Prev Sci (2019) 20:21–29
DOI 10.1007/s11121-017-0826-9

Preventive Effect Heterogeneity: Causal Inference in Personalized Prevention

George W. Howe[1]

Published online: 19 July 2017

Abstract This paper employs a causal inference framework to explore two logically distinct forms of preventive effect heterogeneity relevant for studying variation in preventive effect as a basis for developing more personalized interventions. Following VanderWeele (2015), I begin with a discussion of causal interaction involving manipulable moderators that combine to yield more complex nonadditive effects. This is contrasted with effect heterogeneity, which involves variation in causal structure indexed by stable characteristics of populations or contexts. The paper then discusses one particularly promising approach, the baseline target moderated mediation (BTMM) design, which uses theoretically informed baseline target moderators to strengthen causal inference, suggesting methods for using BTMM designs to develop targeting strategies for personalized prevention. It presents examples of recent intervention trials that apply these different forms of moderation, and discusses causal inference and the problem of moderation confounding, reviewing methods for minimizing its impact, including recent advances in the use of propensity score matching.

Keywords Research design · Causal inference · Personalized prevention · Moderation

Successful prevention requires that we alter risk mechanisms, or build protective mechanisms, before disease or disorder can manifest. MacKinnon et al. (2002) refer to these mechanisms as etiologic, and contrast them with action mechanisms involving social or psychological processes that shape etiologic factors. Prevention programs are designed to invoke action mechanisms to bring about such change. Both etiologic and action mechanisms are inherently causal; action mechanisms prevent future disorder through altering etiologic mechanisms, and this change in etiologic mechanism is responsible for later preventive effect.

Etiology and action may operate in the same way for most people. In that case, universal preventive interventions will be most useful. However, if etiologic or action research indicates that these mechanisms operate differently for different people or in different contexts, personalized or precision targeting may be called for (Weiss 2017). There is growing interest in precision targeting for the prevention of a number of disorders, including substance abuse (Ghitza 2015) and major depressive disorder (Falloon et al. 1992). As prevention scientists have developed trial designs for testing preventive efficacy, they have employed a number of methods for enhancing causal inference, including randomized field trials, propensity score analysis (Harder et al. 2010), sensitivity analysis (Liu et al. 2013), and causal mediation designs (Pearl 2012). Studying personalized prevention requires that we extend these methods to assess preventive effect heterogeneity.

In this paper, I employ a causal inference framework to explore two logically distinct forms of preventive effect heterogeneity. Following VanderWeele (2015) I begin by distinguishing causal interaction involving manipulable moderators that combine to yield more complex nonadditive effects. This is contrasted with effect heterogeneity, which involves variation in causal structure indexed by stable characteristics of populations or contexts. The paper then discusses one particularly promising approach, the baseline target moderated mediation (BTMM) design, which uses theoretically informed baseline target moderators to strengthen causal

✉ George W. Howe
ghowe@gwu.edu

[1] Department of Psychology, George Washington University, 2125 G Street NW, Washington, DC 20052, USA

inference. I present examples of recent intervention trials that apply these different forms of moderation. I then discuss causal inference and the problem of moderation confounding, and review research methods for minimizing its impact, including recent advances in the use of propensity score matching.

Causal Inference

Questions addressed by prevention scientists are inherently causal. We seek to test whether interventions have causal impact in reducing future disease or disorder. We design interventions to target factors that have causal impact in leading to such distal outcomes, and test whether such causal impacts are mediated through changing those more proximal targets. We design and conduct studies to test whether inferences about such causal impacts have empirical support. Recent work on the logic of causal inference provides an important resource for building designs that strengthen confidence in causal inferences. This section provides a brief outline of that work, and how it can be applied to research designs for testing effect heterogeneity.

Most recent work on causal inference derives from the Rubin's causal model (Rubin 1974) and Pearl's causal calculus as represented in directed acyclic graphs (Pearl 2009)."Both are consonant with and apply counterfactual reasoning. Consider a putative cause C and a putative effect E (putative because these are conjectures, and could be wrong). Under the simplest counterfactual model, C can be considered a cause of E if three conditions were obtained: the occurrence of C is followed by changes in E; if C had not occurred, E would not have changed; and all other possible causes of E are equal across the two conditions of C.

Causation requires that we observe changes in C (occurrence means going from absent to present) and determine whether changes in E follow, but that is not enough. We must also have evidence concerning the counterfactual condition: that is, if C had not changed, E would not have changed. Given that we can never observe the counterfactual, we must use some methods to approximate the counterfactual condition and study its impact. A number of methods are available for observing both the factual and the approximated counterfactual condition: between-subject designs that assign people to a treatment and control group (Rubin 1974), within-subject designs that assign people to receive both conditions in sequence (Josephy et al. 2015), or natural experiments that identify and observe people who are naturally exposed to a change in C and other people who are not (Rutter 2007).

Counterfactual logic can be extended to much more complex situations. Rather than just occurring, causes can involve change in state (e.g., moving from an urban to a rural setting) or change in intensity (e.g., having less contact with parents after moving away to college). If the effect variable is changing prior to the study (e.g., youth who are becoming more aggressive as they interact with deviant peers), change in C can suppress or prevent further change in E (e.g., growth in aggression stops after exposure to a prevention program). But in all cases, for causation to be inferred, one form of change in C must be contrasted with a different form of change in C, and the pattern of change in E must differ following those different conditions.

The third condition, referred to as the ignorability or nonconfounding assumption, requires that we rule out the possibility that alternate causes could account for observed differences in E. All other influences on E must be equivalent in the factual and approximated counterfactual condition. Although this assumption can never be proven, a variety of methods are available for increasing confidence that it is plausible. These include random assignment to C, suppression of variation in other factors through sampling (e.g., sample only women) or direct control (e.g., use the same intervention staff for both intervention and control conditions), or statistical methods for balancing on measured variables including propensity score matching or covariate adjustment.

This framework has several implications for studying effect heterogeneity, discussed in detail by VanderWeele (2015). Even in the simplest application, where we have two causes, each of which is either present or absent, the counterfactual structure becomes more complicated, because we now have four groups that need to be balanced on potential confounds. We will need to distinguish moderators that can be considered causes from those that cannot, because the latter cannot change. In the latter case, we will need to consider the possibility of moderational confounding, given that observed putative moderators may be associated with unobserved factors that are the true moderators of causal impact. And if we expand our causal models to include mediation, often necessary in understanding variation in long-term preventive effect, we will need to consider methods for reducing mediational confounding (Howe et al. 2002). Next, we turn to two general forms of moderation, discuss their potential application in the pursuit of personalized prevention, and clarify what is needed to strengthen causal inference.

Causal Interaction

Causal interaction, a term introduced by VanderWeele (2015), can occur when we target two or more separate action mechanisms. Causal interaction refers to the condition where two separate manipulable causes have a combined effect that differs from a simple additive effect (where the two causes operate independently). As an example, consider action theories of intervention that emphasize the importance of both trust building and skill building. Prevention interventionists have noted the importance of building trust or therapeutic alliance as a necessary prelude to

building protective skills such as productive coping or successful parenting (Multisite_Violence_Prevention_Project 2014). These theories imply causal interaction. They predict that skill building interventions are only successful in enhancing protective skills if they are combined with trust building. Figure 1 provides a directed graph of such an interaction.

Tests of causal interaction usually involve independent manipulation of both action mechanisms. Rather than creating a single factual and single approximated counterfactual condition, we need to create conditions that vary on all possible combinations for the two causes of interest. We would design a study that assigns participants to one of four conditions involving presence or absence of trust building activities crossed with presence or absence of skill building. In this case, we would predict that only the combined condition would show any impact on protective skills. VanderWeele and Robins (2008) refer to this as a form of sufficient cause interaction, where both conditions are necessary but neither is sufficient to bring about change. We might also hypothesize a weaker form of causal interaction, where skill building has some effect regardless of trust building, but has a stronger impact when it is combined with trust building compared to skill building alone.

Many preventionists would interpret these findings as demonstrating that trust building moderated the impact of skill building exercises on the development of protective skills. However, the results of this design are agnostic to what is the moderator and what the cause, and in fact, it is the combination of both factors that is causal, rather than one or the other. Figure 1 employs a dot as a convention for illustrating that the two causes operate in combination.

This is not necessarily true for more complex causal hypotheses. We might, for example, hypothesize that impact will depend on sequencing: we need to build trust before we can build skills. Testing this requires a more complex design that varies sequencing as well as presence or absence of each component. Here the group receiving both components could be further divided into those that receive the trust building first and those that receive the skill building first. If effects differ for these two groups, sequencing matters. If not, these groups can be collapsed and treated as the combined group in the four-group design.

Sequencing effects have been hypothesized and tested in both prevention and treatment interventions. Based on evidence that the effects of emotion regulation training for young children depended on verbal skills, Teisl et al. (2012) used a brief microtrial to test whether preliminary verbal skills training would enhance readiness for the intervention. They found that verbal skills training increased the effectiveness of the emotion regulation program on declarative knowledge of emotions, but only for those children who came to the program with deficiencies in verbal skills. They did not include a condition that reversed the sequencing, so the causal impact of sequencing could not be disentangled.

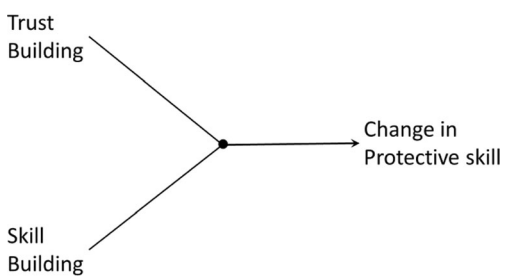

Fig. 1 Directed graph of causal interaction

Pelham et al. (2016) used an adaptive intervention design to test the effects of low-dose medication and behavior parent training on classroom rule violations in children with ADHD. This design allowed for comparison of different sequences of these interventions in a subset of the sample. Findings suggested that adding medication after parent training boosted effectiveness, but adding parent training after medication did not. The authors speculated that medication can lead parents to attribute ADHD to biological causes which cannot be influenced by parenting, reducing motivation to learn new parenting skills when they were introduced later.

Both studies manipulated two causes derived from action theory, each seen as influencing specific etiologic mechanisms. In addition, both studies interpreted findings as reflecting the interaction of those etiologic mechanisms. Tiesl et al. suggested that both verbal skills and exposure to declarative knowledge about emotions were necessary, but neither was sufficient, for the incorporation of that knowledge. Pelham et al. suggested that causal attributions held by parents shaped motivation to engage in new parenting behavior, which in turn moderated the effects of exposure to new parenting skills. This reflects causal interaction among etiologic mediators of intervention impact.

Such interactions will be particularly important to consider when bringing adaptive designs to prevention research, as discussed in other papers in this special issue. Adaptive designs for treatment assess changes in targeted symptoms following an initial intervention, identify participants with no treatment response, and randomly assign that subgroup to a different intervention. Preventive effects for distal outcomes such as substance abuse are unlikely to manifest within the timeframe of the trial itself. As a result, adaptive prevention trials must focus on proximal changes in etiologic risk or protective factors. Developmental research can guide such designs when it points to developmental sequencing of risk or protective mechanisms.

In the absence of strong theory or research, microtrial versions of adaptive designs will likely prove useful in exploring causal interactions of etiologic mechanisms or between etiologic and action processes. Microtrials (Howe et al. 2010) are randomized experiments testing relatively brief and focused environmental manipulations, designed to suppress specific risk mechanisms or enhance specific protective mechanisms,

but not to bring about full preventive effects on distal outcomes. Requiring smaller samples and shorter time frames, microtrials are an efficient means of testing malleability of etiologic mechanisms through brief manipulations built on action theory (Howe et al. 2010). Adaptive microtrials could be used to evaluate sequencing of action components based on change or lack of change in targeted etiologic factors, prior to fielding full-scale adaptive prevention trials.

Tests of causal interaction can clarify how program components targeting different action mechanisms may operate in combination to shape intervention effect. By themselves, however, causal interaction studies do not provide information to help with personalized prevention unless they also attend to person or contextual variables that index different forms of causal impact. The Tiesl et al. study provides such an example. In that study, children differed on level of basic verbal skills, and the sequencing of program components had more of an impact on those who came into the program with lower skill levels. We return to this issue later when we discuss baseline targets as potential moderators of intervention effect.

Confounding of Causal Interaction Random assignment to the four conditions defined by the presence or absence of the two intervention components (or the five conditions when sequencing is of interest) will enhance confidence that the ignorability or nonconfounding assumption is not violated. Constrained randomization is commonly used to insure that assignment results in equal sample sizes for all groups. And, as with simple randomized trials, comparing groups on baseline measures can provide further tests of that assumption.

If group differences are found on baseline variables, regression-based models such as analysis of covariance (ANCOVA) can be employed as a means of controlling for one or more potential confounds. It is important to note that simple adjustment using such covariates will not necessarily correct for potential confounding of the interaction effect if the equal slope assumption of ANCOVA is violated (i.e., if covariates operate differently across groups). This would reflect moderator confounding, discussed in more detail later. In this case, within-group covariate adjustment would be necessary.

Designs that sequence action components also present more challenges for causal inference. The five-group design discussed earlier is a partial factorial design. It assumes that order of condition may be important, but does not test all possible combinations of timing and duration. For example, suppose we require that each condition exposes participants to some action for the same length of time. This controls for duration, but leads us to compare a sequenced condition (e.g., trust-building followed by skill building) where each action occurs for only half the time, with a single condition (e.g., only trust-building) that occurs for the entire time of the intervention. A more complete design would include sequence and order information, with each intervention variable taking one of three conditions: absent, first in sequence, and last in sequence, leading to nine separate conditions. Such designs are very uncommon because of sample sizes required, although they may be more feasible when employed in microtrials.

The study by Pelham et al. suggests that adaptive designs may provide information about causal interaction, but causal inference with such designs is complicated. Adaptive designs use initial response to intervention to determine whether to shift to a different intervention: poor responders are randomly assigned to either continue the old intervention or switch to a new one (Collins et al. 2004). Effects in the second phase can be interpreted as causal within the group of poor responders, but comparison of the different sequences (in the Pelham et al. study, medication followed by parent training, or parent training followed by medication) cannot. Poor responders to one initial intervention may differ from poor responders to a different initial intervention. For example, if girls were more responsive to parent training and boys to medication, the two nonresponder groups will differ on gender as well as on initial intervention status. This will not present a problem for interpreting effects of subsequent intervention within each nonresponder group, given random assignment, but would confound the comparison of different sequences across those groups. Statistical methods may prove useful in dealing with such moderational confounding, but this is clearly an area needing further development.

Effect Moderation

Effect moderation occurs when intervention effects vary across people or contexts. VanderWeele also referred to this as effect modification (VanderWeele 2009) or effect heterogeneity (VanderWeele 2015), although Brand and Thomas (2013) have applied the latter term in more limited fashion to variation in treatment effects due to factors that influence selection into treatment in nonrandomized studies, also referred to as treatment-selection interaction by Shadish et al. (2002).

Unlike causal interaction, the population or contextual factor is not a cause, because it cannot change.[1] Rather it indexes variation in causal structure. That is, different values of the

[1] The causal status of factors that cannot change within individuals, but can vary across individuals, has been the source of some controversy, as witnessed by a recent exchange on race and gender (Glymour and Glymour 2014; Kaufman 2014; VanderWeele and Robinson 2014a, b). Variables such as race can clearly be causes in models where we define them as stimulus conditions for the behavior of others. In that case, they are in fact changeable, since we could vary those conditions and assess the impact of those different conditions. And if the behavior of others in turn has an impact on those who are the "stimuli," then personal characteristics could have an indirect impact on the person who has that characteristic, mediated through the behavior of others. In this case, variation in race across individuals would provide some index, albeit indirect, of rates of exposure to discrimination by others.

effect moderator indicate that different causal processes are in operation. This can happen in several ways. Differences in some stable variables can simply index differences in some other causal variables. For example, gender can index differences in history of exposure to significant stressors including victimization, a risk factor for later depression (Nolen-Hoeksema 2001), with women exposed to higher levels than men. In this case, gender is not an effect moderator, but rather an indirect index of variation in something that could be a cause, since exposure to victimization can change. However, general population characteristics such as gender can indirectly index variation in many different types of exposure, greatly limiting their utility as stand-ins for more concrete causes.

Effect moderators are stable factors that index variation in the cause-effect relationships embodied in action or etiologic mechanisms. In its simplest form, this would reflect variation in intervention impact. For example, Cardemil et al. (2002), in a randomized field trial of a school-based prevention program for youth depression, found preventive effects for Latino children but not African-American children. Figure 2 uses a slightly different graph syntax to represent effect moderation, reflecting the notion that the effect moderator is not a cause, but rather indexes variation in the causal impact of intervention on outcome.

Effect moderators may also index variation in more specific etiologic or action mechanisms. For example, in a longitudinal study of etiologic mechanisms, Cho et al. (2016) found that variation in a candidate gene involved in the dopamine system (DRD4) moderated the association of responsive/supportive parenting with increases in youth self regulation, such that youth with the long DRD4 allele showed stronger effects, consonant with the hypothesis that they were more susceptible to the influence of changes in their social environments. Consistent with this moderated etiologic process, Brody et al. (2014) found that youth DRD4 variation moderated the impact of the Strong African American Teens Program on adolescent substance use for males, and these effects were mediated through differential impact on supportive parenting practices.

Prevention scientists have focused on effect moderation associated with population characteristics, given that it can provide information concerning which groups are most likely to benefit from a prevention program. Much of this work attends to "the usual suspects" involving broad demographic characteristics such as gender (Clarke et al. 2001) or ethnicity (Cardemil et al. 2002). This may be profitably extended through attention to

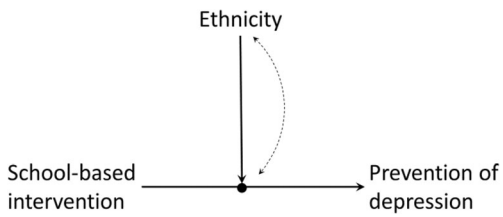

Fig. 2 Directed graph of effect moderation. The *dashed arc* indicates the possibility of moderational confounders

effect moderators derived from theory and research on action and etiologic mechanism. Likely candidates include variation in sensitivity to context (Brody et al. 2014) and developmental readiness (Bierman et al. 2014).

Although broad general person characteristics may prove to be important moderators of preventive intervention effects, research on personality variables as moderators has not always shown strong effects. We may gain more mileage by focusing on more specific risk and protective mechanisms that already inform prevention programs. As such, there is growing interest in studying baseline levels of etiologic targets as effect moderators of preventive impact.[2]

Baseline Target Moderation Howe et al. (2016) suggested that the etiologic targets of an intervention should always be considered as likely candidates for effect moderation. Participants entering a prevention trial almost always vary on the level of etiologic targets. That is, some participants are likely to begin a study with higher levels of risk factors or lower levels of protective factors than others. For example, in a study combining data from three randomized trials of the Familias Unidas program, Perrino et al. (2014) found that Hispanic families entering the program varied substantially in the quality of communication between parents and adolescents. Effective communication was considered an important protective factor, and the Familias Unidas program included components designed to target such communication.

Howe et al. suggested that baseline targets should moderate the impact of the intervention because those participants who begin with less of a protective factor targeted by the intervention will gain more, while those with higher levels of that factor will have less to gain.[3] Perrino et al. (2014) found

[2] Baseline targets are usually effect moderators rather than part of a causal interaction. The status of a target variable at baseline is historical, that is, when the intervention occurs, that baseline value is in the past, and is unchangeable. The future status of that variable can change as a result of the intervention, and therefore can be considered causal of later changes in more distal outcomes, but the baseline value of that target is immutable when the intervention commences. More complex effects are also possible, although infrequently hypothesized. For example, a target may change dynamically in the period prior to intervention, and prior rates of change (rather than status at baseline) could in principle moderate intervention impact on subsequent rates of change in the target. An intervention might exacerbate that rate of change, or dampen it, and do so in different ways depending on the rate of change at the time the intervention begins. This would reflect a causal interaction, although preintervention rates of change in the target would usually be observed rather than induced.

[3] A reviewer raised concerns as to whether BTMM effects could simply reflect regression to the mean. Values on target variables observed over two or more occasions may change for many reasons, including regression to the mean due to unreliability of measurement, decay of unobserved causes that led to initial increase, or exposure to other unobserved causes that shape change in the target. Unless there is reason to believe that measurement reliability differs for intervention and control groups, leading to differential regression to the mean for a target variable, regression artifacts are unlikely to account for findings of baseline target moderation of intervention impact on change in the target.

evidence for such moderation: the Familias Unidas program reduced future adolescent internalizing symptoms more for those families who began the program with less effective family communication, compared to those who began with better communication.

The baseline target moderation design can be extended through testing whether effect moderation is in turn mediated by changes in the target variable. Howe et al. referred to this as a baseline target moderated mediation (BTMM) design. Figure 3 presents a directed graph summarizing this model. Perrino et al. (2014) also found support for moderated mediation: increases in family communication mediated the impact of the intervention on reductions in internalizing, and this effect was strongest for those with the lowest baseline levels of communication, diminishing for those who began with higher baseline levels. Gonzales et al. (2012) found similar effects in a prevention trial of the Bridges/Puentes program for Mexican American families, reporting baseline target moderation for several targeted parenting behaviors including harshness, positive reinforcement, and monitoring, as well as targeted adolescent behaviors involving active coping and school involvement.[4]

Causal inference with BTMM models is enhanced by longitudinal designs that allow for direct assessment of sequential change in mediators and outcomes. For example, Perrino et al. (2016) studied the impact of the Familias Unidas program on internalizing symptoms in an eighth grade cohort of Hispanic youth. They assessed family communication, externalizing, and internalizing at baseline and at 6-, 18-, and 30-month postintervention. Using autoregressive models, they found evidence for a chained mediation effect. The intervention led to changes in communication from baseline to 6 months; changes in communication predicted changes in externalizing from 6 to 18 months, and those changes in turn predicted changes in internalizing from 18 to 30 months. They also tested whether early changes in symptoms mediated impact on later changes in communication. These alternate models were not supported, providing stronger support for causal interpretation of the primary finding. Confounding of mediation paths is still possible, although lagged change-to-change findings rule out many plausible historical confounds (Gunasekara et al. 2011). Such models can be further strengthened through use

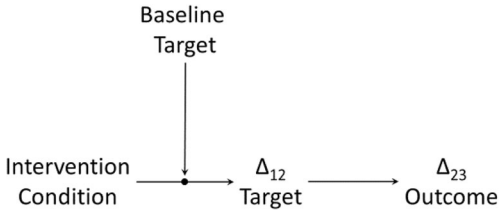

Fig. 3 Directed graph of the baseline target moderated mediation model. Δ_{12} references change in target variable from pretest to posttest; Δ_{23} references change in outcome from posttest to follow-up

of latent change modeling including latent change scores or growth models, which allow for further control of initial levels, should they co-vary with rates of change.

Baseline target moderation has practical implications for personalized prevention. Prevention programs often incorporate components designed to influence multiple targets. If, as in the case of the Bridges/Puentes trial, baseline target moderated mediation is found for several targets, it will be profitable to explore personalization based on baseline target assessment. Families may well differ in their profiles of risk and protective targets, opening the possibility of tailoring intervention components to specific profiles. The logic of this approach is similar to that of the Family Checkup intervention (Stormshak and Dishion 2009), which assesses risk and protective factors as a basis for referring families to relevant services.

Tailoring based on such profiles will be useful only to the degree that specific components do in fact change specific targets. BTMM designs assessing multiple targets can provide an initial test of target specificity by assessing whether baseline target moderation is mediated only through subsequent changes in that specific target. For example, if baseline parent communication moderates impact only through changes in parenting, and baseline child coping moderates impact only through changes in coping, we have stronger evidence for specificity of impact. In addition, microtrials can play an important intermediate role in testing the impact of individual components on specific etiologic targets.

Matching specific intervention components to specific risk or protective factors assumes that these factors have independent effects on outcomes. However, risk and protective mechanisms often co-occur. For example, the parenting variables targeted by the Bridges/Puentes program can intercorrelate strongly (Dumka et al. 2009). When mechanisms are highly correlated, then effective intervention will depend on the underlying causes of that correlation. Some mechanisms may be causally prior to others; for example, parental harshness may suppress effective parent-child communication. In this case, interventions for some may require sequencing, beginning with components that help parents become less harsh and progressing to those that build communication skills. As discussed earlier, microtrials can be designed to test sequenced causal interactions of this sort.

[4] A reviewer noted that promotive interventions often seek to build on existing strengths as a way of promoting well-being, suggesting that this is antithetical to the notion that those with baseline deficits will improve the most. This raises some interesting issues. Can a strength be a "latent" protective factor, in that it is present but inactive when risk mechanisms are in operation, and must be activated to become protective? Can strengths in one life arena be activated through intervention to reduce risks in another arena where they are not currently being employed? Both seem plausible, and could easily be incorporated into the BTMM framework. Activation level of a strength could be a reasonable baseline moderator (when the strengths are already activated and are functioning to protect from risk, interventions to activate them will be less necessary and have less impact).

Risk and protective mechanisms may also mutually influence each other, leading to stable clustering. In this case, we may need to target clusters rather than individual risk or protective mechanisms. Prevention trials using BTMM designs would again prove useful as a means of testing whether moderated mediation operates at the level of the cluster (using composites or latent variables) rather than uniquely for each specific target.

Combining Effect Moderation with Causal Interaction Personalizing prevention will be important if action mechanisms operate differently for different people or in different contexts. If we find evidence for causal interaction among intervention components, it will be important to study whether this more complex action mechanism operates in the same way for all participants, or whether its impact in turn varies across person or context. Tests of such variation require that we add measures of putative effect moderators to our causal interaction design.

For example, we might hypothesize that girls will respond more quickly to trust building components than boys. As a result, the combined effects of trust building and skill building may also vary by gender. We could study such moderation by randomizing participants to the four intervention groups (crossing trust building with skill building) and stratifying that random assignment within gender. This would allow us to test for three forms of effect moderation by gender: differential effects of trust building, of skill building, or of the interaction between the two (a three-way interaction, in analysis of variance terms).

Hybrid designs that include baseline targets as moderators of causal interaction among program components are also possible. Given that larger sample sizes may be needed to test more complex interactions, it may prove more efficient to begin with microtrial studies, given that tight experimental control and more rigorous measurement can reduce sample sizes necessary for detecting such interaction effects (McClelland and Judd 1993).

Causal Inference and Moderational Confounding Unlike causal interaction, effect moderation is inherently asymmetric: we conjecture that the causal impact of one variable, usually the preventive intervention, varies across values of some other variable, the effect moderator. Effect moderators are not causes: as noted earlier, they index variation in causal impact among other variables.

However, causal moderation also involves inherent ambiguities similar to those of cause-effect confounding. We may find that a preventive intervention has stronger impact for males than females, but we cannot conclude on this evidence alone that gender is the actual effect moderator. This would require that our study design eliminates the possibility that gender is associated with, and acts as a proxy for, some other unobserved variable that is the actual index of variation in effect. This could occur due to chance (e.g., by chance we sample more older females and more younger males) or through selection effects (e.g., more girls than boys from higher SES backgrounds choose to enroll in our study). I refer to this as moderational confounding to distinguish it from cause-effect confounding. The dashed arc in Fig. 2 represents this possibility. This problem holds for any potential moderator that cannot be randomly assigned, including baseline targets. It cannot be solved through stratified assignment (randomly assigning within each level of the moderator), which eliminates any confounding of the moderator with intervention status but has no impact on the association of the moderator with other unobserved factors.

Moderational confounding may or may not be a problem, depending on our goals. VanderWeele and Robins (2007) note that proxies can be useful when we wish to identify subgroups who would benefit more from an intervention. This does however assume that the association between the proxy variable and the true moderator is invariant across populations where we want to apply the intervention.

Moderational confounding is a problem when we wish to understand more about the mechanisms through which effect moderation operates. For example, in a trial of the Strong African American Teens program discussed earlier (Brody et al. 2014), presence of the long DRD4 allele was interpreted as a proxy for heightened sensitivity to the environment. Based on etiologic research, Brody et al. hypothesized three possible moderating mechanisms: sensitivity to positive social environments, sensitivity to negative social environments, and difficulty in developing behavioral control. Variation in the DRD4 gene was chosen because it had been associated with differences in brain dopamine function and risk for problems of cognitive processing. Notably, Brody et al. found evidence for a different moderating mechanism: teens with the long DRD4 allele (conjectured to confer greater sensitivity) had parents who responded more to the intervention, altering their parenting to include more monitoring, problem solving, and communication. These changes in parenting mediated the effects of the intervention on reducing teen substance use. Observed parenting may in this case be indexing the state of a dyadic process, involving parent-teen interaction, itself shaped by the intervention program more for those dyads involving a teen with the long DRD4 variant.

However, this study could also be subject to moderational confounding. Different gene variants often co-occur in specific subpopulations, a condition known as absence of linkage disequilibrium (Kidd 1993). Sample variation for one gene may well be associated with sample variation in other genes, and so it is possible that moderation is due to variation in other unmeasured genes. And if that is the case, those genes may be indexing different etiologic mechanisms, including those that do not involve the dopamine system.

Propensity score matching, originally developed to reduce confounding in observational studies of binary causes (such as presence or absence of treatment), has recently been applied to reduce moderational confounding. Green and Stuart (2014), in a study testing whether lifetime occurrence of major depressive disorder affected later substance use differently for men and women, used 24 variables to predict propensity for major depressive disorder, and estimated propensity scores for each participant. They applied these scores in five different propensity score matching strategies. Estimating separate models for men and women yielded the best balance across the four conditions (presence or absence of disorder across both genders). More traditional regression models using covariate adjustment resulted in the worst balance. Although Green and Stuart applied these methods to observational data, they should be applicable in tests of causal moderation involving a binary moderator and random assignment to presence or absence of the intervention condition.

Traditional propensity score matching has focused on binary causes (or in this case, binary moderators), given that the method becomes intractable as the number of causal conditions increases (Rassen et al. 2013). Prevention scientists are often interested in multi-valued moderators, particularly when studying baseline target moderation. In this case, more traditional methods employing regression analysis with covariates (and employing within-group covariate adjustment) could be compared with propensity matching methods that rescale the moderator as binary, usually through a median split. The former method will be more sensitive to graded moderation effects, but is more likely to be confounded: the latter will not be sensitive to graded moderation, but is likely to provide better control of confounding.

Recent attempts to extend propensity score methods to multi-valued causes may also prove useful. Yang et al. (2016) introduce a generalized propensity score model based on the idea of weak unconfoundedness. Matching based on this approach leads to balance across all observed covariates across multiple groups (Yang et al. 2016). This allows for using many more covariates than is feasible in standard regression analysis, although results could still be biased by unmeasured confounders. This approach may be particularly useful when moderator designs include more than two intervention conditions, even when the moderator is binary, as in the case where we wish to test whether causal interaction between two intervention components will vary across some effect moderator.

Conclusions

To develop personalized preventive interventions, we need to know more about which intervention components influence key risk or protective mechanisms, whether their causal impact varies when we combine components or when we apply them to different populations or in different contexts. Causal interaction designs can help determine how intervention components operate in combination; effect moderation designs can help us determine when and how to tailor interventions for maximal impact. Baseline targets are important effect moderators to consider, and moderated mediation designs can provide information about tailoring, particularly if they measure multiple targets and use longitudinal assessments to track change. All of these methods require close attention to issues of causal inference, and investigators need to consider methods to reduce confounding at each step of the design process.

Compliance with Ethical Standards

Funding This manuscript was supported in part by National Institute of Mental Health grant number R01-MH040859.

Conflicts of Interest The author declares that he has no conflict of interest.

Ethical Approval The manuscript does not report any empirical findings; no study was conducted requiring ethical approval.

Informed Consent The manuscript does not report any empirical findings; no study was conducted requiring informed consent.

References

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Welsh, J. A., & Gest, S. D. (2014). Integrating evidence-based preschool programs to support social-emotional and cognitive development. In M. Boivin, K. L. Bierman, M. Boivin, & K. L. Bierman (Eds.), *Promoting school readiness and early learning: Implications of developmental research for practice* (pp. 231–252). New York: Guilford Press.

Brand, J. E., & Thomas, J. S. (2013). Causal effect heterogeneity. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 189–214). New York: Springer.

Brody, G. H., Chen, Y.-f., Beach, S. R. H., Kogan, S. M., Yu, T., DiClemente, R. J., et al. (2014). Differential sensitivity to prevention programming: A dopaminergic polymorphism-enhanced prevention effect on protective parenting and adolescent substance use. *Health Psychology, 33*, 182–191. doi:10.1037/a0031253.

Cardemil, E. V., Reivich, K. J., & Seligman, M. E. P. (2002). The prevention of depressive symptoms in low-income minority middle school students. *Prevention & Treatment, 5*, Article 8. Retrieved from http://psycnet.apa.org/journals/pre/5/8. doi: 10.1037/1522-3736.5.1.58a.

Cho, J., Kogan, S. M., & Brody, G. H. (2016). Genetic moderation of transactional relations between parenting practices and child self-regulation. *Journal of Family Psychology*. doi:10.1037/fam0000228.

Clarke, G. N., Hornbrook, M., Lynch, F., Polen, M., Gale, J., Beardslee, W., et al. (2001). A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. *Archives of General Psychiatry, 58*, 1127.

Collins, L. M., Murphy, S. A., & Bierman, K. L. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science: The Official Journal Of The Society For Prevention Research, 5*, 185–196.

Dumka, L. E., Gonzales, N. A., Bonds, D. D., & Millsap, R. E. (2009). Academic success of Mexican origin adolescent boys and girls: The role of mothers' and fathers' parenting and cultural orientation. *Sex Roles, 60*, 588–599. doi:10.1007/s11199-008-9518-z.

Falloon, I. R. H., Shanahan, W., & Laporta, M. (1992). Prevention of major depressive episodes: Early intervention with family-based stress management. *Journal of Mental Health, 1*, 53–60. doi:10.3109/09638239209034511.

Ghitza, U. E. (2015). A commentary on 'A new initiative on precision medicine'. *Frontiers in Psychiatry, 6*.

Glymour, C., & Glymour, M. R. (2014). Commentary: Race and sex are causes. *Epidemiology (Cambridge, Mass.), 25*, 488–490. doi:10.1097/EDE.0000000000000122.

Gonzales, N. A., Dumka, L. E., Millsap, R. E., Gottschall, A., McClain, D. B., Wong, J. J., et al. (2012). Randomized trial of a broad preventive intervention for Mexican American adolescents. *Journal of Consulting and Clinical Psychology, 80*, 1–16. doi:10.1037/a0026063.

Green, K. M., & Stuart, E. A. (2014). Examining moderation analyses in propensity score methods: Application to depression and substance use. *Journal of Consulting and Clinical Psychology, 82*, 773–783. doi:10.1037/a0036515.

Gunasekara, F. I., Carter, K., & Blakely, T. (2011). Change in income and change in self-rated health: Systematic review of studies using repeated measures to control for confounding bias. *Social Science & Medicine (1982), 72*, 193–201. doi:10.1016/j.socscimed.2010.10.029.

Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods, 15*, 234–249. doi:10.1037/a0019623.

Howe, G. W., Reiss, D., & Yuh, J. (2002). Can prevention trials test theories of etiology? *Development and Psychopathology, 14*, 673–694. doi:10.1017/S0954579402004029.

Howe, G. W., Beach, S., & Brody, G. (2010). Microtrial methods for translating gene-environment dynamics into preventive interventions. *Prevention Science, 11*, 343–354. doi:10.1007/s11121-010-0177-2.

Howe, G. W., Beach, S. R. H., Brody, G. H., & Wyman, P. A. (2016). Translating genetic research into preventive intervention: The baseline target moderated mediator design. *Frontiers In Psychology*, 1–9. doi:10.3389/fpsyg.2015.01911.

Josephy, H., Vansteelandt, S., Vanderhasselt, M.-A., & Loeys, T. (2015). Within-subject mediation analysis in AB/BA crossover designs. *The International Journal of Biostatistics, 11*, 1–22. doi:10.1515/ijb-2014-0057.

Kaufman, J. S. (2014). Commentary: Race: Ritual, regression, and reality. *Epidemiology (Cambridge, Mass.), 25*, 485–487. doi:10.1097/EDE.0000000000000117.

Kidd, K. K. (1993). Associations of disease with genetic markers: Déjà vu all over again. *American Journal of Medical Genetics, 48*, 71–73.

Liu, W., Kuramoto, S. J., & Stuart, E. A. (2013). An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. *Prevention Science, 14*, 570–580. doi:10.1007/s11121-012-0339-5.

MacKinnon, D. P., Taborga, M. P., & Morgan-Lopez, A. A. (2002). Mediation designs for tobacco prevention research. *Drug and Alcohol Dependence, 68*, S69–S83. doi:10.1016/S0376-8716(02)00216-8.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114*, 376–390. doi:10.1037/0033-2909.114.2.376.

Multisite_Violence_Prevention_Project. (2014). Implementation and process effects on prevention outcomes for middle school students. *Journal of Clinical Child and Adolescent Psychology, 43*, 473–485. doi:10.1080/15374416.2013.814540.

Nolen-Hoeksema, S. (2001). Gender differences in depression. *Current Directions in Psychological Science, 10*, 173.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). New York: Cambridge University Press.

Pearl, J. (2012). The causal mediation formula—A guide to the assessment of pathways and mechanisms. *Prevention Science, 13*, 426–436. doi:10.1007/s11121-011-0270-1.

Pelham, W. E., Fabiano, G. A., Waxmonsky, J. G., Greiner, A. R., Gnagy, E. M., Coxe, S., et al. (2016). Treatment sequencing for childhood ADHD: A multiple-randomization study of adaptive medication and behavioral interventions. *Journal of Clinical Child & Adolescent Psychology, 45*, 396–415. doi:10.1080/15374416.2015.1105138.

Perrino, T., Pantin, H., Prado, G., Huang, S., Brincks, A., Howe, G., et al. (2014). Preventing internalizing symptoms among Hispanic adolescents: A synthesis across Familias Unidas trials. *Prevention Science, 15*, 917–928. doi:10.1007/s11121-013-0448-9.

Perrino, T., Brincks, A., Howe, G., Brown, C. H., Prado, G., & Pantin, H. (2016). Reducing internalizing symptoms among high-risk, Hispanic adolescents: Mediators of a preventive family intervention. *Prevention Science: The Official Journal Of The Society For Prevention Research, 17*, 595–605. doi:10.1007/s11121-016-0655-2.

Rassen, J. A., Shelat, A. A., Franklin, J. M., Glynn, R. J., Solomon, D. H., & Schneeweiss, S. (2013). Matching by propensity score in cohort studies with three treatment groups. *Epidemiology (Cambridge, Mass.), 24*, 401–409. doi:10.1097/EDE.0b013e318289dedf.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*(5), 688–701. doi:10.1037/h0037350.

Rutter, M. (2007). Proceeding from observed correlation to causal inference: The use of natural experiments. *Perspectives on Psychological Science, 2*, 377–395. doi:10.1111/j.1745-6916.2007.00050.x.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin Co..

Stormshak, E. A., & Dishion, T. J. (2009). A school-based, family-centered intervention to prevent substance use: The family check-up. *The American Journal of Drug and Alcohol Abuse, 35*, 227–232. doi:10.1080/00952990903005908.

Teisl, M., Wyman, P. A., Cross, W., West, J., & Sworts, L. (2012). *Adaptive intervention to address the needs of children with language delays and behavior problems: Proximal impact on emotion-regulation skill knowledge*. Paper presented at the Annual meeting of the Society for Prevention Research, Washington, DC.

VanderWeele, T. J. (2009). On the distinction between interaction and effect modification. *Epidemiology (Cambridge, Mass.), 20*, 863–871. doi:10.1097/EDE.0b013e3181ba333c.

VanderWeele, T. J. (2015). *Explanation in causal inference. Methods for mediation and interaction*. Oxford: Oxford University Press.

VanderWeele, T. J., & Robins, J. M. (2007). Four types of effect modification: A classification based on directed acyclic graphs. *Epidemiology (Cambridge, Mass.), 18*, 561–568.

VanderWeele, T. J., & Robins, J. M. (2008). Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika, 95*, 49–49.

VanderWeele, T. J., & Robinson, W. R. (2014a). On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology (Cambridge, Mass.), 25*, 473–484. doi:10.1097/EDE.0000000000000105.

VanderWeele, T. J., & Robinson, W. R. (2014b). Rejoinder: How to reduce racial disparities?: Upon what to intervene? *Epidemiology (Cambridge, Mass.), 25*, 491–493. doi:10.1097/EDE.0000000000000124.

Weiss, N. S. (2017). What findings are needed to advocate personalized (precision) prevention of disease? *American Journal of Public Health, 107*, 86–87. doi:10.2105/AJPH.2016.303513.

Yang, S., Imbens, G. W., Cui, Z., Faries, D. E., & Kadziola, Z. (2016). Propensity score matching and subclassification in observational studies with multi-level treatments. *Biometrics, 72*, 1055–1065.