Journal of the American Medical Association*. 2014; 311:2508–2517. [PubMed: 25058085]
66. Swanson SA, Hernán MA. Think globally, act globally: an epidemiologist's perspective on instrumental variable estimation. *Statistical Science*. 2014; 29:371–374. [PubMed: 25580054]
67. Friedman, LM.; Feinberg, CD.; Demets, DL. *Fundamentals of Clinical Trials*. Boston: John Wright; 1981. p. 24–25.

A randomized trial with all-or-none compliance

Fig. 1.

Paired availability design

Fig. 2.

Fig. 3.

Table 1
Assumptions for basic latent class IV method

Setting	Assumptions	
Restricted latent class IV	Exclusion restriction	The probability of outcome in never-takers does not depend on group.
Latent class IV	Exclusion restriction	The probability of outcome does not depend on randomization group in never-takers and always-takers.
	Monotonicity	There are no defiers.

Table 2

Model summary

Latent class			Probability of latent class		Probability $Y=1$ given latent class and assigned treatment		Assumption
	Treatment received if assigned			Treatment received if assigned			
	$T0$	$T1$		$T0$	$T1$		
Never-taker (N)	$T0$	$T0$	π_N	β_N	β_N	Exclusion restriction	
Complier (C)	$T0$	$T1$	π_C	β_{C0}	β_{C1}		
Defier (D)	$T1$	$T0$	0			Monotonicity	
Always-taker (A)	$T1$	$T1$	π_A	β_A	β_A	Exclusion restriction	

Table 3
Assumptions for paired availability design

Stable population	The characteristics of the eligible population related to the probability of outcome do not change over time
Stable ancillary care	Patient management affecting the probability of outcome does not change over time.
Stable disease	The time courses of disease-related events do not change over time in the absence of treatment.
Stable evaluation	Eligibility criteria and definitions of outcome over time do not change over time.
Stable treatment effect	The effect of treatment on the probability of outcome does not change over time among always-receivers and never-receivers (exclusion restriction)
Stable preference	The preference for treatment does not change over time (monotonicity under fixed availability; randomness under random availability)

Table 4
Latent classes under stable preference with fixed availability of $T1$

Treatment preference	Availability of $T1$ at arrival		Latent class
	Time period 0	Time period 1	
$T0$	irrelevant	irrelevant	never-receiver
$T1$	no	no	never-receiver
	no	yes	consistent-receiver
	yes	yes	always-receiver

Table 5
Latent classes under stable preference with random availability of $T1$

Treatment preference	Availability of $T1$ at arrival		Latent class
	Time period 0	Time period 1	
$T0$	irrelevant	irrelevant	never-receiver
$T1$	no	yo	never-receiver
	no	yes	consistent-receiver
	yes	no	inconsistent-receiver
	yes	yes	always-receiver

Table 6
Assumptions for latent class IV method with censored or partially missing binary outcomes

Setting	Assumptions	
Latent class IV with survival outcomes in the presence of death from competing risks	Latent ignorability	The cause-specific hazard for death from competing risks does not depend on an unobserved outcome given the latent class.
	Compound exclusion restriction	The cause-specific hazard rates for the outcome and for death from competing risk do not depend on randomization group in never-takers and always-takers.
	Monotonicity	There are no defiers.
Latent class IV with partially missing binary outcomes	Latent ignorability	The probability of missing in outcome does not depend on outcome given latent class.
	Compound exclusion restriction	The probabilities of outcome and missing in out outcome do not depend on randomization group in never-takers and always-takers.
	Monotonicity	There are no defiers.
Latent class IV with partially missing binary outcomes and auxiliary variable	Latent ignorability	The probability of missing in outcome does not depend on outcome given latent class and auxiliary variable.
	Compound exclusion restriction	The probabilities of outcome, missing in out outcome, and auxiliary variables do not depend on randomization group in never-takers and always-takers.
	Monotonicity	There are no defiers.

Table 7
Assumptions for latent class IV method with three randomization groups

Extended exclusion restriction	The probability of outcome does not depend on randomization group for never-takers.
	The probability of outcome does not depend on randomization group for partial-takers receiving $T1$.
Extended monotonicity	Everyone assigned group 0 receives $T0$.
	A person who receives $T1$ in group 2 would receive $T1$ in group 1.
	A person who receives $T2$ in group 2 would receive $T1$ in group 1.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript