AIDS and Behavior (2021) 25:2441–2454
https://doi.org/10.1007/s10461-021-03207-x

ORIGINAL PAPER

Assessing Potential Outcomes Mediation in HIV Interventions

Heather L. Smyth[1] · Eileen V. Pitpitan[2] · David P. MacKinnon[1] · Robert E. Booth[3]

Accepted: 20 February 2021 / Published online: 19 March 2021

Abstract

Knowledge of causal processes through mediation analysis can help improve the effectiveness and reduce costs of public health programs, like HIV prevention and treatment interventions. Advancements in mediation using the potential outcomes framework provide a method for estimating the causal effect of interventions on outcomes via a mediating variable. The purpose of this paper is to provide practical information about mediation and the potential outcomes framework that can enhance data analysis and causal inference for intervention studies. Causal mediation effects are defined and then estimated using data from an HIV intervention randomized trial among people who inject drugs (PWID) in Ukraine. Results from a potential outcomes mediation analysis show that the intervention had a total causal effect on incident HIV infection such that participants in the experimental group were 36% less likely to become infected during the 12-month study than those in the control arm, but that neither self-efficacy nor network communication mediated this effect. Because neither putative mediator was significant, measurement and confounding issues should be investigated to rule out these mediators. Other putative mediators, such as injection frequency, route of administration, or HIV knowledge can be considered. Future research is underway to examine additional, multiple mediators explaining efficacy of the current intervention and sensitivity to confounding effects.

Resumen

El conocimiento de procesos causales a través del análisis de mediación puede ayudar a mejorar la eficiencia y reducir los costos de programas de salud pública, incluyendo la prevención y el tratamiento del VIH. Avances en mediación, utilizando el enfoque de resultados potenciales ofrece un método para estimar los efectos causales de las intervenciones en variables dependientes a través de variables mediadoras. El objetivo de este artículo es ofrecer información acerca del análisis de mediación y del enfoque de resultados potenciales, el cual permite el análisis y la inferencia causal de las intervenciones. Los efectos causales de mediación son definidos y estimados utilizando los datos de un ensayo clínico con asignación aleatoria para disminuir riego del VIH entre usuarios de drogas inyectables (UDI) en Ucrania. Los resultados del análisis de mediación desde el enfoque de resultados potenciales muestran que la intervención tuvo un efecto causal total en la incidencia de infección del VIH, tal que, durante los 12 meses de estudio, fue 36% menos probable que los participantes del grupo experimental se infectaran en comparación con aquellos en el grupo control. Sin embargo, ni la autoeficacia ni la red comunicación mediaron el efecto. Dado que ninguno de los mediadores resultó ser significativo, sería necesario investigar problemas con la medición y efectos de confusión para poder descartarlos. Otros mediadores podrían ser considerados, tales como frecuencia de la inyección, ruta de administración, o el conocimiento acerca del VIH. Futuras investigaciones podrían estudiar diferentes y múltiples mediadores para explicar la eficacia de esta intervención y realizar un análisis de sensibilidad de efectos de confusión.

Keywords Mediation · Causal mediation · Potential outcomes · HIV intervention · HIV prevention

✉ Heather L. Smyth
Heather.Smyth@asu.edu

[1] Department of Psychology, Arizona State University, PO Box 871104, Tempe, AZ 85287-1104, USA

[2] School of Social Work, San Diego State University, San Diego, CA, USA

[3] Department of Psychiatry, University of Colorado, Denver, CO, USA

Introduction

The field of HIV prevention has undergone significant advances in treatment and intervention approaches. For example, scientists have made great strides in developing and evaluating biomedical primary and secondary prevention strategies (e.g., pre-exposure prophylaxis, or PrEP, antiretroviral treatment). However, health disparities in HIV continue to exist, as key populations and marginalized communities around the world still face significant barriers to accessing (let alone adhering to) such HIV prevention tools [1–3]. Behavioral interventions to prevent HIV remain critical in curbing the epidemic, especially among key populations at higher risk, such as people who inject drugs (PWID) [4, 5].

The purpose of this paper is to provide practical information about data analysis and causal interpretation of mediation effects using a modern mediation analysis approach—the potential outcomes framework [6–8]. As a relatively new analytical tool, our goal is to make this causal mediation analysis accessible to researchers investigating processes involved with HIV prevention. Understanding causal processes related to reducing HIV risk behaviors and disease incidence is important for informing policy decisions and future research. The potential outcomes framework represents a new and more comprehensive approach to estimating the effects of interventions [9] compared to other widely used approaches [10, 11]. Although the potential outcomes framework for causal mediation is widely accepted in epidemiology and biostatistics, it has rarely been applied in intervention analysis [6, 12–16].

Overview of Mediation Analysis

Mediation analysis is used to examine the effects of interventions on behavioral outcomes, specifically, how an intervention influences outcomes through an intermediary variable. Investigating causes and mechanisms is important in HIV intervention research because it can be used to test both theoretical models of behavior [e.g., Theory of Planned Behavior [17, 18], the Information, Motivation, and Behavioral Skills Model [19, 20]] as well as individual components of intervention programs, informing public health programs that are more efficient and effective [21–27]. Behavioral HIV prevention interventions are inherently mediation models designed to reduce HIV by first changing intermediate mechanisms of theory-based psychological or social determinants [e.g. self-efficacy [28]], and subsequently changing behavior or impacting STI/HIV biological outcomes. Consequently, mediated effects would be expected in theory-based behavioral HIV prevention interventions, making mediation analysis a necessary step in their evaluation.

Traditional mediation analysis using the product of path coefficients is conducted by estimating two regression equations, shown below [11]. Equation 1 estimates the X-M association, where the *a* coefficient is the effect of X on M. In Eq. 2, the *c'* coefficient is the effect of X on Y adjusted for M, the *b* coefficient is the effect of M on Y adjusted for X, and the *h* coefficient is the effect of the interaction between X and M. This interaction may be included if there are substantive hypotheses to be tested, but it is often ignored in psychology and other social sciences as treatment effects are assumed to be consistent across control and treatment conditions [29]. However, the XM interaction plays an important role in decomposing effects in the potential outcomes framework [29, 30]. In addition, the XM interaction is important in linking traditional mediation to the potential outcomes framework, and highlighting differences between the two frameworks when mediators and outcomes are not continuous [29, 31]. In Eqs. 1 and 2, i_1 and i_2 represent intercepts and e_1 and e_2 are residuals. The mediated effect is the product of the *a* and *b* coefficients. Standard errors and confidence intervals for the distribution of the product of coefficients can be obtained using the RMediation shiny app allowing for a statistical test for mediation [32].

$$M = i_1 + aX + e_1 \tag{1}$$

$$Y = i_2 + c'X + bM + hXM + e_2 \tag{2}$$

Causal Inference

An important development in mediation analysis is the application of the potential outcomes framework, which considers actual and counterfactual conditions [12, 16], which are defined as combinations of treatment and mediator values that could have been observed even though only some of the potential outcomes are actually observed. Counterfactual conditions are not simply an extension of a repeated measures design where a participant receives different treatments at different times, but rather a case where the same individual simultaneously serves in both the treatment and control condition. Creating an exact duplicate of an individual to serve in both conditions is not possible and therefore we cannot collect data on both observed and counterfactual outcomes at the same time for an individual participant. This is known as "the fundamental problem of causal inference" [33].

However, given a set of identifying assumptions, we can estimate counterfactual mediation conditions using expected values. In order to infer causality using expected values, it is necessary to consider three specific criteria: consistency, positivity, and exchangeability [34]. Consistency requires a well specified intervention in which the treatment is unambiguously defined [35, 36]. When the treatment is well defined, then it becomes clear when differing levels of the treatment are assigned to different potential outcomes. Essentially, the consistency criteria states that the potential outcome at a given level of treatment is equal to the observed outcome of the individual at that given level of treatment. Positivity states that in order to compare two potential outcomes, an individual must have a non-zero probability of assignment to the treatment condition in each of the potential outcomes [37, 38]. Finally, exchangeability states that the potential outcomes among treatment groups are comparable. The exchangeability criteria can be described with four "no unmeasured confounder" assumptions [39]. In all four assumptions, the relations among X, M, and Y may be conditional on confounders such as pretreatment covariates. Two of the confounder assumptions can be satisfied by randomizing the intervention variable, namely that there are no unmeasured confounders of the X–Y relation or the X–M relation. The remaining confounder assumptions, that there are no unmeasured confounders of the M–Y relation, either conditional on treatment, or that are affected by X (called post-treatment confounders), are more difficult to satisfy because although X is randomized it cannot be assumed that M is also randomized. Measures of possible confounders of M to Y are often included in statistical analyses to address these confounding assumptions. Advanced design and statistical approaches are also available to address the non-randomization of M, such as double-randomization designs, inverse probability weighting, G-estimation, and sensitivity analysis [40, 41].

The potential outcomes framework redefines the causal effects of a mediation analysis as the difference between these two potential outcomes. Although there are situations in which traditional and causal estimated effects are the same [29, 31], the differences between traditional and potential outcomes mediation is reflected in the associational versus causal interpretations of the results. One benefit of the potential outcomes framework is the explicit focus on identifying assumptions, which when met, allow for causal interpretations of mediation effects [12, 16, 39]. In addition, this framework is generalizable and easily accommodates non-linear models and outcomes of many data types.

Mediation analysis in the potential outcomes framework results in six causal effects of interest [16]. The formulas for the following causal effects include the nesting of the effect of X on M within the effect of X on Y. Each term represents the outcome, Y, given a certain value of X as well as the mediator, M, given a certain value of X. As can be seen from the formula, the values of X that are associated with Y and M are not always the same. Each term in the formula has the following structure Y(x, M(x)), where capital letters represent a variable, and lower case letters represent specific values of the variable.

The causal indirect effects are the Total Natural Indirect Effect (TNIE) and the Pure Natural Indirect Effect (PNIE). The TNIE is the effect of the intervention on the outcome, conditional on all individuals being in the treatment group, and is computed as the difference between being in the treatment group with a mediator value estimated as the mean for the treatment group versus being in the treatment group with a mediator value estimated as the mean for the control group, as in Eq. 3. The PNIE is the effect of the intervention on the outcome, conditional on all individuals being in the control group, and is computed as the difference between being in the control group with a mediator value estimated as the mean for the treatment group versus being in the control group with a mediator value estimated as the mean for the control group, as in Eq. 4. In both effects, it is the value of the mediator that is manipulated in the formulas, corresponding to the effect that the mediator has on the outcome for the treatment and control groups, respectively. When both the mediator and the outcome are continuous and there is no interaction between X and M, then the TNIE and PNIE are equivalent. If there is an interaction, then the TNIE is equivalent to the simple mediated effect of X on Y through M when X is fixed to the treatment group and the PNIE is equivalent to the simple mediated effect of X on Y through M when X is fixed to the control group [29].

$$\mathrm{TNIE} = \mathrm{E}[\mathrm{Y}(1,\ \mathrm{M}(1)) - \mathrm{Y}(1,\ \mathrm{M}(0))] \tag{3}$$

$$\mathrm{PNIE} = \mathrm{E}[\mathrm{Y}(0,\ \mathrm{M}(1)) - \mathrm{Y}(0,\ \mathrm{M}(0))] \tag{4}$$

Two causal effects that represent the direct effect of an independent variable on an outcome are the Total Natural Direct Effect (TNDE) and the Pure Natural Direct Effect (PNDE). The TNDE is the effect of the intervention on the outcome, conditional on the mediator being fixed to the value that would have been observed had the individual been in the treatment group, which is estimated as the mean value of the mediator for the observations that were actually in the treatment group, as in Eq. 5. The PNDE is the effect of the intervention on the outcome, conditional on the mediator being fixed to the value that would have been observed had the individual been in the control group, which is estimated as the mean value of the mediator for the observations that were actually in the control group, as in Eq. 6. In both the TNDE and PNDE, it is the value of the intervention variable that is manipulated in the formulas, corresponding to the effect an intervention has on an outcome at two different levels of the mediator (i.e., the average mediator value in the treatment group and the control group, respectively.) When both the mediator and the outcome are continuous and there is no interaction between X and M, then the TNDE and PNDE are equivalent. If there is an interaction, the TNDE is equivalent to the simple direct effect of X on Y when M is at the mean value from the treatment group and the PNDE is equivalent to the simple direct effect of X on Y when M is at the mean value from the control group [29].

$$\mathrm{TNDE} = \mathrm{E}[\mathrm{Y}(1,\ \mathrm{M}(1)) - \mathrm{Y}(0,\ \mathrm{M}(1))] \tag{5}$$

$$\mathrm{PNDE} = \mathrm{E}[\mathrm{Y}(1,\ \mathrm{M}(0)) - \mathrm{Y}(0,\ \mathrm{M}(0))] \tag{6}$$

The Controlled Direct Effect (CDE) is the effect of the intervention on the outcome, conditional on a particular value of the mediator, that is, fixing the mediator to a specific value for all individuals (e.g., the mean of the mediator, or a clinically important cutoff), as in Eq. 7. When both the mediator and the outcome are continuous and there is no interaction between X and M, the CDE is equivalent to a simple effect of X on Y computed at a value of the mediator specified by the researcher. Common software packages generally default to the mean value of the mediator [42].

$$\mathrm{CDE}\ =\ \mathrm{E}[\mathrm{Y}(1,\ \mathrm{M}(m)) - \mathrm{Y}(0,\ \mathrm{M}(m))] \tag{7}$$

The Total Effect (TE) is the effect of the intervention on the outcome and is computed as the difference in means between the treatment group and the control group, as in Eq. 8e. When both the mediator and the outcome are continuous and there is no interaction between X and M, this effect is equivalent to the difference between group means. However, when there is non-linearity in the model, such as an XM interaction or categorical outcomes, then the TE and traditional effects may differ [29, 31].

$$\mathrm{TE}\ =\ \mathrm{E}[\mathrm{Y}(1,\ \mathrm{M}(1)) - \mathrm{Y}(0,\ \mathrm{M}(0))] \tag{8}$$

In summary, the TNIE is interpreted as the indirect causal effect of X on Y through M if everyone had been in the treatment group and the PNIE is interpreted as the indirect causal effect of X on Y through M if everyone had been in the control group. The TNDE is interpreted as the causal direct effect of X on Y if all individuals had received treatment and the Pure Natural Direct Effect (PNDE) is interpreted as the causal direct effect of X on Y if all individuals had been in the control condition. Finally, the CDE is interpreted as the causal effect of X on Y at the specified value of the mediator and the TE is interpreted as the causal effect of X on Y.

The potential outcomes framework also produces a quantity known as the *mediated interaction,* which represents the difference between TNIE and PNIE, as well as the difference between TNDE and PNDE [30, 43, 44]. The mediated interaction tests whether the effects in the control group and treatment group are equivalent. Unlike the interaction in a traditional mediation analysis, which tests the simple direct and mediated effects of the control and treatment groups at the same value of the mediator, the mediated interaction instead tests whether the effect at one value of the mediator is equivalent to the effect at a different value of the mediator [29].

The representation of effects in the potential outcomes framework is nonparametric, meaning that the formulas apply for any distribution of mediator or outcome. A strength of the potential outcomes approach is that the formulas apply to both a linear regression model with continuous M and Y, and a logistic regression model with binary M and Y. The different categorical or continuous distributions of M and Y have the same potential outcomes formula, however different methods are used to estimate the potential outcomes (e.g. linear regression for continuous measures and logistic regression for binary measures.)

Empirical Demonstration

The following demonstration uses data from a recent randomized trial testing an HIV prevention intervention for PWID in Ukraine [45]. In this trial, PWID from three cities were randomized into a peer-leader social network intervention or a control condition. The main outcome paper described how PWID in the intervention arm had significantly lower HIV incidence than PWID in the control arm [45]. However, more can be learned about how, or through what mediating mechanism, the intervention reduced HIV incidence. The purpose of this paper is to illustrate how to conduct a mediation analysis within the potential outcomes framework, while testing two possible mediators that may explain lower HIV incidence as a result of the intervention with PWID in Ukraine. The first mediator, self-efficacy to practice safer behavior, is common in HIV prevention and intervention research [46]. The second, communication with network members, is a putative mechanism specific to the intervention because the intervention was designed to promote communication about safer behavior among network members in the experimental condition.

Methods

Overview

The data for the empirical example comes from a randomized trial of a social network intervention to affect HIV incidence among PWID in the Ukraine [45]. In this demonstration, some of the data have been simplified, for example, geographic clustering has been ignored. In addition, self-efficacy, a common variable for trials using an individual approach rather than a social network approach to behavior change, has been specified as a mediator in one of the examples, while the putative mediator, communication with network members by peer leaders, is specified as the mediator in the second example. Finally, in the original analysis by Booth et al., (2016), incident HIV infection was predicted using a Cox proportional hazards model [45], however, survival analysis in the potential outcomes framework is outside the scope of the current paper [47].

Measures

Self-efficacy was measured as the sum of eight items that measured the degree of respondent's agreement. The items were measured as an index where 1 = disagree and 2 = agree. The range of responses in this data set were 0–2, with 0 representing missing data that were not included in the sum total. Example items include, "You would feel uncomfortable talking to a sex partner about using condoms;" and "If someone 'new' offered you their used syringe, you would not use it without cleaning." The KR-20 internal reliability coefficient was 0.46, which suggests low reliability. Poor reliability in a mediator could lead to underestimation of the mediated effect [48]. Several methods to account for the unreliability of the measure were considered, including a synthetic reliability adjustment, modeling the mediator as a latent variable, and removing scale items that showed low correlations with the sum score. In each case, the final conclusions regarding the mediated effect remained unchanged and each method introduced new statistical limitations to the analysis, which are discussed in the limitations section.

Network communication was measured as the sum of 15 items that each counted the number of times the respondent talked with their primary or secondary network members about risk behaviors related to sexual activity or drug use. Example items include, "Since your last interview, how many times did you talk with your network member about how HIV can be spread from person to person?" and "Since your last interview, how many times have you talked with your network member about the problems that make it hard for them to use safer injection practices?" Cronbach's alpha was 0.87, suggesting good reliability. Findings from the self-efficacy and network communication examples are primarily intended for illustrative purposes, thus inferences should be made with caution.

Data Analysis

The original trial included 1200 PWID (25.1% female) who were HIV negative at baseline and randomized into intervention and control groups. To ensure temporal precedence of the mediator to outcome relation, the current analyses include only those participants who were also HIV negative at the 6-month follow-up, leaving $n = 927$ cases used in this example. There were 15 cases that were missing data on the self-efficacy baseline covariate, and 4 additional cases that were missing data on the self-efficacy mediator, which is a 2.1% rate of missing responses, leading to an analyzed sample of $n = 908$ for the first model. There were 65 cases with missing data on the network communication mediator, which is a 7.0% rate of missing responses, leading to an analyzed sample size of $n = 862$ for the second model. The first model included a binary intervention variable (X; 0 = control, 1 = intervention), self-efficacy score measured at 6 months (time 2; M), and the occurrence of incident HIV infection at some point between the 6- and 12-month follow-up (Y; 0 = no infection, 1 = HIV infection). The self-efficacy sum score at baseline was included as a covariate to control for individual differences in self-efficacy that are unrelated to the intervention. A linear regression model for the mediator is specified in Eq. 9 that includes the intervention and baseline self-efficacy. The logistic regression model for incident HIV infection is shown in Eq. 10 and includes the intervention, self-efficacy at time 2, the treatment by mediator interaction, and baseline self-efficacy. Note that the XM interaction has been included in the models. Typically, if this interaction is not significant it would be left out of a final model. However, in the potential outcomes framework, it is advisable to retain the interaction as it is a part of the decompositions of the total effect, as described by VanderWeele [30]. Figure 1 shows the mediation model as a direct acyclic graph.

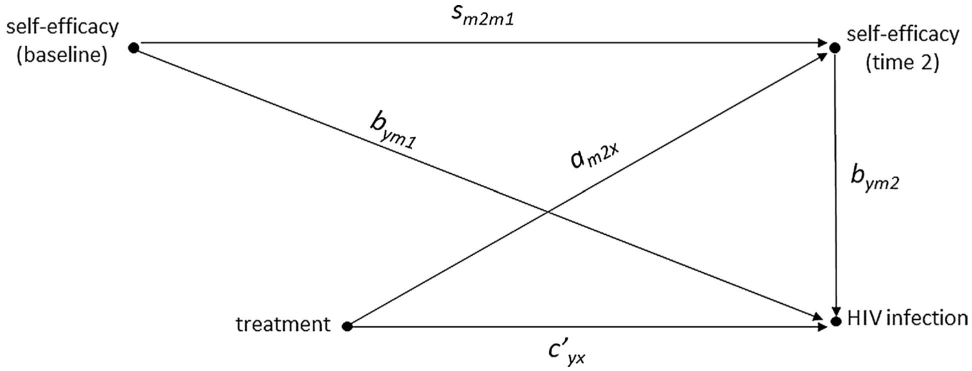

Fig. 1 Direct acyclic graph with self-efficacy at time 2 as a mediator and self-efficacy at baseline as a covariate

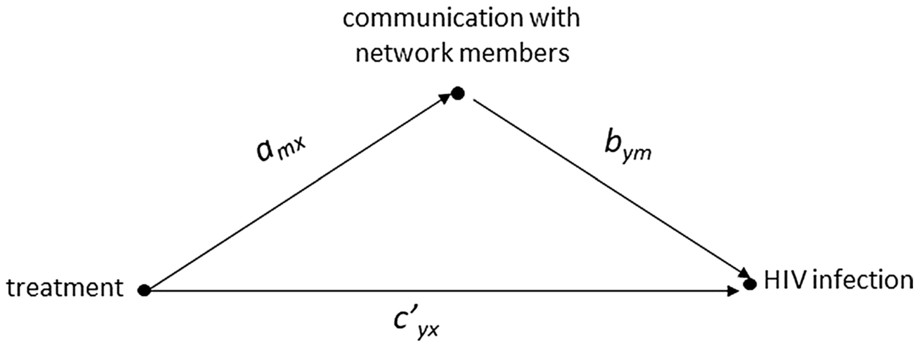

Fig. 2 Direct acyclic graph with network communication as a mediator

$$\widehat{self - eff}. = i_m + a_{m2x}(intervention) + s_{m2m1}(s.e.baseline) \tag{9}$$

$$logit(Pr(\widehat{HIVinfection} = 1)) = i_y + c'_{yx}(intervention) + b_{ym2}(self - eff) + b_{ym1}(s.e.baseline) + h(intervention * self - eff) \tag{10}$$

The second model included the same intervention and HIV status variables, as well as a composite variable that represents communication among network members. Equations 11 and 12 show the linear regression models for the network communication mediator, and the logistic regression model for incident HIV infection outcome, respectively. Figure 2 shows the mediation model as a direct acyclic graph.

$$\widehat{communication} = i_m + a_{mx}(intervention) \tag{11}$$

$$logit(PR(\widehat{HIVinfection} = 1)) = i_y + c'_{yx}(intervention) + b_{ym}(communication) + h(intervention * communication) \tag{12}$$

Data were analyzed with PROC CAUSALMED in SAS 9.4, a program for estimating causal mediation effects. The treatment and outcome variables were listed as categorical variables using the class statement with the descending option specified in order to model the probability of incident HIV infection (y = 1). Incident HIV infection was modeled using a binomial distribution and log link function. The log link function was specified because HIV incidence in the sample was 21.67% and would not be considered a rare event. By default, the estimation procedures in PROC CAUSALMED assumes that the binary outcome is rare and uses logistic regression to estimate Y [16, 30]. The outcome model included the treatment by mediator interaction and baseline self-efficacy as a covariate. Reported effects were

Table 1 Excess relative risk and regression parameters for self-efficacy mediator

Effect	Estimate	Bootstrap	Percentile Bootstrap		*Z*	*Pr*>|Z|
		Standard Error	95% Confidence Limits			
TE	−0.35	0.13	−0.56	−0.05	−2.65	0.01
CDE	−0.34	0.13	−0.56	−0.05	−2.57	0.01
TNDE	−0.34	0.13	−0.56	−0.05	−2.60	0.01
PNDE	−0.35	0.13	−0.56	−0.06	−2.64	0.01
TNIE	−0.00	0.01	−0.02	0.02	−0.08	0.94
PNIE	−0.01	0.01	−0.03	0.02	−0.61	0.54
Mediated Interaction	0.01	0.01	−0.02	0.04	0.48	0.63
Parameter[b]	Estimate	Standard Error	Wald 95%		*Wald* χ^2	*Pr*>|Z|
			Confidence Limits			
a	0.11	0.12	−0.13	0.34	0.81	0.37
b	−0.05	0.06	−018	0.07	0.68	0.41

[a] Effect labels correspond to the following effects in PROC Causalmed output: *TNDE* Total Direct, *PNDE* Natural Direct, *TNIE* Natural Indirect, *PNIE* Pure Indirect

[b] The regression coefficient for the *a*-path shows whether the action theory led to a significant effect, while the *b*-path shows whether the conceptual theory led to a significant effect

Table 2 Excess relative risk and regression parameters for network communication mediator

Effect[a]	Estimate	Bootstrap	Percentile Bootstrap		*Z*	*Pr*>|Z|
		Standard Error	95% Confidence Limits			
TE	−0.40	0.13	−0.64	−0.12	−3.14	<.00
CDE	−0.34	0.15	−0.56	0.05	−2.34	0.02
TNDE	−0.32	0.13	−0.55	−0.03	−2.39	0.02
PNDE	−0.42	0.13	−0.66	−0.14	−3.38	<.00
TNIE	0.02	0.03	−0.04	0.07	1.19	0.23
PNIE	−0.08	0.08	−0.29	0.03	−1.13	0.26
Mediated Interaction	0.10	0.09	−0.03	0.32	1.41	0.16
Parameter[b]	Estimate	Standard Error	Wald 95%		*Wald* χ^2	*Pr*>|Z|
			Confidence Limits			
a	6.51	1.48	3.61	9.41	19.31	<.00
b	−0.01	.01	−0.03	0.01	1.26	0.26

[a] Effect labels correspond to the following effects in PROC Causalmed output: *TNDE* Total Direct, *PNDE* Natural Direct, *TNIE* Natural Indirect, *PNIE* Pure Indirect

[b] The regression coefficient for the *a*-path shows whether the action theory led to a significant effect, while the *b*-path shows whether the conceptual theory led to a significant effect

estimated at the sample means of the mediator and covariate, which are the default software settings. The percentile bootstrap procedure was used to compute 95% confidence intervals for the causal effects using 1,000 bootstrap replications.

Results

The empirical example investigated the effect of a social network intervention on incident HIV infection through self-efficacy related to safer behavior practices, and through communication between the peer leader and social network members. In the first example, self-efficacy scores at baseline were included as a covariate and effects were estimated at the sample mean of this score. When the outcome is binary PROC CAUSALMED presents results for the six causal effects on the excess relative risk scale. Results for a model with self-efficacy as the mediator are reported first, and then for a model with communication as the mediator. Tables 1 and 2 summarize the excess relative risk estimates and regression coefficients for self-efficacy and network communication, respectively. The tables are an example for researchers who wish to present causal mediation effects in a tabular form.

Model with Self-Efficacy Mediator

The total effect of the intervention on HIV conversion with self-efficacy in the model was significant, with TE = −0.35, 95% CI [−0.56, −0.06], $z = -2.65$, $p = 0.01$. There was a total reduced risk of 35% due to the intervention for those in the treatment group with mediator values equal to the treatment mean compared to those in the control group with self-efficacy values equal to the control group mean.

The controlled direct effect of the intervention on incident HIV infection calculated at the sample mean of the mediator, self-efficacy, was significant, with CDE = −0.34, 95% CI [−0.56, −0.05], $z = -2.57$, $p = 0.01$. There was a 34% reduced risk of incident HIV infection due to the intervention among those with mean mediator scores and mean self-efficacy at baseline; if everyone had self-efficacy values equal to the grand mean, then those in the treatment group would be 34% less likely to become infected than those in the control group.

The direct effect of the intervention on incident HIV infection, had everyone's mediator value been equal to the mean self-efficacy score in the treatment group, was significant, with TNDE = −0.34, 95% CI [−0.56, −0.05], $z = -2.60$, $p = 0.01$. The TNDE can be interpreted as a reduced risk of incident HIV infection of 34% due to the intervention among those with mediator scores equal to the treatment group mean and grand mean scores on self-efficacy at baseline; if everyone had mediator scores equal to the treatment group mean, those in the treatment group would be 34% less likely to become infected compared to those in the control group.

The direct effect of the intervention on infection, had everyone's mediator value been equal to the mean self-efficacy score in the control group, was significant, with PNDE = −0.35, 95% CI [−0.56, −0.06], $z = -2.64$, $p = 0.01$. There was a reduced risk of incident HIV infection of 35% due to the intervention among those with mediator scores equal to the control group mean and grand mean scores on self-efficacy at baseline; if everyone had mediator scores equal to the control group mean, those in the treatment group would be 35% less likely to become infected compared to those in the control group.

The indirect effect of the intervention on incident HIV infection through self-efficacy, had everyone been in the treatment group, was not significant, with TNIE = – 0.00, 95% CI [– 0.02, 0.02], z = – 0.08, p = 0.94. For illustrative purposes, the TNIE can be interpreted as a reduced risk of incident HIV infection of < 1% due to the intervention's effect on self-efficacy among the treated with mean scores on self-efficacy at baseline; if everyone had been in the intervention, those who had mediator values at the treatment group mean would be < 1% less likely to become infected than those whose mediator value was at the control group mean.

The indirect effect of the intervention on infection through self-efficacy, had everyone been in the control group, was also not significant, with PNIE = – 0.01, 95% CI [– 0.03, 0.02], z = – 0.61, p = 0.54. For illustrative purposes, there was a reduced risk of incident HIV infection of 1% through the intervention's effect on self-efficacy among those in the control group with mean scores on self-efficacy at baseline. If everyone had been in the control group, those who had mediator values at the treatment group mean would be 1% less likely to become infected than those whose mediator value was near the control group mean.

The mediated interaction of the intervention on incident HIV infection was not significant which suggests the indirect effects of self-efficacy in the treatment and control groups do not differ, MI = 0.01, 95% CI [– 0.02, 0.04], z = 0.48, p = 0.63. Both the TNIE and PNIE in the example data were non-significant and they were not significantly different.

Model with Network Communication Mediator

The total effect of the intervention on incident HIV infection with communication with network members in the model was also significant, with TE = – 0.40, 95% CI [– 0.64, – 0.12], z = – 3.14, p = 0.00. There was a total reduced risk of 40% due to the intervention for those in the treatment group with mediator values equal to the treatment mean compared to those in the control group with communication scores equal to the control group mean.

The controlled direct effect of the intervention on incident HIV infection calculated at the sample mean of the mediator, communication with network members, was significant according to the normal theory p-value, with CDE = – 0.34, 95% CI [– 0.56, 0.05], z = – 2.34, p = 0.02. The discrepancy between the confidence limits and the p-value can be attributed to the difference between assessing significance using a bootstrap method versus a normal theory method. While the confidence limits do contain zero, they do not drastically differ from the conclusion using the z-test. There was a 34% reduced risk of incident HIV infection due to the intervention among those with mean mediator scores; if everyone had communication scores equal to the grand mean, then those in the treatment group would be 34% less likely to become infected than those in the control group.

The direct effect of the intervention on incident HIV infection, had everyone's mediator value been equal to the mean communication score in the treatment group, was significant, with TNDE =– 0.32, 95% CI [– 0.55, – 0.03], z = – 2.39, p = 0.02. The TNDE can be interpreted as a reduced risk of incident HIV infection of 32% due to the intervention among those with mediator scores equal to the treatment group mean; if everyone had mediator scores equal to the treatment group mean, those in the treatment group would be 32% less likely to become infected compared to those in the control group.

The direct effect of the intervention on infection, had everyone's mediator value been equal to the mean communication score in the control group, was also significant, with PNDE =– 0.42, 95% CI [– 0.66, – 0.14], z = -3.38, p = 0.00. There was a reduced risk of incident HIV infection of 42% due to the intervention among those with mediator scores equal to the control group mean; if everyone had mediator scores equal to the control group mean, those in the treatment group would be 42% less likely to become infected compared to those in the control group.

The indirect effect of the intervention on incident HIV infection through communication, had everyone been in the treatment group, was not significant, with TNIE = 0.02, 95% CI [– 0.04, 0.07], z = 1.19, p = 0.24. For the purpose of illustration, the TNIE can be interpreted as a reduced risk of incident HIV infection of 2% due to the intervention's effect on communication among the treated. In other words, if everyone had been in the intervention, those who had mediator values at the treatment group mean would be 2% less likely to become infected than those whose mediator value was at the control group mean.

The indirect effect of the intervention on infection through communication, had everyone been in the control group, was also not significant, with PNIE =– 0.08, 95% CI [– 0.29, 0.03], z = – 1.13, p = 0.26. For illustrative purposes, there was a reduced risk of incident HIV infection of 8% through the intervention's effect on communication among those in the control group; if everyone had been in the control group, those who had mediator values around the treatment group mean would be 8% less likely to become infected than those whose mediator value was near the control group mean.

The mediated interaction of the intervention on incident HIV infection was not significant which suggests the indirect effects of network communication in the treatment and control groups do not differ, MI = 0.10, 95% CI [-0.03, 0.32], $z = 1.41$, $p = 0.16$. Both the TNIE and PNIE in the example data were non-significant and they were not significantly different.

In addition to the six causal effects, PROC CAUSALMED also gives Wald X^2 tests of the regression coefficients, which allows for evaluation of the action and conceptual theories behind the intervention. The effect of the intervention on network communication was significant $X^2(1, n = 862) = 19.31$, $p < 0.00$, 95% Wald CI [3.61, 9.41]. However, neither the effect of the intervention on self-efficacy while controlling for baseline self-efficacy, nor the effects of either mediator on incident HIV infection were significant. These results show support for action theory that the intervention increases network communication, but not self-efficacy. Conceptual theories for the effects of self-efficacy and network communication on incident HIV infection by themselves are not supported. However, it would be expected that the conceptual theory that links the mediator to the outcome consists of a mediation chain, where the tested mediators cause changes in risk behaviors, which has a more direct influence on incident HIV infection.

Overall, the results for this demonstration reflect the findings from the original study that the intervention resulted in reduced risk of incident HIV infection [45]. Using the potential outcomes framework, the intervention had a positive causal effect on incident HIV infection such that individuals in the treatment group had a reduced risk of infection during the last 6 months of the study. However, potential outcomes mediation analysis shows the intervention did not have a causal effect on HIV through self-efficacy or through network communication.

Discussion

In this paper, we illustrate how to use the potential outcomes framework to conduct a mediation analysis with example data from an HIV prevention intervention for PWID in Ukraine. Results show that neither self-efficacy to practice safer behaviors nor communication with network members causally mediate the effect of the intervention on HIV incidence. However, the intervention did have significant direct effects, which suggests that there are other possible mediating processes. Indeed, the intervention was based on multiple theories [45], and therefore a number of other variables may serve as significant mediators (which is beyond the goal and scope of this paper).

One of the greatest strengths of mediation analysis is that it directly tests the theories upon which prevention programs are based [49, 50]. The theory-based variables that are shown to be significant mediators explaining intervention efficacy lend support to the theory, whereas non-significant theory-based hypothesized mediators suggest the need for improved measurement, further testing, or theory re-development [50]. Chen [48] described how mediation analysis informs and evaluates action theory and conceptual theory. Action theory is the theory relating the intervention components to the mediators targeted by the intervention, and conceptual theory is the theory relating mediating variables to the outcome variable [49]. In the examples above, the intervention had significant effects on the network communication mediator in the regression models, but the mediators did not have significant effects on the outcome. This suggests a failure of conceptual theory. Evaluating action and conceptual theory increases understanding of the mechanisms of behavior change and identifies the most effective components of an intervention. If an intervention fails to significantly change targeted mediators, the intervention may have been ineffective, or measures of the mediators may have been inadequate. If the mediating variables show no relationship with the outcome, the mediators may not cause the outcome, confounder bias may have occurred, or measurement may be inadequate. It is also possible that a more complex sequential mediation model is necessary to capture the mediation process. In this example, additional mediators that measure risk behaviors, such as receptive needle sharing or sharing injection equipment could help explain the link between self-efficacy or network communication and incident HIV infection.

The results from the examples in this paper support the original research showing that the HIV intervention had a positive effect in reducing incident HIV infections. However, we are left with a question of how the intervention had its effect on incident HIV infections. Identifying the mechanism at work can help future intervention programs focus on the most important aspects of the intervention. Although the proposed mediators for the examples in this paper were not significant, it does not suggest that a mediator does not exist. Rather, it is possible that either there is another aspect of the intervention that is responsible for transmitting its effect to the outcome, or there are modeling concerns that limit our ability to detect the mediational process of the proposed mediators.

The non-significant TNIE and PNIE of self-efficacy and network communication on incident HIV infection may also be a result of unmeasured confounding variables. Unmeasured confounders can obscure the true relations among the independent variable, the mediator, and the outcome. In most studies, it is not feasible to measure every potential confounder in order to control for its effect in a model, however, sensitivity analysis can provide evidence for the likelihood that an unmeasured confounder is exerting a substantial effect on the variables that are included in the model.

The current social network intervention for PWID in Ukraine was based on theories including Social Cognitive Theory [51], Theory of Reasoned Action [52], and Social Diffusion Theory [53, 54]. Similar to other social network interventions, the current intervention aimed to promote changes among the networks of PWID; for example, by changing social norms around injection drug use, and communication among network members, specifically in conversations around injection practices [55–57]. This is in contrast to interventions that solely target individuals, and not their social context, such as interventions aimed at improving an individual's HIV knowledge, motivation to practice safer behavior, behavioral skills [58], or self-efficacy to practice safer behavior. Other possible mediators may operate more at the network level, including perceived norms around injection risk behaviors, or communication among network members about practicing safe injection. Thus, we sought to test both self-efficacy to practice safer behavior and network communication as two conceptually different potential mediators. As the primary purpose of this paper was to describe how to conduct mediation analysis using the potential outcomes framework, we believed that analyzing these two simple mediators would be most pragmatic and easiest to interpret. Future research is underway to examine additional, multiple mediators explaining efficacy of the current intervention.

Additional hypothesized mediators can be added to the single mediator model and tested simultaneously with a multiple mediator potential outcomes model. The potential outcomes framework for multiple mediators introduces many new counterfactuals that must be considered because of the increased number of variables and possible relations among the mediators themselves [59]. Potential outcomes analysis of multiple mediators is an active area of research that can increase the information available about causal mechanisms of interventions [60–71]. However, there are a limited number of software programs that can accommodate multiple mediators in a causal mediation analysis, including the *mediation* and *medflex* packages in R [42, 72, 73].

Limitations

The low reliability of the self-efficacy mediator is a limitation in this example. In this study, self-efficacy was operationalized equally in terms of *both* injection and sexual risk behavior. Results may differ when the two risk domains are separated. Self-efficacy scores were the sum of eight binary items, however there were varying degrees of missingness on each of the eight items. The sum scores treated missing values as zero, which could introduce an excess of measurement error. Other measurement options could include the imputation of missing data, IRT scores, or average scores for the items that did have responses.

In traditional mediation analysis, unreliability in the mediator can result in underestimation of the mediator's effect on the outcome, and overestimation of the independent variable's effect, controlling for the mediator [48, 59]. Specifying a latent variable for the mediators is a preferred method for handling measurement unreliability in the mediator, and if this is not feasible, then adjustments for unreliability could be made using a manifest variable model where estimates of reliability from previous research are used to adjust for unreliability in the study [59].

Several methods to account for the unreliability in the self-efficacy mediator were investigated, but each introduced new limitations to the analysis while leading to the same final conclusions concerning the mediated effect. First, self-efficacy was modeled as a latent variable. Estimating latent mediators is relatively new in the potential outcomes framework and there is still work to be done to include the XM interaction in software programs, and in verifying the estimation and interpretation of causal effects. This work is beyond the scope of the current paper. Next, a synthetic reliability adjustment was attempted. This method involves estimating the mediator as a latent variable with the sum score as a single indicator. As such, the synthetic adjustment is also limited in the ability to include the XM interaction in current software packages. Finally, an EFA was estimated which showed that three items had correlations with the sum total that were near zero, and when these three items were removed, the KR-20 internal reliability coefficient increased to 0.59. However, two of the three items were related to drug use behavior, and removing them from the scale would create an imbalance between items related to drug use and sexual activity, which may lead to inappropriate inferences about the mediating effect.

Conclusions

A primary purpose of this paper is to introduce and illustrate how to conduct a counterfactual causal mediation analysis in the realm of HIV intervention research. An HIV intervention for PWID in Ukraine was used as an empirical example. Self-efficacy and network communication were not found to be mediating mechanisms in this analysis, however future research will continue to investigate possible mechanisms. The method described in this paper can be used to evaluate the causal effects of interventions with a variety of outcomes types, including binary outcomes as demonstrated in the example. Next steps include investigating both parallel and sequential mediator models, testing latent variable mediation models in the potential outcomes framework, conducting psychometric analysis and adjustments to improve reliability of measured mediators and outcomes, sensitivity analysis, and conducting a survival analysis in the potential outcomes framework.

The potential outcomes framework is a major advancement for mediation analysis with practical benefits to HIV prevention research. HIV primary prevention scientists have spent over 30 years developing and evaluating interventions designed to change behavior. Meta-analyses have described the efficacy of behavioral interventions to reduce risky behavior for various populations (e.g. PWID, Black women, heterosexual couples), regions (e.g. Asia, Latin America), and types of interventions (e.g. peer-education, brief single-session) [74–79]. Across meta-analyses, behavioral interventions were consistently effective at changing behavior (e.g. condomless sex), but were less consistently effective for STI/HIV biological outcomes, including HIV incidence [80, 81]. The variability and inconsistency demonstrated by behavioral interventions might be perceived as a significant challenge to HIV prevention, and yet, the differences between studies present a unique opportunity to disentangle the heterogeneity of efficacy findings in HIV prevention trials. Identifying mediators through which effective programs influence behavioral outcomes is an important step toward increasing effectiveness in preventing HIV. Mediation analysis from the potential outcomes framework can be a useful tool to help shed light on the causal mechanisms underlying intervention efficacy, informing why some interventions are efficacious, and why some are not.

Estimating causal mediation effects increases the information that can be drawn from intervention study data and clarifies causal interpretations of those effects. A pragmatic benefit of potential outcomes mediation is an emphasis on model assumptions and strategies for handling assumption violations, such as how confounder bias influences causal inferences [40, 59, 82–84]. With this focus, researchers can better address the theoretical questions at the heart of an intervention program. Investigating causation through potential outcomes mediation can lead to better decisions during program design and implementation, and ultimately more effective and efficient public health programs.

Appendix

PROC CAUSALMED Program

```
/*The following program performs a causal mediation analysis*/
/*X is treatment (0=control, 1=treatment), M is a continuous mediator measured at time 2, Y is a binary outcome (0=no disease, 1=disease), and C is the baseline measurement of the mediator as a covariate*/
/*The class statement specifies the two categorical variables. The descending option is used to predict the probability Y=1*/
/*A binary distribution for Y and the log link function are specified in the model statement. The log link is used because the outcome is not rare.*/
/*bootstrap confidence intervals are specified using 1000 bootstraps and a specified random seed.*/
Title 'Single mediator with baseline covariate';
proc causalmed data=use.data pall alpha = .05;
class X Y/descending;
model Y = X|M / dist=bin link=log;
mediator M = X;
covar C;
bootstrap CI (all) nboot = 1000 seed = 08012019;
run;
quit;
```

Acknowledgements This research was supported in part by the National Institute on Drug Abuse (NIDA; R37DA09757) and (R01DA042666). Data are from a NIDA-funded randomized trial (R01DA026739). We acknowledge the staff and directors of the organization who participated in the original intervention and data collection, including Dimitry Kryzhko with Health of Nation in Makeyevka/Donetsk; Olga Kostyuk and Tatiana Semikop with Faith, Hope and Love in Odessa; and Elena Goryacheva with the Charity Foundation Vykhod in Nikolayev. We would also like to thank the participants who gave their time for the project.

Authors' Contributions Heather Smyth was responsible for conducting the statistical analysis. David MacKinnon and Eileen Pitpitan assisted with the analysis and data interpretation. Heather Smyth, Eileen Pitpitan, and David MacKinnon wrote the initial draft of the manuscript. Robert Booth secured funding for and led the original intervention trial. All authors contributed to and approved the final manuscript.

Funding This research was supported in part by the National Institute on Drug Abuse (NIDA; R37DA09757) and (R01DA042666). Data are from a NIDA-funded randomized trial (R01DA026739).

Data Availability N/A

Code Availability A SAS program for analysis is available in the appendix, or by contacting the first author.

Declarations

Conflict of interest The authors declare they have no conflict of interest.

Ethical Approval This study represents a secondary data analysis of a study, including informed consent, that was approved by the Colorado Multiple Institutional Review Board at the University of Colorado Denver and by the Ukrainian Institute on Public Health Policy. As a secondary analysis of de-identified data, the current analysis was not deemed as Human Subjects research by the Institutional Review Board at San Diego State University.

Consent to Participate N/A

Consent for Publication N/A

References

1. Risher K, Mayer KH, Beyrer C. HIV treatment cascade in MSM, people who inject drugs, and sex workers. Curr Opin HIV AIDS. 2015;10(6):420–9.
2. Macdonald V, Verster A, Baggaley R. A call for differentiated approaches to delivering HIV services to key populations. J Int AIDS Soc. 2017;20(S4):28–31.
3. Lall P, Lim SH, Khairuddin N, Kamarulzaman A. Review: An urgent need for research on factors impacting adherence to and retention in care among HIV-positive youth and adolescents from key populations. J Int AIDS Soc. 2015;18(2S1):41–53.
4. Mathers BM, Degenhardt L, Phillips B, Wiessing L, Hickman M, Strathdee SA, *et al*. Global epidemiology of injecting drug use and HIV among people who inject drugs: A systematic review. The Lancet. 2008;372(9651):1733–45.
5. Degenhardt L, Mathers B, Vickerman P, Rhodes T, Latkin C, Hickman M. Prevention of HIV infection for people who inject drugs: Why individual, structural, and combination approaches are needed. The Lancet. 2010;376(9737):285–301.
6. Pearl J. Interpretation and identification of causal mediation. Psychol Methods. 2014;19(4):459.
7. Pearl J. The causal mediation formula: A guide to the assessment of pathways and mechanisms. Prev Sci. 2012;13(4):426–36.
8. VanderWeele TJ. Mediation and mechanism. Eur J Epidemiol. 2009;24(5):217–24.
9. Hernàn MA, Robins JM. Causal Inference. Boca Raton: Chapman & Hall/CRC; 2019. (**forthcoming**).
10. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986;51(6):1173.
11. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychol Methods. 2002;7(1):83–104.
12. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010;15(4):309.
13. Jo B. Causal inference in randomized experiments with mediational processes. Psychol Methods. 2008;13(4):314.
14. Liu W, Kuramoto SJ, Stuart EA. An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. Prev Sci. 2013;14(6):570–80.
15. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. Prev Sci. 2015;16(3):475–85.
16. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods. 2013;18(2):137–50.
17. Ajzen I, Fishbein M. Understanding attitudes and predicting social behavior. New Jersey: Prentice-Hall; 1980.
18. Albarracin D, Johnson BT, Fishbein M, Muellerleile PA. Theories of reasoned action and planned behavior as models of condom use: a meta-analysis. Psychol Bull. 2001;127(1):142.
19. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull. 1992;111(3):455.
20. Fisher JD, Fisher WA, Williams SS, Malloy TE. Empirical tests of an information-motivation-behavioral skills model of AIDS-preventive behavior with gay men and heterosexual university students. Health Psychol. 1994;13(3):238.
21. Bryan A, Schmiege SJ, Broaddus MR. Mediational analysis in HIV/AIDS research: Estimating multivariate path analytic models in a structural equation modeling framework. AIDS Behav. 2007;11(3):365–83.
22. O'Rourke HP, MacKinnon DP. Reasons for testing mediation in the absence of an intervention effect: A research imperative in prevention and intervention research. J Stud Alcohol Drugs. 2018;79(2):171–81.
23. O'Rourke HP, MacKinnon DP. The importance of mediation analysis in substance-use prevention. Prevention of Substance Use: Springer; 2019. p. 233–46.
24. Kisbu-Sakarya Y, MacKinnon DP, O'Rourke HP. Statistical models of mediation for drug program evaluation. In: Scheier LM, editor. Handbook of adolescent drug use prevention: Research, intervention strategies, and practice. Washington: American Psychological Association; 2015. p. 459–78.
25. Hardnett FP, Pals SL, Borkowf CB, Parsons J, Gomez C, O'Leary A. Assessing mediation in HIV intervention studies. Public Health Rep. 2009;124(2):288–94.
26. Pitpitan EV, Kalichman SC, Garcia RL, Cain D, Eaton LA, Simbayi LC. Mediators of behavior change resulting from a sexual risk reduction intervention for STI patients, Cape Town, South Africa. J Behav Med. 2015;38(2):194–203.
27. Pitpitan EV, Patterson TL, Abramovitz D, Vera A, Martinez G, Staines H, *et al*. Policing behaviors, safe injection self-efficacy, and intervening on injection risks: Moderated mediation results from a randomized trial. Health Psychol. 2016;35(1):87–91.
28. Bandura A. Self-Efficacy: The Exercise of Control. New York: Worth Publishers; 1997.
29. MacKinnon DP, Valente MJ, Gonzales O. The correspondence between causal and traditional mediation analysis: The link is the mediator by treatment interaction. Prev Sci. 2020;21(2):147–57.
30. VanderWeele TJ. A unification of mediation and interaction: a four-way decomposition. Epidemiology (Cambridge, Mass). 2014;25(5):749–61.
31. Rijnhart JJM, Valente MJ, MacKinnon DP, Twisk JWR, Heymans MW. The use of traditional and causal estimators for mediation models with a binary outcome and exposure-mediator interaction. Structural Equation Modeling: A Multidisciplinary Journal. 2018. https://doi.org/10.1080/10705511.2020.1811709.
32. Tofighi D, MacKinnon DP. RMediation: An R package for mediation analysis confidence intervals. Behav Res Methods. 2011;43(3):692–700.
33. Holland PW. Statistics and causal inference. J Am Stat Assoc. 1986;81(396):945–60.
34. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology. 2009;20(1):3–5.
35. Hernán MA. Does water kill? A call for less casual causal inferences. Ann Epidemiol. 2016;26(10):674–80.

36. VanderWeele TJ. Concerning the consistency assumption in causal inference. Epidemiology. 2009;20(6):880–3.
37. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health. 2006;60(7):578–86.
38. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008;168(6):656–64.
39. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Statistics and its Interface. 2009;2(4):457–68.
40. Valente MJ, Pelham WEI, Smyth HL, MacKinnon DP. Confounding in statistical mediation analysis: What it is and how to address it. J Couns Psychol. 2017;64(6):659–71.
41. Pirlott AG, MacKinnon DP. Design approaches to experimental mediation. J Exp Soc Psychol. 2016;66:29–38.
42. Valente MJ, Rijnhart JJM, Smyth HL, Muniz FB, MacKinnon DP. Review and comparison of software for the estimation of causal mediation effects. Structural Equation Modeling. 2020.
43. Ikram MA, VanderWeele TJ. A proposed clinical and biological interpretation of mediated interaction. Eur J Epidemiol. 2015;30(10):1115–8.
44. VanderWeele TJ. A three-way decomposition of a total effect into direct, indirect, and interactive effects. Epidemiology (Cambridge, Mass). 2013;24(2):224–32.
45. Booth RE, Davis JM, Dvoryak S, Brewster JT, Lisovska O, Strathdee SA, *et al*. HIV incidence among people who inject drugs (PWIDs) in Ukraine: Results from a clustered randomised trial. The Lancet HIV. 2016;3(10):e482–9.
46. Casey MK, Timmermann L, Allen M, Krahn S, Turkiewicz KL. Response and Self-Efficacy of Condom Use: A Meta-Analysis of this Important Element of AIDS Education and Prevention. Southern Communication Journal. 2009;74(1):57–78.
47. VanderWeele TJ. Causal Mediation Analysis With Survival Data. Epidemiology. 2011;22(4):582–5.
48. Hoyle RH, Kenny DA. Sample size, reliability, and tests of statistical mediation. In: Hoyle RH, editor. Statistical strategies for small sample research. Thousand Oaks, CA: Sage; 1999. p. 195–222.
49. Chen HT. Theory-driven evaluations. Newbury Park, CA: Sage; 1990.
50. MacKinnon DP. Analysis of mediating variables in prevention and intervention research. NIDA Res Monogr. 1994;139:127–53.
51. Bandura A. Social cognitive theory: An agentic perspective. Annu Rev Psychol. 2001;52(1):1–26.
52. Fishbein M. A theory of reasoned action: Some applications and implications. Nebr Symp Motiv. 1979;27:65–116.
53. Winett RA, Anderson ES, Desiderato LL, Solomon LJ, Perry M, Kelly JA, *et al*. Enhancing social diffusion theory as a basis for prevention intervention: A conceptual and strategic framework. Appl Prev Psychol. 1995;4(4):233–45.
54. Valente TW. Network models of the diffusion of innovations. Computational & Mathematical Organization Theory. 1996;2(2):163–4.
55. Latkin CA, Mandell W, Vlahov D, Oziemkowska M, Celentano DD. The long-term outcome of a personal network-oriented HIV prevention intervention for injection drug users: The SAFE study. Am J Community Psychol. 1996;24(3):341–64.
56. Latkin CA. Outreach in natural settings: The use of peer leaders for HIV prevention among injecting drug users' networks. Public Health Rep. 1998;113(Suppl 1):151.
57. Latkin CA, Knowlton AR. New directions in HIV prevention among drug users: Settings, norms, and network approaches to AIDS prevention (SNNAAP): A social influence approach. In: Latkin CA, editor. Emergent Issues in the Field of Drug Abuse. Bingley: Emerald Group Publishing Limited; 1999. p. 261–87.
58. Fisher JD, Fisher WA, Misovich SJ, Kimble DL, Malloy TE. Changing AIDS risk behavior: Effects of an intervention emphasizing AIDS risk reduction information, motivation, and behavioral skills in a college student population. Health Psychol. 1996;15(2):114–23.
59. MacKinnon DP. Introduction to statistical mediation analysis. New York: Routledge; 2008.
60. Albert JM, Nelson S. Generalized causal mediation analysis. Biometrics. 2011;67(3):1028–38.
61. Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. UCLA: Department of Statistics, UCLA. 2005:Retrieved from https://escholarship.org/uc/item/45x689gq.
62. Daniel R, De Stavola B, Cousens S, Vansteelandt S. Causal mediation analysis with multiple mediators. Biometrics. 2015;71(1):1–14.
63. Imai K, Yamamoto T. Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments. Political Analysis. 2013;21(2):141–71.
64. Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. Am J Epidemiol. 2013;179(4):513–8.
65. Nguyen TQ, Webb-Vargas Y, Koning IM, Stuart EA. Causal mediation analysis with a binary outcome and multiple continuous or ordinal mediators: Simulations and application to an alcohol intervention. Structural equation modeling: a multidisciplinary journal. 2016;23(3):368–83.
66. Taguri M, Featherstone J, Cheng J. Causal mediation analysis with multiple causally non-ordered mediators. Stat Methods Med Res. 2018;27(1):3–19.
67. VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. Epidemiologic Methods. 2014;2(1):95–115.
68. Wang W, Nelson S, Albert JM. Estimation of causal mediation effects for a dichotomous outcome in multiple-mediator models using the mediation formula. Stat Med. 2013;32(24):4211–28.
69. Yu Q, Fan Y, Wu X. General multiple mediation analysis with an application to explore racial disparities in breast cancer survival. Journal of Biometrics and Biostatistics. 2014;5(2):1–9.
70. Zheng C, Zhou XH. Causal mediation analysis in the multilevel intervention and multicomponent mediator case. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2015;77(3):581–615.
71. Mayer A, Thoemmes F, Rose N, Steyer R, West SG. Theory and analysis of total, direct, and indirect causal effects. Multivar Behav Res. 2014;49(5):425–42.
72. Steen J, Loeys T, Moerkerke B, Vansteelandt S. Medflex: An R package for flexible mediation analysis using natural effect models. J Stat Softw. 2017. https://doi.org/10.18637/jss.v076.i11.
73. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. J Stat Softw. 2014;59(5):38.
74. Medley A, Kennedy C, O'Reilly K, Sweat M. Effectiveness of peer education interventions for HIV prevention in developing countries: A systematic review and meta-analysis. AIDS Educ Prev. 2009;21(3):181–206.
75. Crepaz N, Marshall KJ, Aupont LW, Jacobs ED, Mizuno Y, Kay LS, *et al*. The efficacy of HIV/STI behavioral interventions for African American females in the United States: A meta-analysis. Am J Public Health. 2009;99(11):2069–78.
76. LaCroix JM, Pellowski JA, Lennon CA, Johnson BT. Behavioural interventions to reduce sexual risk for HIV in heterosexual couples: A meta-analysis. Sexually transmitted infections. 2013;89(8):620–7.
77. Eaton LA, Huedo-Medina TB, Kalichman SC, Pellowski JA, Sagherian MJ, Warren M, *et al*. Meta-analysis of single-session behavioral interventions to prevent sexually transmitted infections: Implications for bundling prevention packages. Am J Public Health. 2012;102(11):e34–44.

78. Tan JY, Huedo-Medina TB, Warren MR, Carey MP, Johnson BT. A meta-analysis of the efficacy of HIV/AIDS prevention interventions in Asia, 1995–2009. Soc Sci Med. 2012;75(4):676–87.
79. Albarracin J, Albarracin D, Durantini M. Effects of HIV-prevention interventions for samples with higher and lower percents of Latinos and Latin Americans: a meta-analysis of change in condom use and knowledge. AIDS Behav. 2008;12(4):521–43.
80. Fishbein M, Pequegnat W. Evaluating AIDS prevention interventions using behavioral and biological outcome measures. Sex Transm Dis. 2000;27(2):101–10.
81. Lyles CM, Kay LS, Crepaz N, Herbst JH, Passin WF, Kim AS, *et al*. Best-evidence interventions: Findings from a systematic review of HIV behavioral interventions for US populations at high risk, 2000–2004. Am J Public Health. 2007;97(1):133–43.
82. Coffman DL. Estimating causal effects in mediation analysis using propensity scores. Structural equation modeling: a multidisciplinary journal. 2011;18(3):357–69.
83. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Belmont: Wadsworth Cengage Learning; 2002.
84. Vansteelandt S. Estimating direct effects in cohort and case–control studies. Epidemiology. 2009;20:851–60.