Path Specific Effects Under Monotonic Mediation

Aleksei Opacic and Xiang Zhou

Harvard University

January 23, 2023

Abstract

Identification and estimation of causal mechanisms is an area that has seen rapid innovation in recent years across the social sciences. Yet, to date, mediation analysis has been implicitly concerned exclusively with settings where levels of the mediator are accessible to all individuals, irrespective of their treatment status. Thus, conventional approaches are ill equipped for studying many demographic events, actions and milestones that are characterized by a setup in which each "transition" can be considered "monotonic" in nature. In this article, we introduce a general framework for tracing causal mediation effects in the context of multiple monotonic mediators. Our framework considers the effect of any given demographic transition as operating indirectly, via subsequent transitions, and directly, net of these transitions. We demonstrate that the average treatment effect (ATE) of a transition can be additively decomposed into mutually exclusive components that capture these direct and indirect effects. Our proposed decomposition has a number of special properties which distinguish it from conventional decompositions of the ATE via multiple sequential mediators, properties which facilitate less restrictive identification assumptions as well as identification of all of the causal paths in the decomposition. We propose both parametric and semiparametric methods for estimating our decomposition, and illustrate the proposed framework using two empirical examples drawn from the education and sociology literature.

1 Introduction

Tracing the causal mechanisms through which a treatment affects an outcome is a central goal of the social sciences. Epidemiologists often explore the extent to which the effect of a particular drug or intervention on health outcomes is mediated via certain types of physiological changes (Vanderweele et al 2014). In political psychology, interest often lies in examining how differential framing of a particular political issue affect attitudes towards that issue via either changing individuals' beliefs about or perceived importance of it (e.g. Brader et al 2008). In sociology, a large body of literature for instance investigates whether the effects of neighborhood disadvantage on youths' educational outcomes occur primarily through aspects of school disadvantage, or alternatively through different mechanisms such as levels of crime or environmental pollution. Despite their substantive particularities, concern with causal mediation across these disciplines is unified by the goal of testing and evaluating theories that aim to explain a range of social and political regularities.

Compared with its beginnings in the 1980s, the field of mediation analysis has become attentive to complex issues of identification and estimation in ways that rival scholarly innovation in identifying and estimating the ATE. Indeed, the range of mediation estimands and methods for mediation estimation has grown rapidly in recent years. Researchers are now equipped with strategies to not only decompose the total effect of a treatment into not only its direct and indirect effects via a single mediator, but to redefine their mediation effects based on a preference to either assume away (Pearl 2001, Imai et al. 2010, 2011) or explicitly model (Imai and Yamamoto 2013), Vanderweele et al 2014) intermediate confounders for the effect of the mediator on the outcome, trace the multiple causal pathways of a treatment through causally dependent mediators (Vanderweele et al 2014), and examine mediation that varies across time (Vanderweele and Tchetgen Tchetgen 2019). Estimation strategies also abound; researchers are able to use simple linear models for these decompositions (e.g. Vanderweele and Vansteelandt 2009, Wodtke and Zhou 2021), a range weighting and imputation estimators (Vanderweele et al 2014, Imai et al 2011, Zhou and Yamamoto 2022), or alternatively can employ semiparametric approaches (Tchetgen Tchetgen Shipster 2012, Miles et al 2015, Farbmacher et al 2020, Zhou 2022).

Despite the plethora of extant approaches for the study of causal mediation, this explosion of work is implicitly concerned only with settings where all values of a putative mediator are accessible to individuals irrespective of their treatment status. Neighborhood poverty, for example, may make school disadvantage more likely, but school disadvantage is by no means restricted to neighborhoods that are poor. Similarly, health insurance may increase incidence of regular medical checkups, but some individuals may undertake regular checkups even if they are uninsured. Nevertheless, applied researchers are often engaged in settings in which sequential transitions are "monotonic" in nature - that is, in which some demographic transition may exclusively, or almost exclusively, be "accessed" by individuals after they have accessed or attained the previous transition. This setting is particularly salient for demographic-related questions, which often involve sequential transitions over the life course. For example, in the context of educational attainment, individuals can typically attend college only if they graduate high school. Certain demographic events are more rigid in their monotonicity as a result of their very definition. Clearly, an individual must have married in order to divorce. Similarly, individuals can only graduate college if they attend college. Researchers may be interested in decomposing the total effect of marriage on various outcomes earnings or life satisfaction, into its direct and and indirect components via divorce; divorce can only be "attained" among individuals who are already married. Similarly, the effect of parenthood on earnings can be seen as operating directly, through the effect of having a single child net of subsequent children, as well as operating indirectly through the effects of having future children on earnings, transitions which are clearly monotonic in nature. A similar perspective may be taken to decompositions of the total effect on later stage outcomes of criminal offending via recidivism, or arrest via incarceration. Because typical approaches to causal mediation rely on variation in mediator statuses across treatment status, they are ill-equipped for exploring scientific questions with this particular characteristic.

Recognition of the sequential nature of many demographic and social transitions implies that a causal mediation framework may be taken to study their effects on various life outcomes. Specifically, we can consider the first transition in a demographic sequence of interest as a treatment variable,

A, and subsequent transitions as mediators that "transmit" the effects of the treatment and of prior transitions, M_k ($1 \leq k \leq K$). For example, if we are interested in the total effect of high-school completion on earnings, we may ask to what extent this total effect operates indirectly, through the effect of college attendance (a putative mediator) on earnings, or directly, through alternative causal pathways. The insight that the total causal effect of a demographic transition can be decomposed into its direct and indirect effects opens up a range of important research questions and opportunities, including the ability to explicitly trace the causal paths by which an event or treatment affects some outcome, and to evaluate potentially countervailing patterns of heterogeneity in direct and indirect effects. Nevertheless, existing research in these types of demographic contexts often close the door to such research questions by either failing to distinguish between the direct and indirect components of transition effects, or by using approaches that lead to estimation bias. For example, research on the effects of marriage on later life outcomes almost exclusively takes a dichotomous approach to the measurement of marriage, disregarding divorce. By conflating the direct and indirect effects of marriage, such research is thus inattentive to whether the positive effects of marriage may be attenuated by the associated risk of divorce down the line. Further, in educational settings, researchers resort to one of several strategies, which include using a continuous measure of years of schooling, or coding educational attainment into one of several mutually exclusive categories. Neither of these approaches is optimal: the former focuses on the "direct effect" of each additional year of schooling, while the latter is at risk of severe estimation bias because the potential outcomes under each level of attainment are likely not independent of the categories conditional on baseline confounders.

In this article, we show that demographic events and milestones characterized by such monotonicity can be analyzed via causal mediation, and introduce a causal mediation framework for analyzing the effects of monotonic mediators. Specifically, we make two contributions. First, we develop a general formula that decomposes the total effect of a treatment into a direct effect, and K mutually exclusive gross effects: $A \to M_1 \cdots \to M_k \to Y$. This is in contrast to previous literature on this particular monotonic setting which has focused exclusively on the case of two mediators (e.g., Zhou 2021). The $K + 1$ monotonic path-specific effects (MPSEs) are all non-parametrically identified under the assumption of sequential ignorability, which allows for the effect of each educational transition to be confounded by a different set of covariates. Second, we introduce a range of estimation approaches for our proposed decomposition, including a simple linear model-based regression-with-residuals (RWR) procedure, as well as a non-parametric estimation approach.

Our mediation-based decomposition of the total effect of education is similar to conventional mediation analysis which involves multiple mediators, in particular a decomposition of the total treatment effect into a set of path-specific effects (PSEs) (see Miles et al 2015, Zhou 2021). However, our proposed decomposition is characterized by an important feature that distinguishes it from decompositions of treatment effects more generally, namely, the fact that for any individual to experience a given transition, that individual must have completed all prior transitions. Thus, while in the more general context, only composite path-specific effects of the form $A \to M_k \rightsquigarrow Y$ are identified, in the context of sequential monotonic mediation, *all* of the path-specific effects can in fact be identified. Further, the decomposition of educational effects can be identified under weaker assumptions than those required for the general decomposition of the ATE. Specifically, our decomposition accommodates the presence of a distinct set of observed intermediate confounders for each transition. Overall, our proposed approach constitutes a new framework for analyzing causal mechanisms in a range of demographic settings characterized by mediator monotonicity. Applied researchers can adopt our framework to make inferences about causal processes that were previously invisible to analysts. In the following sections, we first describe two empirical settings that serve as running examples throughout the paper, and introduce a decomposition in the case of a single monotonic mediator, before generalizing the approach to an arbitrary number of causally ordered, monotonic mediators mediators. Next, we discuss identification of the decomposition under the assumption of sequential ignorability, and introduce both parametric and non-parametric estimation procedures. Finally, we illustrate the proposed framework and methods with an empirical example using two empirical examples from the social sciences.

2 Monotonic Path-Specific Effects (MPSEs)

One of the most resilient social scientific findings across a range of national contexts is the strong associations between educational attainment and a variety of life outcomes, such as earnings, health, social capital, and family stability (Hout 2012). Conventionally, research has taken one of two approaches to conceptualizing the social and economic returns to education: the first employs a "years of schooling" measure of educational attainment (e.g. Card 1999, 2001), while the second dichotomizes attainment as a point-in-time treatment. This latter approach has been particularly influential in the study of higher education effects, where the treatment considered is often an indicator for whether a high-school graduate has attended, or graduated from, college (e.g., Brand and Xie 2010; Carneiro et al. 2011; Zimmerman 2014). Despite the important insights this literature has made into establishing the causal effect of these two operationalizations of attainment on important social and economic outcomes, existent work has been inattentive to the sequential process by which people make educational transitions (Mare 1980).[1] At the end of high-school, individuals decide to enroll at college or not. Among college enrollees, only about 60% of students receive a BA within six years of initial college entry, and the proportion is even lower for black students (Snyder, de Brey, and Dillow 2016). Amidst higher educational expansion in the US, graduates must increasingly choose whether to enter the labour market or to enroll in post-graduate education. Increasingly, therefore, educational attainment in the US has become a field of multiple levels with sequential transitions, all of which are independently consequential for individuals' labour market outcomes, and are therefore of independent scientific interest for scholars of education.

The benefits of marriage for family income gains are well documented. Income sharing during marriage enables each partner to benefit from the spouses' income; such benefits tend to be greater for women, since men's market incomes are on average higher than women's (Becker 1994). Marriage rather than cohabitation appears to be key for this poolign benefit: cohabiting couples are less likely

[1] For this particular example, we use the term "educational transition" to refer both to vertical transitions (e.g. enrollment at a secondary or tertiary institution), as well as to the attainment of a qualification at that institutional level (e.g. high-school graduation or BA completion).

to pool their incomes than are married couples and exhbit lower levels of commitment than married couples do (Kenney 2004; Smock 2000). Nevertheless, while the positive effects of marriage on income must be considered alongside the risk of divorce and the associated implications for earnings, especially since rates of marital dissolution increased for cohorts born in the latter half of the 20th century (Bumpass et al. 1991. Martin 2006, Sweeney & Phillips 2004), research on marriage effects often simply disregards divorce (e.g. Light 2003).

A unifying study of marriage and divorce promises to shed light on the countervailing forces that shape marriage's overall effects. While marriage's positive effects are well known, that divorce and separation pose a primary source of family income loss comes as little surprise, as dissolved spouses lose their pooling of household resources and economies of scale (Sorensen 1994). Such losses are typically estimated as being especially pronounced for women, as the economic inactivity of one partner can no longer be buttressed by the family income of the household breadwinner (Bane & Ellwood 1986, Burkhauser & Duncan 1988). Family income lossses for women one year after divorce are estimated in the range of 20 to 40% (e.g. Avellar and Smock 2005, Holden and Smock 1991, Tach and Eads 2015). Overall, it seems clear that the negative effects of divorce are less pronounced for men: some studies report gains for men after marital dissolution (Smock 1994, Galarneau and Sturrock 1997), while others report losses, particularly for men whose wife was the primary breadwinner (McManus and DiPrete 2001). These negative effects on family incomes are long lasting and only partially mitigated by remarriage, although the longevity of the negative effect appears to vary considerably across countries (Andress et al 2006).

2.1 A Single Monotonic Mediator

We first consider the case of two monotonic transitions (that is, an initial transition and a single monotonic mediator). Suppressing subscripts, let A denote an indicator for the initial transition, M denote an indicator for the second transition (the monotonic mediator), and Y a binary or continuous outcome of interest such as earnings. Our single-transition decomposition thus assesses the educational sequence $A \to M \to Y$. In our running example of the effect of marriage on earnings, for example, we thus treat divorce as a mediator of the total effect of marriage on earnings, in relation to which the total effect of marriage can be decomposed into an indirect effect (flowing through divorce), and a direct effect (net of divorce).

Using the potential outcomes notation, let $M(a)$ denote an individual's potential value of the mediator if their treatment status were set to a, and let $Y(a, m)$ denote that individual's potential outcome if their treatment and mediator statuses were set to a and m, respectively,. We first note that, although this is a mediation-style setting, in this special case we assume that $Y(a, m)$ is only defined in cases where $a \geq m$. In other words, for individuals to experience the monotonic transition M, we require that the treatment be present, too. For example, while $Y(0, 0)$ is defined (for instance, the potential outcome if a person does not marry and does not divorce, the quantity $Y(0, 1)$ is undefined (the potential outcome if a person divorces but does not marry). This sequential nature of monotonic transitions thus implies three potential outcomes: $Y(0, 0)$, $Y(1, 0)$, $Y(1, 1)$, while ignoring M, we have two potential outcomes, $Y(0)$ and $Y(1)$. It also means that, whereas in general, $Y(0) = Y(0, M(0)) \neq Y(0, 0)$, in the case of such *monotonic transitions*, the composition assumption implies $Y(0) = Y(0, 0)$, and

$$Y(1) = Y(1, 0) + M(1)[Y(1, 1) - Y(1, 0)].$$

Given the above equation, the total effect of A on Y can be decomposed as

$$Y(1) - Y(0) = Y(1, 0) - Y(0, 0) + M(1)[Y(1, 1) - Y(1, 0)]. \tag{1}$$

Since these quantities are unidentified at the individual level, we focus on the population-level equivalent. By taking the expectation of Equation 1, we obtain the following decomposition of the ATE (see Zhou 2021):

$$\text{ATE} = \mathbb{E}[Y(1) - Y(0)]$$

$$= \underbrace{\mathbb{E}[Y(1,0) - Y(0,0)]}_{\text{direct effect of treatment}} + \mathbb{E}[M(1)] \underbrace{\mathbb{E}[Y(1,1) - Y(1,0)]}_{\text{net effect of monotonic mediator}} + \underbrace{\text{cov}[M(1), Y(1,1) - Y(1,0)]}_{\text{selection into monotonic mediator}} \tag{2}$$

$$= \underbrace{\Delta_0}_{\text{A} \rightarrow Y} + \underbrace{\pi_1 \Delta_1 + \eta_1}_{\text{A} \rightarrow M \rightarrow Y}, \tag{3}$$

where Δ_0 and Δ_1 denote the direct effects of the first and second transition, $A \rightarrow Y$ and $M \rightarrow Y$, respectively, π_1 denotes $\mathbb{E}[M_1(1)]$, and η_1 denotes the covariance between the effect of the initial transition on completion of the second and the effect of the second transition on Y. In our running example on marriage effects, η_1 is positive if those who would divorce given marriage (i.e., $M(1) = 1$) benefit more from divorce in terms of their later earnings (i.e., have a larger $Y(1,1) - Y(1,0)$) than those who do not (i.e., $M(1) = 0$), and negative if the opposite holds. The composite term $(\pi_1 \Delta_1 + \eta_1)$ thus captures the average indirect effect of the initial transition via the subsequent transition ($A \rightarrow M \rightarrow Y$). For example, in the example of the effects of high-school graduation via completion completion, it comprises the sum of the probability of college enrollment if an individual graduated high-school multiplied by the direct of college enrollment and the covariance between college enrollment and its direct effect on earnings.

2.2 Generalization to K Monotonic Mediators

We now generalize the approach introduced in Section 2.1 to the case of K intermediate educational transitions. In the following, we denote the initial educational transition of interest, high-school graduation, by A and use $M_1, \dots M_K$ to refer to the K subsequent transitions of interest, where we assume that all of $M_1, \dots M_K$ are binary and that for any $i < j$, M_i temporally precedes M_j.

Let τ_0 denote the ATE of A and τ_k, the gross effect of the kth mediator, i.e.,

$$\tau_k = \mathbb{E}[Y(\overline{1}_{k+1}) - Y(\overline{1}_k, 0)].$$

In addition, let Δ_0 denote the direct effect of A and Δ_k, the direct effect of the kth mediator, i.e.,

$$\Delta_k = \mathbb{E}[Y(\overline{1}_{k+1}, \underline{0}_{k+2}) - Y(\overline{1}_k, \underline{0}_{k+1})].$$

To explicate our approach, we observe that the gross effect of the kth mediator includes not only the direct effect $M \rightarrow Y$ net of subsequent educational transitions (i.e.,Δ_k) but also the indirect effects of M via subsequent transitions ($M \rightsquigarrow Y$, where a squiggly arrow denotes a combination of multiple paths). This insight motivates us to further decompose τ into its direct and indirect components. Thus, under the composition assumption [2] τ_k can be decomposed as

$$\tau_k = \Delta_k + \pi_{k+1}\tau_{k+1} + \eta_{k+1}, \tag{4}$$

where

$$\begin{aligned}
\pi_{k+1} &= \mathbb{E}[M(\overline{1}_{k+1})], \\
\eta_{k+1} &= \text{cov}[M(\overline{1}_{k+1}), Y(\overline{1}_{k+2}) - Y(\overline{1}_{k+1}, 0)].
\end{aligned}$$

For $k = K-1, K-2, \ldots 1$, iteratively substituting equation 4 into the corresponding expression for τ_{k-1} yields

$$\tau_0 = \underbrace{\Delta_0}_{A \rightarrow Y} + \sum_{k=1}^{K} \underbrace{(\Pi_{j=1}^{k}\pi_j)\Delta_k + (\Pi_{j=1}^{k-1}\pi_j)\eta_k}_{A \rightarrow M_1 \ldots \rightarrow M_k \rightarrow Y},$$

where $\Delta_K = \tau_K$ (i.e., the Δ_K term will include indirect effects via subsequent educational transitions not considered in the decomposition), unless of course we have "saturated" our decomposition with all possible subsequent educational transitions, in which case $\Delta_K$can precisely be labelled a direct effect. [3]

[2] Which, to reiterate, in the context of monotonic education transitions, implies that a necessary condition for $M_k = 1$ is that $\{A, \overline{M}_{k-1}\} = 1, \forall k \in \{K\}$.

[3] More specifically, the $K-1$ Δ terms are insensitive to the number of education transitions considered in the decomposition since, by the composition assumption, terms such as $Y(1,1,0)$ can be equivalently written as $Y(1,1,0,0,0)$, etc. By contrast, the K^{th} Δ term will be a composite path (fea-

2.3 Identification of the Δ_k, π_k and η_k terms

In order to estimate the above decomposition, it suffices to identify three types of quantity: the expectation of composite counterfactuals such as $Y(a, m_1, \dots m_K)$, and $M_k(a, m_1, \dots m_{k-1})$, as well as covariance terms of the form $\mathrm{cov}[M_k(\overline{1}_k), Y(\overline{1}_k, \underline{0}_{k+1}) - Y(\overline{1}_{k-1}, \underline{0}_k)]$. Let an overbar denote a vector of variables, so that $M_k = (M_1, M_2, \dots M_k)$, and $m_k = (m_1, m_2, \dots m_k)$. These quantities are then: $\psi_{a\overline{m}_k} \equiv \mathbb{E}[Y(a, \overline{m}_k)]$ for any, $\varphi_{am_1} \equiv \mathbb{E}[M_k(a, \overline{m}_{k-1})]$, $\varphi_a \equiv \mathbb{E}[M_1(a)]$, for any $a, m_k, k \in [K]$. The latter quantity is simply the average treatment effect (ATE) on the mediator, and the first two are identifiable via the g-formula, which has typically been applied in the case of time-varying treatments (Robins 1986). Under sequential ignorability (namely, the assumption that M_k is independent of potential outcomes of its descendants, conditional on all antecedent variables), Δ_k and π_k are identified via the g-formula and the identification formula for the ATE, under the assumptions of consistency, sequential ignorability and positivity:

1. consistency: for any unit, if $A = a$ and $\overline{M}_K = \overline{m}_K$, then $Y = Y(a, \overline{m}_K)$;
2. sequential ignorability: $Y(a, \overline{m}_K) \perp\!\!\!\perp A | X, \forall a, k \in [K]$, and $Y(a, \overline{m}_k) \perp\!\!\!\perp M_k | X, A, \overline{Z}_k, \overline{M}_{k-1}, \forall a, k \in [K]$;
3. Positivity: $p_{A|X}(a|x) > \epsilon > 0$, $p_{M_k|X,A,\overline{Z}_k,\overline{M}_{k-1}}(m_k|x, a, \overline{z}_k, \overline{m}_{k-1}) > \epsilon > 0 \, \forall k \in [K]$.

Under these assumptions, we can identify Δ_k and π_k terms in the decomposition, which we denote as $\psi_{a\overline{m}_K}$ and $\varphi_{a\overline{m}_K}$, respectively, as:

turing the direct and continuation effects of subsequent transitions) and will depend on the number of transitions chosen.

Thus, the paths $A \rightarrow Y$ and $A \rightarrow M_k \cdots \rightarrow M_{K-1} \rightarrow Y$, for $k \in [K-1]$ as *direct* continuation effects, while K^{th} single path $A \rightarrow M_k \cdots \rightarrow M_K \rightarrow Y$ can more accurately be referred to as *gross* or composite continuation effect, since this latter path is a composite path in that it contains all of the residual paths omitted in the decomposition through educational transitions subsequent to K. An exception to this would of course be the case where the K paths considered in the decomposition 'saturate' the total number of educational transitions available to individuals.

$$\psi_{a\overline{m}_k} = \mathbb{E}[Y(a, \overline{m}_k)] = \int_x \int_{\overline{z}_k} \mathbb{E}[Y|a, x, \overline{z}_K, \overline{m}_k] \big[\prod_{j=1}^{k} dP(z_j|x, a, \overline{z}_{j-1}, \overline{m}_{j-1}) \big] dP(x) \tag{5}$$

$$\varphi_{a\overline{m}_k} = \mathbb{E}[M_{k+1}(a, \overline{m}_k)] = \int_x \int_{\overline{z}_k} \mathbb{E}[M_{k+1}|a, x, \overline{z}_K, \overline{m}_l] \big[\prod_{j=1}^{k} dP(z_j|x, a, \overline{z}_{j-1}, \overline{m}_{j-1}) \big] dP(x) \tag{6}$$

The covariance (η_k) components in the decomposition are identified as the "residual" terms such as in equation 4. This follows directly from the fact that all other components in these equations are identified. Thus, for $k \in [1, \dots K]$, we can identify η_k as $\eta_k = \tau_{k-1} - \Delta_{k-1} - \pi_k \tau_k$.

2.4 A Comparison with Conventional Mediation Analysis with Multiple Causally Ordered Mediators

Our decomposition has an analog in the context of a mediation–based decomposition of the ATE with multiple ordered mediators, but differs from conventional mediation analysis in important ways. To illustrate the differences, consider a binary treatment, A, an outcome of interest, Y , and a vector of pretreatment covariates, X, and let $M_1, M_2, \dots M_K$ denote K causally ordered mediators, assuming that for any $i < j$, M_i precedes M_j, as above. Moreover, let an overbar denote a vector of variables, so that$\overline{m}_k = (m_1, m_2, \dots m_k)$, and $\bar{a}_k = (a_1, a_2, \dots a_k)$. Using the potential outcomes notation as above, we can define the following expectation of a nested counterfactual,

$$\psi_{a\bar{a}_k} \triangleq \mathbb{E}\left[Y(a, \bar{M}_k(\bar{a}_k)) \right],$$

where $\bar{M}_k(\bar{a}_k) \triangleq \big(\bar{M}_{k-1}(\bar{a}_{k-1}), M_k(a_k, \bar{M}_{k-1}(\bar{a}_{k-1})) \big), \forall k \in [K]$. Under Pearl's (2009) non-parametric structural equation model (NPSEM), the ATE of A on Y can be decomposed into $K + 1$ identifiable PSEs corresponding to each of the causal paths $A \to Y$ and $A \to M_k \rightsquigarrow Y$ $(k \in [K])$:

$$\text{ATE} = \psi_{\overline{1}} - \psi_{\overline{0}} = \underbrace{\psi_{1,\overline{0}_K,} - \psi_{\overline{0}_{K+1}}}_{A \to Y} + \sum_{k=1}^{K} \underbrace{\big(\psi_{\overline{1}_{k+1},\underline{0}_{k+1}} - \psi_{\overline{1}_k,\underline{0}_{k+1}} \big)}_{A \to M_k \rightsquigarrow Y}. \tag{7}$$

This decomposition holds algebraically when mediators are not monotonic in nature. Yet, the monotonic characteristic of our proposed decomposition leads to a number of important differences between the two decompositions. First, for general mediation settings, the PSE decomposition is not algebraically unique, and thus the PSEs defined under alternative decompositions will differ if the effects of the treatment and each mediator differs across levels of the other mediators. In fact, depending on the order in which the paths $A \rightarrow Y$ and $A \rightarrow M_k \rightsquigarrow Y$ are considered, there are $(K+1)!$ identifiable different ways of decomposing the ATE; the decomposition shown in Equation 7 is just one such decomposition. Consider the case of two causally dependent mediators. In this setting, the causal pathway $A \rightarrow M_2 \rightsquigarrow Y$ can be defined with respect four different combinations of treatment and mediator 1 status: under (i) $A = 1$ and $M_1(a) = 1$, (ii) $A = 1$ and $M_1(a) = 0$, (iii) $A = 0$ and $M_1(a) = 1$, or (iv) $A = 0$ and $M_1(a) = 0$. In our setting, however, the decomposition shown is in fact the unique PSE decomposition of the ATE under monotonicity. This is a direct consequence of the monotonicity assumption: under monotonicity, however, $M_2(1, M_1(0)) = M_2(0, M_1(1)) = M_2(0, M_1(0)) = 0$, and thus, only the first definition of the PSE $A \rightarrow M_2 \rightsquigarrow Y$ is defined.

Second, for general PSE decompositions, the identifiable decompositions are merely a small subset of the total number of decompositions algebraically possible. In particular, the identifiable decomposition does not enable us to disentangle the mediating effects of M_k that are direct (net of subsequent mediators) and indirect (through different combinations of subsequent mediators). This restriction is a result of the fact that a potential outcome is identified (in expectation) only if the value that a mediator M_k takes, i.e., $M_k(a_k)$, is carried over to all future mediators. For example, in the case of 2 causally dependent mediators, to assess the mediating role of M_1, only the composite path $A \rightarrow M_1 \rightsquigarrow Y$ is identified, and not the individual paths $A \rightarrow M_1 \rightarrow Y$, and $A \rightarrow M_1 \rightarrow M_2 \rightarrow Y$ (as before, the squiggle arrow encompasses all possible causal paths from M_1 to Y). By contrast, in the monotonic setting, each PSE is in fact separately identifiable.This results from the fact that many non-identifiable paths in the general causal mediation context simply do not exist under monotonicity. In particular, in the case of 2 causally dependent mediators, the causal path $A \rightarrow M_2 \rightarrow Y$ does not exist. and each of the paths $A \rightarrow Y$, $A \rightarrow M_1 \rightarrow Y$ and $A \rightarrow M_1 \rightarrow M_2 \rightarrow Y$ are identifiable. Figure 1 illustrates the causal pathways defined under our decomposition in the case of two monotonic mediators.

Finally, the sequential ignorability assumption required to identify our decomposition is weaker than the ignorability assumption required in order to identify the PSE decomposition of the ATE. Specifically, the latter requires Pearl's (2009) NPSEM, which stipulates that, for each $M_k, k \in [K]$, conditional on antecedent mediators $\overline{M}_{k-1}$and the treatment A, $M_k \perp\!\!\!\perp (A, \overline{M}_{k-1})|X)$. This assumption, sometimes referred to as the "cross-world" independence assumption, is stronger than sequential ignorability because it does not allow for confounders of the $M_k \rightarrow Y$ effect, be they observed or unobserved. By contrast, our proposed decomposition accommodates observed intermediate confounding without altering the substance of the decomposition.[4] Finally, we also note that there are several additional differences between the two decompositions, which we summarize in Table 1.

Table 1: Comparison between proposed decomposition for monotonic mediators with path-specific effect (PSE) decomposition of mediating effects.

	Monotonic Mediator Decomposition	**Path-Specific Effect (PSE) Decomposition (Zhou, 2020)**
Identifiable Decomposition	$A \rightarrow Y + \sum_{k=1}^{K} A \rightarrow M_1 \ldots \rightarrow M_k \rightarrow Y$ (Unique)	$A \rightarrow Y + \sum_{k=1}^{K} A \rightarrow M_k \rightarrow Y$ (Not unique)
Ignorability Assumption	Sequential Ignorability (✓ observed intermediate confounders)	Cross-World Independence (✗ observed intermediate confounders)
Mediators	Binary	Multivariate / Continuous
Mediator Terms	Decomposable (Δ_k, π_k and η_k)	Not decomposable

[4] In this regard, our proposed decomposition is closer in its identification assumptions to the controlled direct and mediation effects (CDE, CMEs), another estimand considered in the mediation literature (e.g. Vanderweele et al 2014) which measures the strength of the causal relationship between a treatment and outcome when a mediator is fixed at a given value for all units.

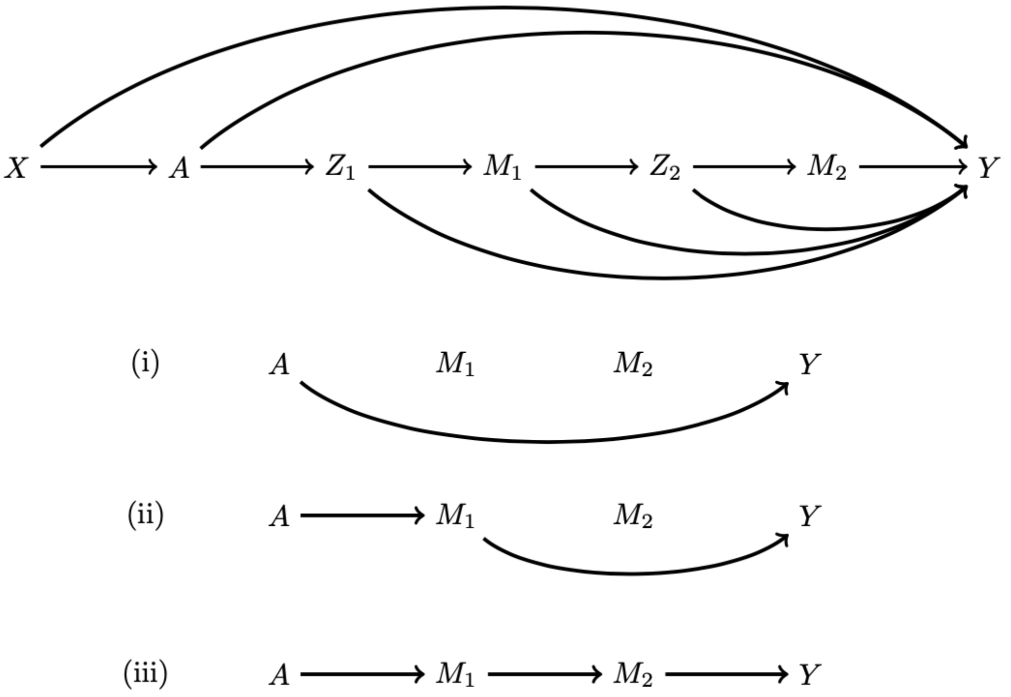

Figure 1: Causal Relationships with Two Monotonic Mediators Shown in a Directed Acyclic Graph (DAG) and the 3 Monotonic Path Specific Effects (MPSEs).

3 Estimation

3.1 Regression-With-Residuals Estimation

Our identification formulae suggest that we can estimate our proposed decomposition via several approaches, including outcome-based modeling, models for the treatment and mediators via inverse probability weighting (VanderWeele 2009), as well as doubly robust approaches. In this section, we show how a linear regression-with-residuals (RWR) approach can be taken for the MPSE decomposition. This approach relies on estimating three (sets of) linear models. The first model is simply a model for the ATE, and can be written as

$$\mathbb{E}[Y \mid X, A] = \lambda_0 + \lambda_1 A + \alpha_1^T X^{\perp} + \alpha_2^T A X^{\perp}, \tag{8}$$

where $X^{\perp} = X - \mathbb{E}[X]$. The second set of models is for the conditional mean of the outcome given the treatment, baseline confounders, k mediators, and k posttreatment confounders for all $k \in \{1, \ldots, K\}$, i.e.

$$\begin{aligned}\mathbb{E}[Y, A, \overline{Z}_k, \overline{M}_k] =& \beta_0 + \beta_1 A + \sum_{j=1}^{k} \mu_{j+1}\beta_j + \eta_1^T X^\perp + \sum_{j=1}^{k} \eta_{j+1}^T Z_j^\perp + \gamma_1 A X^\perp \\ &+ \sum_{j=1}^{k-1} \gamma_{j+1} M_j X^\perp + \sum_{j=1}^{k-1} M_j \sum_{l=1}^{j} \gamma_{k-1+j+l} Z_l^\perp,\end{aligned} \tag{9}$$

where $X^\perp$ is as above, $Z_k^\perp = M_{k-1}\big[Z_k - \mathbb{E}[Z_k \mid X, \overline{Z}_{k-1}, M_{k-1} = 1]\big]$, $M_0 = A$, and $M_k = 0$ if $M_{k-1} = 0, \forall k \in \{1, \ldots, K\}$, as described in Section X. This model differs from conventional linear regression in that (i) pre-treatment variables are centered around their marginal means, and (ii) post-treatment confounders $Z_k \forall k \in \{1, \ldots K\}$ have been centered around their conditional means given all antecedent variables. The second model is for the conditional mean of the mediators given treatment and pre-treatment confounders, i.e:

$$\mathbb{E}[M_{k+1} \mid X, A, \overline{Z}_k, \overline{M}_k] = \theta_0 + \delta_1^T X^\perp + \sum_{j=1}^{k} \delta_{j+1}^T Z_j^\perp, \tag{10}$$

where $X^\perp$ and $Z_k^\perp$ as are above, and where $M_0 = A$. Under assumptions X and assuming that the outcome and mediators are linear in pre- and post-treatment confounders, the treatment, and prior mediators, then the ATE τ_0 can be obtained from the linear model $\mathbb{E}[Y \mid X, A]$ as λ_1, and coefficients from the models $\mathbb{E}[Y \mid X, A, \overline{Z}_k, \overline{M}_k]$ and $\mathbb{E}[M_{k+1} \mid X, A, \overline{Z}_k, \overline{M}_k]$ yield estimates of components of the decomposition as follows:

$$\tau_k = \mathbb{E}[Y(\overline{1}_{k+1}) - Y(\overline{1}_k, 0)] = \beta_{k+1},$$

$$\Delta_{k-1} = \mathbb{E}[Y(\overline{1}_{k+1}, \underline{0}_{k+2}) - Y(\overline{1}_k, \underline{0}_{k+1})] = \begin{cases} \beta_k & \forall k \in \{2, \ldots, K\}, \\ \beta_1 & \text{for } k = 1 \end{cases}$$

$$\pi_{k+1} = \mathbb{E}[M(\overline{1}_{k+1})] = \theta_0.$$

RWR estimation of the decomposition can be implemented as follows:

1. For each of the baseline confounders, compute $\hat{X}^{\perp} = X - \hat{\mathbb{E}}[X]$.

2. Fit $\hat{\mathbb{E}}[Y \mid X, A]$ using the linear specification shown above; an estimate of τ_0 is given by $\hat{\lambda}_1$.

3. For each set of post-treatment confounders Z_k, $k \in \{1, \ldots, K\}$, compute $Z_k^{\perp} = M_{k-1}\big[Z_k - \mathbb{E}[Z_k \mid X, \overline{Z}_{k-1}, M_{k-1} = 1]\big]$ where an overbar denotes a vector of variables such that $\overline{Z}_k = (Z_1, \ldots, Z_k)$, by fitting a regression of Z_k on X and $\overline{Z}_{k-1}$ among units with $M_{k-1} = 1$ and then calculating the residuals.

4. For each $k \in \{1, \ldots K\}$:

   (a) compute least squares estimates of equations 9 and 10, using estimates of $X^{\perp}$ and $Z_k^{\perp}$.

   (b) compute $\hat{\tau_k} = \hat{\beta}_{k+1}$, $\hat{\Delta}_{k-1} = \hat{\beta}_k$ if $k \in \{2, \ldots, K\}$, else $\hat{\Delta}_{k-1} = \hat{\beta}_1$, and $\hat{\pi}_{k+1} = \hat{\theta}_0$.

5. Compute the decomposition using $\hat{\tau}_k$, $\hat{\Delta}_k$ and $\hat{\pi}_{k+1}$, and estimating the covariance terms as $\hat{\eta}_k = \hat{\tau}_{k-1} - \hat{\Delta}_{k-1} - \hat{\pi}_k \hat{\tau}_k$.

These estimates are consistent under sequential ignorability; standard errors and confidence intervals can then be obtained via the non-parametric bootstrap.

3.2 Semiparametric Estimation

RWR is attractive because of its conceptual simplicity and ease of implication. However, In practice, when X and $\overline{Z}_K$ are high-dimensional, parametric estimators which require a user-defined specification of the data-generating process may suffer biases resulting from model misspecification. In order to reduce model dependency, in this section we provide a nonparametric estimation approach which draws on a debiased machine learning (DML) approach for estimating Δ_k, π_k and η_k, for all $k \in [K]$ (see Rotnitzky and Robins 2017, Chernozhukov et al. 2018). Our DML approach is characterized by two components: first, the use of a Neyman orthogonal estimating equation based on the efficient influence function (EIF) for the targeted parameter, which makes estimates of the parameter 'locally robust' to estimates of the nuisance function; second, the use of a K-fold cross-fitting algorithm (**?**). For our proposed decomposition, we are required to estimate only three kinds of quantity, Δ_k,π_k, and τ_k (since all other terms in the decomposition can be derived from these quantities), for which it suffices to estimate the $\psi_{a\overline{m}_K} \triangleq \mathbb{E}[T(a, \overline{m}_K)\, \forall (a,m) \in [0,1], \forall T \in (M_1 \dots M_K, Y)$.

We first note that $\psi_{a\overline{m}_K}$ can be written in terms of iterated expectations, as:

$$\psi_{a\overline{m}_K} = \mathbb{E}_X \mathbb{E}_{Z_1|X,a} \dots \mathbb{E}_{Z_K|X,a,\overline{Z}_{K-1},\overline{m}_{K-1}} \mathbb{E}[T|a, X, \overline{Z}_K, \overline{m}_K]],$$

and thus define $\mu_k\left(X, \bar{Z}_k\right)$ as

$$\mu_K\left(X, \bar{Z}_K\right) \triangleq \mathbb{E}\left[T \mid X, \bar{Z}_K, \overline{m}_K\right]$$

$$\mu_k\left(X, \bar{Z}_k\right) \triangleq \mathbb{E}\left[\mu_{k+1}\left(X, \bar{Z}_{k+1}\right) \mid X, a, \bar{Z}_k, \overline{m}_k\right], \quad k = K-1, \dots, 0,$$

where $K \equiv K-1$ when $T = M_K$. In the event that our identification strategy requires no distinct set of intermediate confounders Z_k for the k^{th} mediator, $\mu_k = \mu_{k+1}$. The EIF of $\psi_{a\overline{m}_K}$ is closely related to the efficient influence function (EIF) for the g-formula (Rotnitzky and Robins 2017):

$$\varphi_{a\overline{m}_K}(O) = \sum_{k=0}^{K} \varphi_k(O),$$

where $O = (X, A, \overline{Z}_K, \overline{M}_K, Y)$ denotes the observed data, and where

$$\varphi_0(O) = \mu_0(X) - \varphi_{a\overline{m}_K}$$

$$\varphi_k(O) = \frac{\mathbb{I}\left(A=a\right)}{p\left(a \mid X\right)} \left(\prod_{j=1}^{k-1} \frac{\mathbb{I}(M_j = m_j)}{p(m_j|x, a, \overline{z}_j, \overline{m}_{j-1})} \right) \left(\mu_k\left(X, \overline{Z}_k\right) - \mu_{k-1}\left(X, \bar{Z}_{k-1}\right)\right), \quad k \in [1 \dots, K]$$

$$\varphi_{K+1}(O) = \frac{\mathbb{I}\left(A=a\right)}{p\left(a \mid X\right)} \left(\prod_{j=1}^{K} \frac{\mathbb{I}(M_j = m_j)}{p(m_j|x, a, \overline{z}_j, \overline{m}_{j-1})} \right) \left(T - \mu_K\left(X, \bar{Z}_K\right)\right)$$

It can be shown that, in the event that our identification strategy requires no distinct set of intermediate confounders Z_k for the k^{th} mediator, $\varphi_{k-1}(O)$ simply drops out of the estimating equation.

The second component of the debiased-machine learning approach, 'sample-splitting', can then be implemented as follows:

1. Randomly split data into K folds: $\{S_1, ...S_k\}$.
2. For each fold S_k:
   (a) Use the remaining $(k-1)$ folds (training sample) to fit a flexible machine-learning model for each of the nuisance functions involved in the estimating equations. All models for the k^{th} transition should be fitted as a function of X and $\overline{Z}_k$ *only among respondents who have made the $k-1$ prior transitions*, and fitted values computed for *all respondents*. A similar logic applies to the k^{th} outcome model.

   (b) For each observation in k (estimation sample), use estimates of the above models to construct a set of signals for Δ_k π_k, and τ_k.
3. Across all subsamples S_1 through S_K, for all units, compute a set of signals (recentered EIFs) for η_k and the k^{th} continuation effect $\theta_k$via the following equations (see Appendix A for derivations):

$$\eta_k^{\text{r}\hat{\text{EIF}}} = \tau_{k-1}^{\text{r}\hat{\text{EIF}}} - \Delta_{k-1}^{\text{r}\hat{\text{EIF}}} - \hat{\tau}_k \cdot \pi_k^{\text{r}\hat{\text{EIF}}} - \hat{\pi}_k \cdot \tau_k^{\text{r}\hat{\text{EIF}}} + \hat{\pi}_k \hat{\tau}_k,$$

for $k \in \{1, \dots K\}$, and

$$\begin{aligned}\theta_k^{\text{r}\hat{\text{EIF}}} &= (\prod_{j=1}^{k} \hat{\pi}_j)\Delta_k^{\text{r}\hat{\text{EIF}}} + \sum_{j=1}^{k}(\prod_{l:l\neq j}^{k} \hat{\pi}_l)\hat{\Delta}_k \cdot \pi_j^{\text{r}\hat{\text{EIF}}} + (\prod_{j=1}^{k} \hat{\pi}_j)\eta_k^{\text{r}\hat{\text{EIF}}} + \sum_{j=1}^{k-1}(\prod_{l:l\neq j}^{k-1} \hat{\pi}_l)\hat{\eta}_k \cdot \pi_j^{\text{r}\hat{\text{EIF}}} \\ &\quad - k(\prod_{j=1}^{k} \hat{\pi}_j)\hat{\Delta}_k - (k-1)(\prod_{j=1}^{k-1} \hat{\pi}_j)\hat{\eta}_k,\end{aligned}$$

for $k \in \{1, \dots K\}$, with $\theta_0 = \Delta_0$.

resume Compute an estimate of the decomposition by averaging the estimated influence functions across all subsamples S_1 through S_K, for all units, and construct standard errors using the sample variance of the empirical analogs of the EIFs.

4 Empirical Illustrations

4.1 Effects of Marriage on Earnings via Divorce

We implement our proposed MPSE decomposition for our two running examples. We first draw on data from the National Longitudinal Survey of Youth 1979 (NLSY79) to examine the direct and indirect effects of marriage on later life earnings via subsequent divorce. The NLSY79 first surveyed a nationally-representative sample of young adults ages 14-22 in 1979 and has subsequently interviewed them annually or biennially. The NLSY79 is especially appropriate for such an analysis, since it follows birth cohorts from the early 1960s, enabling the examination of marital events and effects which typically unfold over a prolonged period.

We construct our treatment and mediator via respondents' self-reports of their (heterosexual) marital status in each survey wave. Specifically, we construct the treatment as an indicator for whether an individual has married and is still currently married, by age 35, and our mediator as an indicator for whether an individual had divorced by age 40. We also use respondents' self-reported marital status to determine widowhood and remarriage - individuals who experienced either of these events within the treatment-mediator period of interest (i.e. up to age 40) are excluded from the analyses. We also restrict the sample to individuals with a marital length of over 5 years, primarily for the purposes of constructing our intermediate confounders (see below). Thus, our analyses can be interpreted as an analysis of the effects of first marriage, and first divorce in early and mid-career, on their earnings. Annual earnings Y are defined as respondent's self-reported wage and salary income averaged over ages 45-50. The earnings and wage variables are adjusted for inflation to 2019 dollars using the personal consumption expenditures index (PCE). We use the natural logarithm of respondents average annual earnings over this period (inflation-adjusted to 2019 dollars), and accommodate respondents with zero earnings or missing earnings information by bottom-coding average annual earnings at 0 dollars and then adding a small constant ($1,000$ dollars) to the earnings variable before taking the log transformation.

We control for an extensive set of background characteristics (X) that prior research identifies as confounders of the effects of both marriage and divorce on later life earnings. We control for race/ethnicity and nativity by including indicator variables for whether a respondent is Hispanic, Black, or other, and for whether the respondent was born in the United States. Since education is positively associated with both marriage and earnings, but negatively associated with divorce (Cooke 2006; Schoen et al. 2002; Schoen, Rogers, and Amato 2006), we control for respondents' education via a categorical variable that measures whether the respondent had either completed a BA degree, completed an associates degree, attended some college, completed high-school, or not attained a high-school diploma by age 29. We also include controls for several other education-based measures of ability and behavior reported during childhood, including percentile score on the ASVAB test, high school GPA, an index of substance use, an index of delinquency, whether the respondent had any children by age 18), and peer and school-level characteristics (college expectation among peers, and three dummy variables denoting whether the respondent ever had property stolen at school, was ever threatened at school, and was ever in a fight at school. To reflect the fact that family background (for instance, parental income and family structure) has a lingering effect on both adult earnings and marital status (Chetty et al 2017, Hewitt et al. 2006; Tzeng 1992), we control for whether a parent lived with both biological parents during upbringing as well as for parental education, parental income and assets, and presence of a paternal figure. We also measure several pre-marital adult attributes of respondents, including the respondent's average earnings (adjusted to 2019 dollars with the PCE) and hourly wages in the four years prior to the beginning of their marriage (or, if never married, in the four years up to the respondent's 35th birthday), respondents' expressed desire to have children, and gender role attitudes. In particular, we control for gender role attitudes by extracting the first two principal components of individuals' responses to eight questions on this topic in the 1982 survey.

birthday).[5] We also include an array of intermediate confounders of the divorce → earnings effect in our analyses. Several such confounders are defined with respect to a respondent's first marriage or marital partner: we control for whether the spouse was previously married as well as for an indicator for whether the respondent and their spouse cohabited before marriage, and the ages of each partner at the time of marriage. We include controls for several features of the marital household: the total family income of the household unit, divided by the square root of the number of family members in the household, spousal and respondent salary, the number of hours worked by the respondent's spouse, and the number of biological, step- or adopted children present in the household, top-coded at 4, to adjust for the negative association between children and (women's) wages (Budig and England 2001). These latter confounders are distinct in that they are measured with respect to a particular year, rather than being attributes of the marriage per se, and we therefore define them as averages in the 3 years prior to divorce, or, if a respondent had not divorced by age 40, in the 3 years prior to their 40th birthday. Since marital disruption is more likely earlier in marriage (Brines and Joyner 1999; Heckert et al. 1998; Rogers 2004; Schoen et al. 2002, 2006; Smock et al. 1999; South 2001), we also control for a marital duration, measured in years. Individuals who have not divorced by age 40 are assigned the length of their marriage up until this

We handle missing components of background characteristics (X) and intermediate variables (Z) using multivariate imputation via chained equations, with ten imputed data sets, and adjust standard errors of our parameter estimates using Rubin's (1987) method. Imputation-specific standard errors under the DML procedure are obtained via estimates of the influence functions for the decomposition components; under the RWR procedure, they are obtained via the nonparametric bootstrap with 5000 replications. All analyses are weighted using the NLSY79 custom weights. After constructing the analytical sample, we apply both the DML and RWR algorithms described in the previous section to implement our proposed decomposition. For the DML approach, we estimate all nuisance functions, using a super learner (Van der Laan et al. 2007) composed of Lasso and random forest and, following Chernozhukov et al. (2018), use five-fold cross-fitting. Appendix C gives further details about the

[5] We reverse-code items where required so that higher values always indicate the more conservative response.

particular models we are required to fit given our assumed data generation process. After multiple imputation, our analytic sample comprises $N = 8,176$ respondents.

Table 4 shows our estimates of the average total effect (ATE) on marriage on log family earnings and its direct and continuation vis divorce components under both our DML and RWR procedures. The first column shows that the estimated ATE of attending high-school on log earnings under DML (RWR) is 0.413 (0.528), which implies an earnings premium of $50 - 70\%$. The next two columns indicate that the overwhelming majority of the ATE operates directly, i.e. net of divorce: married individuals who do not subsequently divorce. Specifically, individuals who marry and do not divorce ($A \rightarrow Y$) earn on average 55% more than married individuals who go on to divorce. The gross effect of marrying via divorcing ($A \rightarrow M \rightarrow Y$) is estimated at -0.023 under DML, implying an earnings loss of approximately 20%, although under RWR, this effect is estimated as positive. In both instances, however, this gross effect is statistically insignificant. Table 5 shows estimates of the direct effects (Δ_0 and Δ_1), effect of marriage on divorce (π_0) and covariance term (η_1), under both the DML and RWR procedures. Under DML, we find that the net effect of divorce given marriage is in fact strongly negative, implying an income loss of approximately 23%; RWR also estimates this effect to be negative, although the estimate is significantly smaller.

Table 2: Decomposition of the Average Total Effect (ATE) of High-School Graduation on Logged Earnings via Debiased Machine-Learning (DML) and Regression-With-Residuals (RWR).

	ATE (τ_0)	$A \rightarrow Y$	A $\rightarrow M \rightarrow Y$
DML	0.413 (0.026)	0.436 (0.033)	-0.023 (0.019)
RWR	0.525 (0.026)	0.385 (0.092)	0.140 (0.087)

Note: Numbers in parentheses are estimates of standard errors, adjusted for multiple imputation via Rubin‘s (1987) method.

Table 3: Direct Effects (Δ_k), Probabilities (π_k) and Covariance Terms (η_k) Involved in Decomposition via Debiased Machine-Learning (DML) and Regression-With-Residuals (RWR).

	Δ_0	Δ_1	π_1	η_1
DML	0.436	-0.261	0.037	-0.013
	(0.033)	(0.500)	(0.029)	(0.003)
RWR	0.385	-0.019	0.066	0.141
	(0.092)	(0.075)	(0.087)	(0.004)

Note: Numbers in parentheses are estimates of standard errors, adjusted for multiple imputation via Rubin's (1987) method.

4.2 Educational Transition Effects

W next draw on data from the National Longitudinal Survey of Youth 1997 (NLSY97) to parse out the direct effect of high-school graduation on adult earnings and its indirect or continuation effects via (i) college attendance, (ii) college graduation, and (iii) graduate school attendance. Our analytic sample comprises $N = 8,649$ respondents.

We construct four types of variables: educational transitions, adult earnings, a set of confounders for the effect of high-school graduation on subsequent transitions and earnings, and a single set of intermediate confounders for the effect of college completion on subsequent transitions and earnings. Specifically, our educational transition variables contain a binary treatment denoting whether a respondent had graduated high school by age 21, and three binary mediators denoting whether the respondent had attended a four-year college by age 22, whether the respondent had received a BA degree by age 29, and whether the respondent had enrolled in a graduate level program by age 29, respectively. We assume that all individuals who make a given educational transition have made all previous educational transitions. Thus, we code a respondent as a high school graduate (i.e., $A = 1$) if the individual had either graduated high-school by age 22, attended a four-year college by age 22, received a BA degree by age 29, or attended graduate school by age 29, and as a high school dropout otherwise (i.e., $A = 0$). Similarly, we code a respondent as a college goer (i.e., $M_1 = 1$) if the person had either attended a four-year college by age 22, or had received a BA degree or attended graduate school by age 29, and as a high school graduate otherwise (i.e., $M_1 = 0$). Among college goers, we code a respondent as a college graduate (i.e., $M_2 = 1$) if the individual had received a BA degree by age 29, and as a college dropout/stop-out (i.e., $M_2 = 0$) otherwise. Finally, among college graduates, we code a respondent as a graduate school enrollee (i.e., $M_3 = 1$) if the individual had enrolled in an MA or higher degree by age 29, and as a college dropout (i.e., $M_3 = 0$) otherwise. Thus, by construction, our coding strategy disallows for cases which violate the monotonic transition assumption. In other words, we assume away cases in which an individual makes a particular educational transition without having made *all* previous transitions (e.g. if an individual enrolls in college without having completed high-school), an approach which we consider to be a reasonable approximation to reality. Annual earnings Y are defined as the sum of the respondent's self-reported wage and salary income and income from farms and businesses. The earnings and wage variables are adjusted for inflation to 2019 dollars using the personal consumption expenditures index (PCE). We use the natural logarithm of respondents average annual earnings at ages 30-33 (inflation-adjusted to 2019 dollars), and accommodate respondents with zero earnings or missing earnings information by bottom-coding average annual earnings at 0 dollars and then adding a small constant ($1,000$ dollars) to the earnings variable before taking the log transformation.

We include a large array of background characteristics (X) in our models for the effects of high-school attendance on labour market outcomes. These include basic demographic variables (gender, race, ethnicity, age at 1997), socioeconomic background (parental education, parental income, parental asset, co-residence with both biological parents, presence of a paternal figure, rural residence, southern residence), ability and behavior (percentile score on the ASVAB test, high school GPA, an index of substance use, an index of delinquency, whether the respondent had any children by age 18), and peer and school-level characteristics (college expectation among peers, and three dummy variables denoting whether the respondent ever had property stolen at school, was ever threatened at school, and was ever in a fight at school). In particular, parental education is measured using mother's years of schooling; when mother's years of schooling is unavailable, it is measured using father's years of schooling. Parental income is measured as the average annual parental income from 1997 to 2001. Both parental income and parental asset are transformed to 2019 dollars using the PCE.

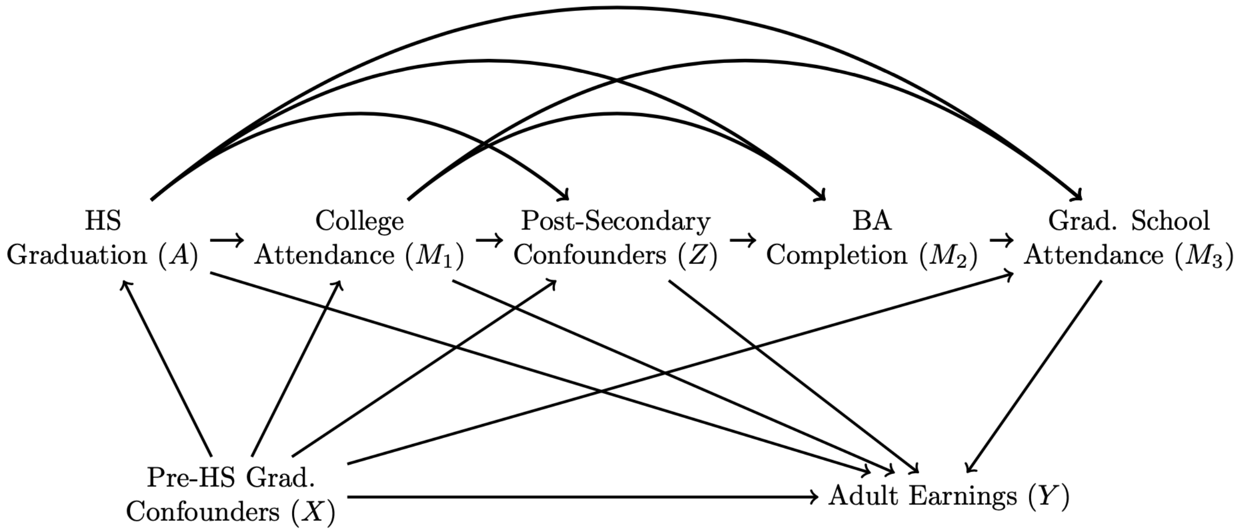
HS Graduation (A)
College Attendance (M_1)
Post-Secondary Confounders (Z)
BA Completion (M_2)
Grad. School Attendance (M_3)
Pre-HS Grad. Confounders (X)
Adult Earnings (Y)

Figure 2: A DAG showing the hypothesized causal relationships between high-school completion A and adult earnings Y via mediators M_1, M_2 and M_3.

Our proposed decomposition enables the inclusion of a distinct set of observed intermediate confounders for each transition to adjust for selection processes that may confound the causal effects of each transition on earnings (i.e., the $A - Y$ and $M_k - Y$ relationships, for $k \in \{1, \dots K\}$). Figure 2 summarizes our assumed data-generating process for this illustration. Here we assume it is plausible to treat a number of the Z_k sets as empty, and assume ignorability conditional on all antecedent variables alone. As such, we assume that the effect of the first transition (college attendance, M_1) on subsequent transitions and adult earnings is conditionally independent given high-school graduation (A) and background characteristics (X). By contrast, we do not assume away intermediate confounders for the causal effect of second transition (college completion M_2), and include a range of postsecondary characteristics (Z) in our models for BA completion and earnings to adjust for selection processes that may confound the causal effect of BA completion. These variables include college type, college quality, field of study, and college GPA. In each survey wave of the NLSY97, respondents were asked to report the names of the colleges in which they were currently or most recently enrolled, if any. Since many respondents attended more than one college, we focus on the college in which the respondent had been enrolled for the longest time by age 29. College type is a trichotomous variable denoting whether the college is a public institution, a private not-for-profit institution, or a for-profit institution. We employ a multi-dimensional measure of college quality that reflects not only admission selectivity, but also graduation rate and the college's record of helping low-income students move up the economic ladder. To gauge college selectivity, we use three dummy variables to denote whether the college is one of the "most competitive," "highly competitive," and "very competitive" colleges in Barron's Profile of American Colleges 2000. To measure graduation rate, we use the percentage of students graduating within six years measured in 2002 , which is available from the IPEDS database. In addition to college selectivity and graduation rate, we extract from the database of the Opportunity Insights project a measure of "upward mobility rate," i.e., the percentage of students who reach the top quintile of the income distribution among those with parents in the bottom quintile of the income distribution (Chetty et al. 2020a). In each survey wave, respondents who were currently or recently enrolled in college were also asked to report their major field of study. We use a dummy variable to denote whether the field of study in which the respondent had majored for the longest time by age 29 is a STEM field. Finally, college GPA is measured using the respondent's cumulative GPA from the Post-Secondary Transcript Study (PSTRAN). Finally, we assume that the effect of third transition (graduate school attendance, M_3) is ignorable conditional on (X, A, M_1, Z, M_2).[6]We handle missing components of background characteristics (X) and intermediate variables (Z) using multivariate imputation via chained equations, with ten imputed data sets, and adjust standard errors

[6] This assumption of ignorability of M_3 without conditioning on intermediate variables Z_3 is perhaps the strongest assumption we make. For example, many individuals take time off to work before enrolling in graduate school, and labour market experience and earnings gained in the interim period between college completion and graduate school enrollment may confound the latter variable's effects on earnings. Nevertheless, including a measure of labour market characteristics for this period is difficult because some respondents enrol directly in graduate school after BA completion. As a consequence, our analyses should be seen as illustrative of our proposed methodology rather than a definitive assessment of the effects in question.

of our parameter estimates using Rubin's (1987) method. After constructing the analytical sample, we apply both the DML and RWR algorithms described in Section X to implement our proposed decomposition. For the DML approach, we estimate all nuisance functions, using a super learner (Van der Laan et al. 2007) composed of Lasso and random forest and, following Chernozhukov et al. (2018), use five-fold cross-fitting. Appendix C gives further details about the particular models we are required to fit given our assumed data generation process.

Table 4 shows our estimates of the average total effect (ATE) on log earnings and its direct and continuation components under both our DML and RWR procedures. Both procedures return very similar estimates. The first column shows that the estimated ATE of attending high-school on log earnings under DML (RWR) is 1.375 (1.338), which implies an earnings premium of approximately 295%. The next two columns indicate that the vast majority of the ATE operates directly, i.e. net of college attendance, BA completion and graduate school attendance. Specifically, high-school graduates who do not go on to college ($A \rightarrow Y$) can be expected to earn approximately 220% more than high-school dropouts. The continuation effects of high-school graduation via college attendance, BA completion and graduate school attendance, respectively, constitute the remainder of the ATE. The continuation effect via college attendance among individuals who do not graduate from college ($A \rightarrow M_1 \rightarrow Y$) is very small, at approximately 3.0%, and fails to be statistically significant. By contrast, the continuation effect via college attendance and graduation, among individuals who do not go on to enroll in graduate school ($A \rightarrow M_1 \rightarrow M_2 \rightarrow Y$), is much higher, at about 0.19% under DML. The continuation effect via graduation ($A \rightarrow M_1 \rightarrow M_2 \rightarrow M_3 \rightarrow Y$) is also very small, and also fails to reach conventional levels of significance. Thus, we can conclude that the vast majority of the total effect of high-school graduation on earnings is determined in large part by the direct effect of high-school graduation, and also in small part by the continuation effect via BA completion.

Table 5 shows estimates of the various components (the direct effects (Δ_k), probabilities (π_k) and covariance terms (η_k)) that compose the continuation effects via college attendance, BA completion, and graduate school enrollment, under both t he DML and RWR procedures. Under DML, We find that, while the probability of BA completion given attendance, π_1 is 0.42, the direct effect of college attendance without completion (Δ_0) on earnings is slightly negative but statistically insignificant, which renders the continuation effect via college completion small and insignificant. Under RWR, Δ_0 is estimated as being much more positive than under the DML procedure, though the estimate is highly imprecise. Overall, given the slightly different point estimates of π_1 and η_1 under DML and RWR, the overall MPSE $A \to M_1 \to Y$ is estimated as being similar under both approaches.

Further, the direct effects of BA completion, without graduate school attendance, and graduate school attendance are large in terms of expected earnings, although RWR suggests lower effect sizes. Of all the covariance components, the covariance between BA completion and payoff is the strongest across both procedures. This finding is consistent with the comparative advantage thesis (Willis and Rosen 1979; Carneiro et al. 2011), which advances that individuals self-select into college education on the basis of their expected economic returns to investment, such that individuals who gain more education are in fact those who would experience the highest returns. The sum of the estimated direct effects for contiguous transitions can be interpreted as the joint effect of undergoing the transitions in question. For example, the sum of the four direct effects is 2.14, which captures the joint effect of undertaking all four educational transitions we consider. In other words, the earnings premium associated with graduating high school, attending and a completing a four-year college and enrolling in a graduate program as compared with not completing high-school is approximately 746.6% ($e^{2.136} - 1 = 7.44$) under DML. Finally, we note that, across components of the decomposition, RWR produces much larger standard errors than DML, which is to be expected given the asymptotic properties of DML.

Table 4: Decomposition of the Average Total Effect (ATE) of High-School Graduation on Logged Earnings via Debiased Machine-Learning (DML) and Regression-With-Residuals (RWR).

	ATE (τ_0)	$A \to Y$	A $\to M_1 \to Y$	$A \to M_1 \to M_2 \to Y$	$A \to M_1 \to M_2 \to M_3 \to Y$
DML	1.375 (0.061)	1.153 (0.064)	0.030 (0.026)	0.190 (0.019)	0.002 (0.004)
RWR	1.338 (0.077)	1.157 (0.081)	0.058 (0.066)	0.124 (0.068)	-0.001 (0.022)

Note: Numbers in parentheses are estimates of standard errors, adjusted for multiple imputation via Rubin‘s (1987) method.

Table 5: Direct Effects (Δ_k), Probabilities (π_k) and Covariance Terms (η_k) Involved in Decomposition via Debiased Machine-Learning (DML) and Regression-With-Residuals (RWR).

	Δ_0	Δ_1	Δ_2	Δ_3	π_1	π_2	π_3	η_1	η_2	η_3
DML	1.153	-0.003	0.737	0.249	0.424	0.268	0.286	0.031	0.252	-0.052
	(0.064)	(0.058)	(0.099)	(0.088)	(0.020)	(0.027)	(0.015)	(0.006)	(0.009)	(0.012)
RWR	1.157	0.124	0.349	0.181	0.373	0.496	0.224	0.012	0.159	-0.044
	(0.081)	(0.178)	(0.311)	(0.357)	(0.016)	(0.122)	(0.072)	(0.006)	(0.014)	(0.123)

Note: Numbers in parentheses are estimates of standard errors, adjusted for multiple imputation via Rubin‘s (1987) method.

5 Conclusion

Despite the large swathe of scientific interest and innovation in causal mediation in recent years, this body of work is implicitly concerned exclusively with settings where all values of a putative mediator are accessible to individuals irrespective of their treatment status. Given that many demographic actions, events and transitions often involve sequential transitions over the life course that are “monotonic” in nature, decomposition the total effect of an initial transition through subsequent monotonic transitions is evidently a research topic of strong interest. Yet, because typical approaches to causal mediation typically rely on variation in mediator statuses across treatment status, they are ill-equipped for exploring scientific questions with settings defined by mediator monotonicity.

To address this gap, in this paper, we have developed a causal mediation framework for analyzing the effects of monotonic transitions. First, we have demonstrated that the total effect of any level of education can be decomposed into a direct effect, net of K subsequent educational transitions, and K mutually exclusive "continuation" or gross effects: $A \to M_1 \cdots \to M_k \to Y$, which are all identifiable under the assumption of sequential ignorability, which allows for the effect of each educational transition to be confounded by a different set of variables. Next, we have introduced a non-parametric estimation strategy for our decomposition via the use of data-adaptive methods. We have highlighted how our mediation-based decomposition of the total effect of education compares with conventional mediation analysis with multiple mediators, in particular a decomposition of the total treatment effect into a set of path-specific effects (PSEs) (see Miles et al 2015, Zhou 2021). In particular, the monotonic nature of the mediators / transitions in our decomposition distinguishes it from conventional decompositions of the average treatment effect (ATE) into mutually exclusive path-specific effects (PSEs), properties which facilitate less restrictive identification assumptions as well as identification of all of the causal paths in the decomposition. In particular, an especially important advantage of our proposed decomposition is that it enables researchers to include a distinct set of intermediate confounders for each mediator, which is not the case in conventional mediation decompositions. While we have introduced our mediation decomposition in the context of a two particular substantive issues, our framework clearly applies more broadly to settings where mediators are monotonic in form. We hope that our decomposition will open the door to important research questions also characterized by monotonicity, which have thus far been limited by mediation estimands and methods targeted at strictly non-monotonic settings.

A Derivation of Efficient Influence Functions (EIFs)

A.1 Influence Function for the Covariance Components

Since the covariance component η_k is identified as $\eta_k = \tau_{k-1} - \Delta_{k-1} - \pi_k \tau_k$, by linearity of the EIF, η_k^{EIF} is given by

$$\eta_k^{\text{EIF}} = \tau_{k-1}^{\text{EIF}} - \Delta_{k-1}^{\text{EIF}} - (\pi_k \tau_k)^{\text{EIF}}$$

Letting P denote the true data distribution and $\mathcal{P}_t$ denote a parametric submodel of the form $\mathcal{P}_t = t 1_o(\tilde{o}) + (1-t)\mathcal{P}$, we can derive the EIF of $\pi_k \tau_k$ as follows:

$$\begin{aligned}(\pi_k \tau_k)^{\text{EIF}}(\mathcal{P}) &= \left.\frac{\partial \pi_k(\mathcal{P}_t)\tau_k(\mathcal{P}_t)}{\partial t}\right|_{t=0} \\ &= \pi_k^{\text{EIF}} \cdot \tau_k + \tau_k^{\text{EIF}} \cdot \pi_k.\end{aligned}$$

Thus, η_k^{EIF} can be written as

$$\begin{aligned}\eta_k^{\text{EIF}} &= \tau_{k-1}^{\text{EIF}} - \Delta_{k-1}^{\text{EIF}} - [\pi_k^{\text{EIF}} \cdot \tau_k + \tau_k^{\text{EIF}} \cdot \pi_k] \\ &= \tau_{k-1}^{\text{EIF}} - \Delta_{k-1}^{\text{EIF}} - \pi_k^{\text{EIF}} \cdot \tau_k - \tau_k^{\text{EIF}} \cdot \pi_k.\end{aligned}$$

This expression can then be rewritten in terms of recentered EIFs:

$$\begin{aligned}\eta_k^{\text{EIF}} = \eta_k^{\text{rEIF}} - \eta_k &= (\tau_{k-1}^{\text{rEIF}} - \tau_{k-1}) - (\Delta_{k-1}^{\text{rEIF}} - \Delta_{k-1}) - (\pi_k^{\text{rEIF}} - \pi_k) \cdot \tau_k - (\tau_k^{\text{rEIF}} - \tau_k) \cdot \pi_k \\ &= \tau_{k-1}^{\text{rEIF}} - \tau_{k-1} - \Delta_{k-1}^{\text{rEIF}} + \Delta_{k-1} - \tau_k \cdot \pi_k^{\text{rEIF}} + \pi_k \cdot \tau_k - \pi_k \cdot \tau_k^{\text{rEIF}} + \tau_k \cdot \pi_k,\end{aligned}$$

and thus

$$\begin{aligned}\eta_k^{\mathrm{rEIF}} &= \tau_{k-1}^{\mathrm{rEIF}} - \tau_{k-1} - \Delta_{k-1}^{\mathrm{rEIF}} + \Delta_{k-1} - \tau_k \cdot \pi_k^{\mathrm{rEIF}} + \pi_k \cdot \tau_k - \pi_k \cdot \tau_k^{\mathrm{rEIF}} + \tau_k \cdot \pi_k + (\tau_{k-1} - \Delta_{k-1} - \pi_k \tau_k) \\ &= \tau_{k-1}^{\mathrm{rEIF}} - \Delta_{k-1}^{\mathrm{rEIF}} - \tau_k \cdot \pi_k^{\mathrm{rEIF}} - \pi_k \cdot \tau_k^{\mathrm{rEIF}} + \tau_k \cdot \pi_k.\end{aligned}$$

The empirical analog of η_k^{rEIF} is therefore

$$\eta_k^{\mathrm{r\hat{E}IF}} = \tau_{k-1}^{\mathrm{r\hat{E}IF}} - \Delta_{k-1}^{\mathrm{r\hat{E}IF}} - \hat{\tau}_k \cdot \pi_k^{\mathrm{r\hat{E}IF}} - \hat{\pi}_k \cdot \tau_k^{\mathrm{r\hat{E}IF}} + \hat{\pi}_k \hat{\tau}_k.$$

A.2 Influence Function for the Continuation Effects

We first note that $\theta_0^{\mathrm{rEIF}} = \Delta_0^{\mathrm{rEIF}}$. Following the same logic as the above, we can derive the recentered influence functions for the continuation effects. For instance, since we can identify the first continuation effect θ_1 as

$$\theta_1 = \pi_1 \cdot \Delta_1 + \eta_1,$$

via the chain and product rules for the product term, we can write θ_1^{EIF} as

$$\theta_1^{\mathrm{EIF}} = \pi_1^{\mathrm{EIF}} \cdot \Delta_1 + \Delta_1^{\mathrm{EIF}} \cdot \pi_1 + \eta_1^{\mathrm{EIF}}.$$

Writing θ_1^{EIF} in terms of recentered influence functions and solving for θ_1^{rEIF} gives that

$$\begin{aligned}\theta_1^{\mathrm{rEIF}} &= \left(\pi_1^{\mathrm{rEIF}} \cdot \Delta_1 - \pi_1 \cdot \Delta_1 + \pi_1 \cdot \Delta_1^{\mathrm{rEIF}} - \Delta_1 \cdot \pi_1 + \eta_1^{\mathrm{rEIF}} - \eta_1\right) + \left(\pi_1 \cdot \Delta_1 + \eta_1\right) \\ &= \Delta_1 \cdot \pi_1^{\mathrm{rEIF}} + \pi_1 \cdot \Delta_1^{\mathrm{rEIF}} + \eta_1^{\mathrm{rEIF}} - \Delta_1 \cdot \pi_1.\end{aligned}$$

Following a similar logic, and after some algebraic manipulation, for the k^{th}continuation effect for $k \in 1, K$, we have that

$$\begin{aligned}\theta_k^{\mathrm{rEIF}} &= (\prod_{j=1}^{k} \pi_j)\Delta_k^{\mathrm{rEIF}} + \sum_{j=1}^{k}(\prod_{l:l\neq j}^{k} \pi_l)\Delta_k \cdot \pi_j^{\mathrm{rEIF}} + (\prod_{j=1}^{k} \pi_j)\eta_k^{\mathrm{rEIF}} + \sum_{j=1}^{k-1}(\prod_{l:l\neq j}^{k-1} \pi_l)\eta_k \cdot \pi_j^{\mathrm{rEIF}} \\ &- k(\prod_{j=1}^{k} \pi_j)\Delta_k - (k-1)(\prod_{j=1}^{k-1} \pi_j)\eta_k,\end{aligned}$$

with an empirical analog as follows:

$$\begin{aligned}\theta_k^{\mathrm{r\hat{E}\hat{I}F}} &= (\prod_{j=1}^{k} \hat{\pi}_j)\Delta_k^{\mathrm{r\hat{E}\hat{I}F}} + \sum_{j=1}^{k}(\prod_{l:l\neq j}^{k} \hat{\pi}_l)\hat{\Delta}_k \cdot \pi_j^{\mathrm{r\hat{E}\hat{I}F}} + (\prod_{j=1}^{k} \hat{\pi}_j)\eta_k^{\mathrm{r\hat{E}\hat{I}F}} + \sum_{j=1}^{k-1}(\prod_{l:l\neq j}^{k-1} \hat{\pi}_l)\hat{\eta}_k \cdot \pi_j^{\mathrm{r\hat{E}\hat{I}F}} \\ &- k(\prod_{j=1}^{k} \hat{\pi}_j)\hat{\Delta}_k - (k-1)(\prod_{j=1}^{k-1} \hat{\pi}_j)\hat{\eta}_k,\end{aligned}$$

B Derivation of RWR Procedures

For expositional clarity, we begin with the simpler single-mediator setting. Assume the following linear specification of the outcome model:

$$\mathbb{E}[Y \mid X, A, Z, M] = \beta_0 + \beta_1 A + \beta_2 M + \alpha_1^T X^{\perp} + \alpha_2^T Z^{\perp} + \alpha_3 X^{\perp} A,$$

where $X^{\perp} = X - \mathbb{E}[X]$ and $Z_1^{\perp} = Z_1 - \mathbb{E}[Z_1 \mid X, A = 1]$.

Then:

$$\begin{aligned}\mathbb{E}[Y(1,1) - Y(1,0)] &= \int \mathbb{E}[Y|x, A = 1, z, M = 1]dP(z_1|A = 1, x)dP(x) \\ &- \int \mathbb{E}[Y|x, A = 1, z, M = 0]dP(z_1|A = 1, x)dP(x) \\ &= \int [(\beta_0 + \beta_1 + \beta_2 + \alpha_1^T x^{\perp} + \alpha_2^T z^{\perp} + \alpha_3 x^{\perp})dP(z_1|A = 1, x)\end{aligned}$$

$$
\begin{aligned}
&- (\beta_0 + \beta_1 + \alpha_1^T x^\perp + \alpha_2^T z^\perp + \alpha_3 x^\perp) dP(z|A=0,x)]dP(x) \\
&= \int [\beta_0 + \beta_1 + \beta_2 + \alpha_1^T x^\perp + \alpha_2^T \mathbb{E}[Z - \mathbb{E}[Z \mid x, A=1]|x, A=1] \\
&- (\beta_0 + \beta_1 + \alpha_1^T x^\perp + \alpha_2^T \mathbb{E}[Z - \mathbb{E}[Z \mid x, A=1]|x, A=1])]dP(x) \\
&= \beta_0 + \beta_1 + \beta_2 + \alpha_1^T \mathbb{E}[X - \mathbb{E}[X]] + \alpha_3^T \mathbb{E}[X - \mathbb{E}[X]] \\
&- (\beta_0 + \beta_1 + \alpha_1^T \mathbb{E}[X - \mathbb{E}[X]] - \alpha_3^T \mathbb{E}[X - \mathbb{E}[X]]) \\
&= \beta_2.
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[Y(1,0) - Y(0,0)] &= \int \mathbb{E}[Y|x, A=1, z, M=0] dP(z|A=1,x) dP(x) \\
&- \int \mathbb{E}[Y|x, A=0, z, M=0] dP(z|A=0,x) dP(x) \\
&= \int [(\beta_0 + \beta_1 + \alpha_1^T x^\perp + \alpha_2^T z^\perp) dP(z|A=1,x) \\
&- (\beta_0 + \alpha_1^T x^\perp + \alpha_2^T z^\perp) dP(z|A=0,x)]dP(x) \\
&= \int [\beta_0 + \beta_1 + \alpha_1^T x^\perp + \alpha_2^T \mathbb{E}[Z - \mathbb{E}[Z \mid x, A=1]|x, A=1] \\
&- (\beta_0 + \alpha_1^T x^\perp + \alpha_2^T \mathbb{E}[Z - \mathbb{E}[Z \mid x, A=0]|x, A=0])]dP(x) \\
&= \beta_0 + \beta_1 + \alpha_1^T \mathbb{E}[X - \mathbb{E}[X]] + \alpha_3^T \mathbb{E}[X - \mathbb{E}[X]] \\
&- (\beta_0 + \beta_1 + \alpha_1^T \mathbb{E}[X - \mathbb{E}[X]] - \alpha_3^T \mathbb{E}[X - \mathbb{E}[X]]) \\
&= \beta_1.
\end{aligned}
$$

For simplicity, throughout the following we let $Z_0 = X$ and $M_0 = A$. We assume the following linear specification of the outcome model:

$$
\mathbb{E}[Y \mid \overline{Z}_k, A, \overline{M}_k] = \beta_0 + \sum_{j=0}^{k} \beta_{j+1} M_j + \sum_{j=0}^{k-1} \eta_{j+1}^T Z_j^\perp + \sum_{j=0}^{k-1} M_j \sum_{l=0}^{j} \eta_{j+l+1} Z_l^\perp, \tag{11}
$$

where $Z_k^\perp = Z_k - \mathbb{E}[Z_k | \overline{Z}_{k-1}, M_{k-1} = 1_{k-1}], \forall k \in [0, \ldots, K]$. In the following derivations, we use the fact thats, $\forall k \in \{1, \dots, K\}$,

$$\begin{aligned}
&\int z_k^{\perp} dP(z_j|\overline{z}_{k-1}, m_{k-1} = 1) \\
&= \mathbb{E}[Z_k - \mathbb{E}[Z_k|\overline{z}_{k-1}, m_{k-1} = 1]|\overline{z}_{k-1}, m_{k-1} = 1] \\
&= 0.
\end{aligned}$$

and

Letting $X = Z_0$, the above also implies that $\int z_0^{\perp} dP(z_0) = \mathbb{E}[Z_0 - \mathbb{E}[Z_0] = 0$. Under sequential ignorability and assuming linearity of the outcome with respect to all antecedent variables, we observe that $\hat{\beta}_{k+1}$ in equation 11 is consistent for $\mathbb{E}[Y(\overline{1}_{k+1}) - Y(\overline{1}_k, 0)]$ since span{$Z^{\perp}$}$\subset$ span{X,A}$^{\perp} \implies$ span$\{[1, A, AM_k, X^{\perp}, AZ^{\perp}]\}$ = span$\{[1, A, AM_k, X, AZ]\}$. Then, we have that

$$\begin{aligned}
\Delta_{k-1} &= \int \mathbb{E}[Y|\overline{M}_{k-1} = \overline{1}_{k-1}, \overline{z}_k, M_k = 0] \prod_{j=0}^{k} dP(z_j|\overline{z}_{j-1}, m_{j-1} = 1) \\
&- \int \mathbb{E}[Y|\overline{M}_{k-2} = \overline{1}_{k-2}, \overline{z}_k, M_{k-1} = 0] \prod_{j=1}^{k} dP(z_j|\overline{z}_{j-1}, m_{j-1} = 1) \\
&= \int \Big[\beta_0 + \sum_{j=0}^{k-1} \beta_{j+1} + \sum_{j=0}^{k} \eta_{j+1}^T z_j^{\perp} + \sum_{j=0}^{k-1} M_j \sum_{l=0}^{j} \eta_{j+l+1} Z_l^{\perp} \\
&- (\beta_0 + \sum_{j=0}^{k-2} \beta_{j+1} + \sum_{j=0}^{k} \eta_{j+1}^T Z_j^{\perp} + \sum_{j=0}^{k-2} M_j \sum_{l=0}^{j} \eta_{j+l+1} Z_l^{\perp})\Big] \prod_{j=0}^{k} dP(z_j|\overline{z}_{j-1}, m_{j-1} = 1) \\
&= \beta_0 + \sum_{j=0}^{k-1} \beta_{j+1} - (\beta_0 + \sum_{j=1}^{k-2} \beta_{j+1}^T) \\
&= \beta_k.
\end{aligned}$$

Further, for $\tau_k \forall k \in \{1, \dots, K\}$ we have that

$$
\begin{aligned}
\tau_k &= \int \mathbb{E}[Y|A=1, \overline{M}_k = \overline{1}_k, \overline{z}_k] \prod_{j=0}^{k} dP(z_j|\overline{z}_{j-1}, m_{j-1}=1) \\
&- \int \mathbb{E}[Y|A=1, \overline{M}_{k-1} = \overline{1}_{k-1}, \overline{z}_k, M_k = 0] \prod_{j=0}^{k} dP(z_j|\overline{z}_{j-1}, m_{j-1}=1) \\
&= \int \Big[(\beta_0 + \sum_{j=0}^{k} \beta_{j+1} + \sum_{j=0}^{k} \eta_{j+1}^T z_j^{\perp} + \sum_{j=0}^{k-1} M_j \sum_{l=0}^{j} \eta_{j+l+1} Z_l^{\perp}) \\
&- (\beta_0 + \sum_{j=0}^{k-1} \beta_{j+1} + \sum_{j=0}^{k} \eta_{j+1}^T z_j^{\perp} + \sum_{j=0}^{k-1} M_j \sum_{l=0}^{j} \eta_{j+l+1} Z_l^{\perp}) \Big] \prod_{j=0}^{k} dP(z_j|\overline{z}_{j-1}, m_{j-1}=1) \\
&= \beta_0 + \sum_{j=0}^{k} \beta_{j+1} - (\beta_0 + \sum_{j=1}^{k-1} \beta_{j+1}^T) \\
&= \beta_{k+1}.
\end{aligned}
$$

Finally, we assume that

$$
\mathbb{E}[M_{k+1} \mid A=1, \overline{Z}_k, \overline{M}_k = \overline{1}_k] = \theta_0 + \sum_{k=0}^{k} \delta_{k+1}^T Z_k^{\perp} \quad .
$$

Then:

$$
\begin{aligned}
\mathbb{E}[M(\overline{1}_{k+1})] &= \int \Big[\theta_0 + \sum_{k=0}^{k} \delta_{k+1}^T Z_k^{\perp} \Big] \prod_{j=0}^{k} dP(z_j|\overline{z}_{j-1}, m_{j-1}=1) \\
&= \theta_0.
\end{aligned}
$$

C Description of EIFs Used in Empirical Illustration

For each component involved in the MPSE, we construct a Neyman-orthogonal "signal" using its EIF, whose exact form depends on whether each set of intermediate confounders is empty or not. For our illustration of the effect of marriage on earnings, we include a distinct set of confounders for both the treatment A as well as for M. Our recentered EIFs for each component in marital effect decomposition are shown below:

$$
\begin{aligned}
M^*(1) &= \gamma_1(X) + \frac{\mathbb{I}(A=1)}{\pi_0(X,1)}(M - \gamma_1(X)), \\
Y^*(a) &= \mu_0(X,a) + \frac{\mathbb{I}(A=a)}{\pi_0(X,a)}(Y - \mu_0(X,a)), \text{ for } a \in \{0,1\} \\
Y^*(1,m_1) &= \nu_1(X,m) + \frac{\mathbb{I}(A=1)\mathbb{I}(M=m)}{\pi_0(X,1)\pi_1(X,m_1)}(Y - \mu_1(X,Z,m)) \\
&\quad + \frac{\mathbb{I}(A=1)}{\pi_0(X,1)}(\mu_1(X,Z,m) - \nu_1(X,m)), \text{ for } m \in \{0,1\}
\end{aligned}
$$

where

$$
\begin{aligned}
\pi_0(X,a) &\triangleq \Pr[A = a \mid X] \\
\pi_1(X,m_1) &\triangleq \Pr[M = m \mid X, A = 1] \\
\gamma_1(X) &\triangleq \mathbb{E}[M \mid X, A = 1] \\
\mu_0(X,a) &\triangleq \mathbb{E}[Y|X, A = a] \\
\mu_1(X,Z,m) &\triangleq \mathbb{E}[Y|X, A = 1, Z, M = m] \\
\nu_1(X,m) &\triangleq \mathbb{E}[\mu_1(X,Z,m)|X, A = 1]
\end{aligned}
$$

Figure 2 in the main text shows a potential data-generating process for the direct and indirect (continuation) effects of high-school graduation on adult earnings, via three transitions: college attendance (M_1), BA completion (M_2), and graduate school attendance (M_3). We assume that a set of pre-college characteristics serve as confounders for the $A-(M_1, M_2, M_3, Y)$ relationships, and that a set of post-secondary confounders Z confound the $M_2-(M_3, Y)$ relationships.

Under these assumptions for the various sets of confounders, our MPSE decomposition implies that, in the case of four transitions (one treatment and three mediators), it suffices to estimate the following three sets of parameters: (i) four direct effects $\Delta_k, k \in [0\ldots,3]$, where $\Delta_3 = \tau_3$, (ii) four gross effects $\tau_k, k \in [0\ldots,3]$, where $\tau_0 = \text{ATE}$, (iii) three mediator terms, $\pi_k, k \in [1\ldots,3]$.

All components in the three-mediator decomposition can then be estimated as functions of these parameters. For each of these target parameters, we construct a Neyman-orthogonal signal using its efficient influence function. Because of our assumed data-generating process, which maintains that there is only a single set of intermediate confounders (as opposed to a separate set of confounders for each mediator), the EIF for each estimand involved in the decomposition simplifies somewhat. Specifically, the recentered EIFs for each component in the decomposition are shown below:

$$\begin{aligned}
M_1^*(1) &= \gamma_1(X) + \frac{\mathbb{I}(A=1)}{\pi_0(X,1)}(M_1 - \gamma_1(X)),\\
M_2^*(1,1) &= \gamma_2(X) + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)}{\pi_0(X,1)\pi_1(X,1)}(M_2 - \gamma_2(X)),\\
M_3^*(1,1,1) &= \mathbb{E}[\gamma_3(X,Z)|X,A=1,M_1=1]\\
&\quad + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)}{\pi_0(X,1)\pi_1(X,1)}(\gamma_3(X,Z) - \mathbb{E}[\gamma_3(X,Z)|X,A=1,M_1=1])\\
&\quad + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)\mathbb{I}(M_2=1)}{\pi_0(X,1)\pi_1(X,1)\pi_2(X,Z,1)}(M_3 - \gamma_3(X,Z)),\\
Y^*(a) &= \mu_0(X,a) + \frac{\mathbb{I}(A=a)}{\pi_0(X,a)}(Y - \mu_0(X,a)),\ \text{for } a \in \{0,1\}\\
Y^*(1,m_1) &= \mu_1(X,m_1) + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=m_1)}{\pi_0(X,1)\pi_1(X,m_1)}(Y - \mu_1(X,m_1)),\ \text{for } m_1 \in \{0,1\}\\
Y^*(1,1,m_2) &= \mathbb{E}[\mu_2(X,Z,m_2)|X,A=1,M_1=1]\\
&\quad + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)}{\pi_0(X,1)\pi_1(X,1)}(\mu_2(X,Z,m_2) - \mathbb{E}[\mu_2(X,Z,m_2)|X,A=1,M_1=1])\\
&\quad + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)\mathbb{I}(M_2=m_2)}{\pi_0(X,1)\pi_1(X,1)\pi_2(X,Z,m_2)}(Y - \mu_2(X,Z,m_2)),\ \text{for } m_2 \in \{0,1\}\\
Y^*(1,1,1,m_3) =&\mathbb{E}[\mu_3(X,Z,m_3)|X,A=1,M_1=1]\\
&\quad + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)}{\pi_0(X,1)\pi_1(X,1)}\left(\mu_3(X,Z,m_3) - \mathbb{E}[\mu_3(X,Z,m_3)|X,A=1,M_1=1]\right)\\
&\quad + \frac{\mathbb{I}(A=1)\mathbb{I}(M_1=1)\mathbb{I}(M_2=1)\mathbb{I}(M_3=m_3)}{\pi_0(X,1)\pi_1(X,1)\pi_2(X,Z,1)\pi_3(X,Z,m_3)}\left(Y - \mu_3(X,Z,m_3)\right)\ \text{for } m_3 \in \{0,1\},
\end{aligned}$$

where

$$
\begin{aligned}
\pi_0(X, a) &\triangleq \Pr[A = a \mid X] \\
\pi_1(X, m_1) &\triangleq \Pr[M_1 = m_1 \mid X, A = 1] \\
\pi_2(X, Z, m_2) &\triangleq \Pr[M_2 = m_2 \mid X, A = 1, M_1 = 1, Z] \\
\pi_3(X, Z, m_3) &\triangleq \Pr[M_3 = m_3 \mid X, A = 1, M_1 = 1, Z, M_2 = 1] \\
\gamma_1(X) &\triangleq \mathbb{E}[M_1 \mid X, A = 1] \\
\gamma_2(X) &\triangleq \mathbb{E}[M_2 \mid X, A = 1, M_1 = 1] \\
\gamma_3(X, Z) &\triangleq \mathbb{E}[M_3 \mid X, A = 1, M_1 = 1, Z, M_2 = 1] \\
\mu_0(X, a) &\triangleq \mathbb{E}[Y | X, A = a] \\
\mu_1(X, m_1) &\triangleq \mathbb{E}[Y | X, A = 1, M_1 = m_1] \\
\mu_2(X, Z, m_2) &\triangleq \mathbb{E}[Y | X, A = 1, M_1 = 1, Z, M_2 = m_2] \\
\mu_3(X, Z, m_3) &\triangleq \mathbb{E}[Y | X, A = 1, M_1 = 1, Z, M_2 = 1, M_3 = m_3].
\end{aligned}
$$