Published in final edited form as:

Sociol Methods Res. 2010 May 1; 38(4): 515–544. doi:10.1177/0049124110366236.

Direct and indirect effects for neighborhood-based clustered and longitudinal data

T.J. VanderWeele

Dept. of Epidemiology, Harvard School of Public Health, 677 Huntington Ave., Boston, MA 02115,
tvanderw@hsph.harvard.edu

Abstract

Definitions of direct and indirect effects are given for settings in which individuals are clustered in groups or neighborhoods and in which treatments are administered at the group level. A particular intervention may affect individual outcomes both through its effect on the individual and by changing the group or neighborhood itself. Identification conditions are given for controlled direct effects and for natural direct and indirect effects. The interpretation of these identification conditions are discussed within the context of neighborhood research and multilevel modeling. Interventions at a single point in time and time-varying interventions are both considered. The definition of direct and indirect effects requires certain stability or no-interference conditions; some discussion is given as to how these no-interference conditions can be relaxed.

Keywords

Causal inference; direct and indirect effects; interference; longitudinal data; multilevel models; neighborhood effects; mediation; potential outcomes

1. Introduction

Our concern here will be with observational studies in which individuals are clustered in groups or neighborhoods and in which the causal effect of a particular cluster-level intervention is under consideration (e.g. Diez Roux et al., 1997; Sampson et al., 1997; Subramanian et al. 2002; Browning et al., 2006). We discuss how the potential outcomes framework (Neyman, 1923; Rubin 1974, 1978) extends to this setting and present the assumptions needed to estimate direct and indirect effects of a cluster- or neighborhood-level intervention. Our focus will be on the identification of direct and indirect effects and we will draw on recent causal inference theory developed for direct and indirect effects in the setting of non-clustered individual level data (Robins and Greenland, 1992; Pearl, 2001; Robins, 2003). This work circumvents many of the criticisms concerning the estimation of direct and indirect effects using structural equations modeling and path analysis (Bollen, 1987; Holland, 1988; Sobel, 1990; Winship and Morgan, 1999; Raudenbush and Sampson, 1999).

In particular, direct and indirect effects can be defined and estimated, and the total effect can be decomposed into a natural direct and indirect effect, even in settings involving interactions and non-linear models (Pearl, 2001; Kaufman et al., 2004; Joffe et al., 2007).

This is important because much of the work in the social sciences concerning mediation presupposes that there are no interactions between the effects of the treatment and the mediator on the outcome. The standard approach consists of regressing the outcome on the treatment, mediator and covariates (Baron and Kenny, 1986) to obtain a direct effect and then subtracting the direct effect from the total effect to obtain an indirect effect. This approach, however, in general only works when there are no interactions between the effects of the treatment and the mediator on the outcome. Kaufman et al. (2004; cf. Appendix of VanderWeele, 2009a) has shown when there are interactions or non-linearities, subtracting what is defined below as a controlled direct effect from a total effect does not in general give a quantity that can be interpreted as an indirect effect e.g. this difference may be non-zero even when there is no mediation and the true indirect effect is in fact zero. The work on direct and indirect effects in causal inference gives more general definitions of direct and indirect effects based on counterfactuals.

In this paper, we consider direct and indirect effects for interventions taking place at a single point in time and also extend our results to interventions that may change over time. The remainder of this paper is organized as follows. Section 2 provides an overview of how the potential outcomes framework relates to the setting of cluster-level interventions. Section 3 considers the identification of controlled direct effects. Section 4 considers the identification of natural direct and indirect effects. Section 5 discusses ways in which no-interference or SUTVA (stable unit treatment value assumption) conditions can be relaxed. Section 6 considers how the results in Sections 3 and 4 can be extended to data that is both clustered and longitudinal. Section 7 offers some concluding remarks.

2. Cluster-level Interventions

Let Y_{ik} denote the individual level outcome for individual i in group or neighborhood or cluster k . Let T_k denote the cluster-level treatment or intervention under consideration. Let X_{ik} denote individual level pretreatment characteristics for individual i in cluster k . Let V_k denote group or neighborhood-level pretreatment characteristics for cluster k . Note that the setting here differs somewhat from those considered elsewhere (Gitelman, 2005; Hong and Raudenbush, 2006) in educational research in which individuals are first assigned to classrooms and then classrooms are assigned to treatments so that there are in effect two stages of selection to consider: classroom assignment and treatment assignment. In this paper do not consider interventions to move individuals from one group or neighborhood to another as, for example, in the MTO study (Del Conte and Kling, 2001; Sobel, 2006; Kling et al., 2007) but rather interventions on the neighborhoods or clusters themselves.

Related work discusses the assumptions necessary estimate total causal effects of neighborhood or cluster-level interventions (VanderWeele, 2008a). See Gitelman (2005) and Hong and Raudenbush (2006) for related discussion. Specifically, let $Y_{ik}(t)$ denote the potential or counterfactual outcome individual i in cluster k would have obtained if, possibly contrary to fact, the cluster-level treatment, T_k , in cluster k were set to t . For the counterfactual variable $Y_{ik}(t)$ to be well-defined one must assume that an individual's outcome does not depend on the treatment assigned to clusters other than the individual's own cluster. We will refer to this assumption as the neighborhood level stable unit treatment

value assumption (or NL-SUTVA). Note that the NL-SUTVA assumption is a weaker condition than the standard no-interference assumption (Cox, 1958) in causal inference because treatment is administered at the cluster level rather than the individual level; see VanderWeele (2008a) for further discussion. For the counterfactual variable $Y_{ik}(t)$ to be well-defined it must furthermore be assumed that the interventions on the various neighborhoods do not change the cluster to which individuals belong. Hong and Raudenbush (2006, 2008) refer to this assumption as "intact clusters." For causal effects to be identified, VanderWeele (2008a) assumed that cluster-level treatment assignment is ignorable given the covariates X_{ik} and V_k ; i.e. that for $t = 0, 1$

$$Y_{ik}(t) \perp\!\!\!\perp T_k | X_{ik}, V_k \quad (1)$$

where $A \perp\!\!\!\perp B/C$ denotes the independence relation that A is independent of B conditional on C . Assumption (1) requires that within strata of the covariates X_{ik} and V_k , the distribution of the potential outcomes, $Y_{ik}(t)$, are independent of the treatment actually received, T_k ; i.e. that within strata of the covariates, the groups are comparable in their potential outcomes. The assumption is often interpreted as requiring that the covariate vectors, X_{ik} and V_k , include all variables that affect both the treatment and the outcome. VanderWeele (2008a) noted that an individual-level covariate need be included in the individual-level covariate vector, X_{ik} , only if it affects neighborhood-level treatment, T_k , in some way different than the neighborhood-level covariate equivalent in V_k ; e.g. if treatment decisions used information only on the *mean* of neighborhood income rather than on the *distribution* of income then mean income would have to be included in V_k but individual income would not have to be included in X_{ik} . A "consistency" assumption was also needed that $T_k = t \Rightarrow Y_{ik}(t) = Y_{ik}$ (Robins, 1986, 1987; Cole and Frangakis, 2009; VanderWeele, 2009b); i.e. that the outcome that would have been observed if treatment were set to what it in fact was is equal to the outcome that was in fact observed. Under these assumptions it was shown (VanderWeele, 2008a) that

$$E[Y_{ik}(t)] = \sum_{x,v} E[Y_{ik} | T_k=t, X_{ik}=x, V_k=v] P(X_{ik}=x, V_k=v). \quad (2)$$

The right hand side of equation (2) is given in terms of observed data and can thus be estimated by appropriate statistical models. Further discussion of NL-SUTVA and of the ignorability assumption (1) in the context of neighborhood effects research is given elsewhere (VanderWeele 2008a).

Multilevel (or hierarchical) modeling is often employed to model conditional expectations of the form $E[Y_{ik} | T_k = t, X_{ik} = x, V_k = v]$ when individuals are clustered in groups or neighborhoods. The typical model employed for a cluster-level intervention will have the following form:

$$Y_{ik} = \beta_0 + \beta_1 X_{ik} + \beta_2 V_k + \gamma T_k + u_k + \varepsilon_{ik} \quad (3)$$

where β_0 denotes an intercept, u_k denotes a neighborhood-level random error and ε_{ik} denotes an individual-level random error. The u_k and ε_{ik} are usually assumed to be independent and identically distributed with mean zero. That the ε_{ik} are independent of one another implies that the correlation between deviations from the mean that arises from subjects within the

same neighborhood is wholly attributable to the neighborhood random term u_k . Model (3) is also often written in hierarchical form with level 1 given by the equation:

$$Y_{ik} = \alpha_k = \beta_1 X_{ik} + \varepsilon_{ik} \quad (4)$$

and level 2 given by the equation:

$$\alpha_k = \beta_0 + \beta_2 V_k + \beta_3 T_k + u_k. \quad (5)$$

Substituting equation (5) into equation (4) gives equation (3). Estimation of the coefficients of (3) or equivalently of (4) and (5) is straightforward using standard statistical software. Equation (3) could of course also be generalized to include interaction terms. If model (3) holds then under assumption (1), $\mathbb{E}Y_{ik}(1) - \mathbb{E}Y_{ik}(0) = \gamma$; i.e. γ has a causal interpretation as the average causal effect for treatment $T_k = 1$ as compared with $T_k = 0$. If model (3) does not hold but some more complex model, possibly including interactions, for the observed data; i.e. for $E[Y_{ik}|T_k = t, X_{ik} = x, V_k = v]$, does hold then (2) can still be used for the estimation of causal effects $\mathbb{E}Y_{ik}(1) - \mathbb{E}Y_{ik}(0)$. Whenever a multilevel model, such as (3), is employed, it assumes a particular functional form and inference is only valid if the model is correctly specified.

The remainder of this paper considers the identification not simply of total causal effects but of direct and indirect effects. For purposes of illustration we will consider two examples. In the first, we will consider the effect of the construction of new hospitals, a neighborhood level intervention, on health outcomes. Let T_k denote the construction of a new hospital in neighborhood k in year r . Let Y_{ik} denote a continuous measure of health status for individual i in neighborhood k in year $r + 1$. Let X_{ik} denote a vector of individual level characteristics for individual i in neighborhood k including variables such as age, sex, race, socioeconomic status, and prior health status, all measured at the beginning of year r . Let V_k denote community-level characteristics for neighborhood k including variables such as number of prior hospitals, neighborhood safety and neighborhood mean income. We will consider a mediator, M_{ik} , that denotes whether individual i in neighborhood k has access to clinical care within 10 miles.

In the second example, we will let T_k be a measure of neighborhood policing in the past year and we will let the outcome, Y_{ik} , be an indicator for whether an individual was a victim of crime any time in the last year; the variables X_{ik} and V_k will once again denote individual and neighborhood characteristics respectively and we will consider a potential mediator, M_{ik} , that denotes the average number of hours per week in the past year that an individual spends walking throughout the neighborhood. Similar examples, outside the context of mediation can be found in Kaufman (2005) and Verbitsky and Raudenbush (2009).

3. Controlled Direct Effects for a Cluster-level Intervention

Now suppose that the individual health outcome Y_{ik} is affected by the construction of a new hospital both by improved health care access at the new hospital and through the changes that the new hospital brings about in the community. With regard to the latter, the construction of the new hospital may provide employment opportunities for the

neighborhood which may improve individual health outcomes through increasing salaries, lowering crime, etc. Let M_{ik} be an indicator that denotes whether individual i in neighborhood k has access to clinical care within 10 miles by some point in time after the hospital has been constructed (e.g. by 3 months after the hospital has been constructed). Note that depending on the size of the areas in a study, building a new hospital may or may not ensure access to clinical care within 10 miles for all individuals in the neighborhood. Suppose that by setting up a temporary clinic any individual could be given access to clinical care within 10 miles and thus that it is hypothetically feasible to intervene on M_{ik} . Suppose further that it were possible to intervene so as to deny an individual care at a newly constructed hospital so that we could potentially intervene on neighborhood k and individual i to set $T_k = 1$ and $M_{ik} = 0$. Throughout the remainder of this paper we will assume that some, possibly hypothetical, intervention on the mediator is available that can change the value of M_{ik} . If the mediator, M_{ik} , cannot be changed through an intervention, an alternative approach based on principal stratification may be employed (Frangakis and Rubin, 2002; Rubin, 2004; VanderWeele 2008b; Gallop et al., 2009).

Let $Y_{ik}(t, m)$ denote the counterfactual value of the health outcome for individual i in neighborhood k if T_k were set to t and if M_{ik} were set to m . Note that for this quantity to be well defined the health of individual i in neighborhood k must not depend on the access to a nearby hospital of any other individual in the same or any other neighborhood nor on the construction of hospitals in neighborhoods other than the individual's own. We will refer to this assumption as the individual and neighborhood level stable unit treatment value assumption. Note that this is a considerably stronger assumption than that of the neighborhood level SUTVA assumption considered above: a neighborhood-level no-interference assumption is required for treatment but a stronger individual-level no-interference assumption (Cox, 1958) is required for the mediator. In Section 5 we discuss how this assumption can be relaxed.

Consider the following two contrasts:

$$\mathbb{E}[Y_{ik}(1, 0) - Y_{ik}(0, 0)] \quad (6)$$

and

$$\mathbb{E}[Y_{ik}(1, 1) - Y_{ik}(0, 1)]. \quad (7)$$

The contrast given in (6) considers the causal effect of the hospital construction intervening so that the individual is not able to attain access to clinical care within 10 miles; the expression in (6) thus can be interpreted as the causal effect of the hospital on an individual that is not due to that individual's access to health care related to the new hospital. Similarly the expression in (7) considers the causal effect of the hospital construction intervening to ensure that the individual is able to attain access to clinical care within 10 miles. Note that contrasts of the form (6) and (7) consider interventions which fix both T_k and M_{ik} to particular values. Expressions such as (6) and (7) are thus referred to as "controlled direct effects" (Pearl, 2001) because contrast (6) and (7) represent the direct effect of hospital construction on health controlling for individual access; i.e. intervening to set individual

access to 0 or 1, respectively. More generally, define the controlled direct effect for individual i in cluster k by $CDE_{ik}^m(t, t') = Y_{ik}(t, m) - Y_{ik}(t', m)$. Note that the definition of controlled direct effects allows the magnitude of the direct effect to vary with m ; this is in contrast to much of the literature on direct effects in the structural equation modeling literature (e.g. Bollen, 1987) in which the direct effect is assumed homogenous for all levels of the mediator. In the following section controlled direct effects will be contrasted to "natural" direct and indirect effects and these remarks concerning relaxing homogeneity (i.e. allowing interaction between the effects of the treatment and mediator on the outcome) pertain also to natural direct and indirect effects. We might also be interested in contrasts of the form

$$\mathbb{E}[Y_{ik}(1, 1) - Y_{ik}(1, 0)] \quad (8)$$

and

$$\mathbb{E}[Y_{ik}(0, 1) - Y_{ik}(0, 0)]. \quad (9)$$

Expression (8) considers the average causal effect of access to clinical care within 10 miles intervening to ensure that a new hospital in the neighborhood is built; expression (9) considers the average causal effect of access to clinical care within 10 miles intervening to ensure that no new hospital is built.

The joint effect of the hospital construction and access to clinical care within 10 miles can be decomposed as follows:

$$Y_{ik}(1, 1) - Y_{ik}(0, 0) = \{Y_{ik}(1, 1) - Y_{ik}(1, 0)\} + \{Y_{ik}(1, 0) - Y_{ik}(0, 0)\}. \quad (10)$$

Note also that the decomposition given in (10) is a decomposition of the joint effect of T_k and M_{ik} rather than of simply the effect of hospital construction T_k alone. The decomposition given in (10) is not unique. For example, the joint effect could also be decomposed as $Y_{ik}(1, 1) - Y_{ik}(0, 0) = \{Y_{ik}(1, 1) - Y_{ik}(0, 1)\} + \{Y_{ik}(0, 1) - Y_{ik}(0, 0)\}$. In the next section when we consider natural direct and indirect effects we will provide a decomposition of the effect of hospital construction T_k alone rather than the joint effect of T_k and M_{ik} together.

We now turn to conditions for the identification of expressions (6)–(9). As noted above, in order to estimate the causal effects of a neighborhood level intervention, the ignorability assumption (1) must hold; that is to say X_{ik} and V_k must contain all individual and neighborhood level variables that confound the relationship between the neighborhood level treatment, T_k , and the outcome, Y_{ik} . In order to estimate contrasts of the form (6) and (7); i.e. controlled direct effects, we will require not only that the treatment-outcome relationship is unconfounded given X_{ik} and V_k but also that the mediator-outcome relationship is unconfounded given T_k , X_{ik} and V_k . Formally we will need the following conditions:

$$Y_{ik}(t, m) \perp\!\!\!\perp T_k \mid X_{ik}, V_k \quad (11)$$

and

$$Y_{ik}(t, m) \prod M_{ik} | T_k, X_{ik}, V_k. \quad (12)$$

Condition (11) requires that X_{ik} , V_k contain all the confounders of the treatment-outcome relationship; condition (12) requires that T_k , X_{ik} , V_k contain all the confounders of the mediator-outcome relationship. Depending on how rich the covariates sets X_{ik} and V_k are, assumptions (11) and (12) may or may not be reasonable. Causal directed acyclic graphs can be useful in assessing these assumptions (Pearl, 1995, 2001). If conditions (11) and (12) do hold then we will be able to estimate controlled direct effects of the form (6) and (7), as stated in the following theorem. Note we will assume throughout this paper that the consistency condition holds for $Y_{ik}(t, m)$; i.e. that $T_k = t, M_{ik} = m \Rightarrow Y_{ik}(t, m) = Y_{ik}$ (VanderWeele and Vansteelandt, 2009). The proofs of all theorems are given in the online appendix. These generally follow those for non-clustered data.

Theorem 1. If for some set of variables X_{ik} and V_k conditions (11) and (12) hold then the average controlled direct effect $\mathbb{E}[CDE_{ik}^m(t, t')] = \mathbb{E}[Y_{ik}(t, m)] - \mathbb{E}[Y_{ik}(t', m)]$ is identified and is given by

$$\begin{aligned} \mathbb{E}[CDE_{ik}^m(t, t')] = & \sum_{x,v} \{ \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t, M_{ik}=m] \\ & - \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t', M_{ik}=m] \} P(X_{ik}=x, V_k=v). \end{aligned} \quad (13)$$

Theorem 1 gives an empirical formula for expression (6) when $t = 1, t' = 0$ and $m = 0$ and for (7) when $t = 1, t' = 0$ and $m = 1$. Note that the expression on the left hand side of (13) is a causal quantity; the expression on the right hand side of equation (13) is given in terms of the observed data and could be estimated from a multi-level model which extends equation (3) so as to include M_{ik} such as the model

$$Y_{ik} = \beta_0 + \beta_1 X_{ik} + \beta_2 V_k + \gamma T_k + \theta M_{ik} + u_k + \varepsilon_{ik} \quad (14)$$

where once again u_k and ε_{ik} are assumed to be independent and identically distributed with mean zero. Under model (14) and the conditions (11) and (12), the controlled direct effect in (13) is given by $\mathbb{E}[CDE_{ik}^m(t, t')] = \gamma(t - t')$. As before, interaction terms could be included in the multi-level model given in (14) if appropriate and, using (13), the controlled direct effect could then be expressed as a different combination of the regression coefficients. If the outcome is not continuous, a generalized linear multilevel model (Lee and Nelder, 1996) could be employed in place of (14). In the neighborhood policing example, for instance, the outcome, Y_{ik} , was a binary variable indicating whether or not an individual was the victim of a crime in the past year. In estimating controlled direct effects from generalized linear multilevel models, equation (13) can be used but care must be taken to transform estimates using the inverse link function so that the expression $\mathbb{E}[Y_{ik} | X_{ik} = x, V_k = v, T_k = t, M_{ik} = m]$, rather than its transform, is used in (13). Furthermore for generalized linear multilevel models, estimates conditional on the cluster u_k will not in general equal effect estimates marginalized over u_k and when employing a generalized linear multilevel model to estimate controlled direct effects, it will be necessary to integrate over u_k in order to use expression (13). Standard errors for the expression in (13) could then be obtained by bootstrap

techniques. Alternatively, a marginal approach using generalized estimating equations could be used and such an approach has been advocated elsewhere on grounds of efficiency (Hubbard et al., 2009).

It can also be verified that if conditions (11) and (12) hold then expressions (8) and (9) are also identified from the data and:

$$\begin{aligned} \mathbb{E}[Y_{ik}(t, m)] - \mathbb{E}[Y_{ik}(t, m')] &= \sum_{x,v} \{ \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t, M_{ik}=m] \\ &\quad - \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t, M_{ik}=m'] \} P(X_{ik}=x, V_k=v). \end{aligned}$$

Note that both conditions (11) and (12) must hold in order to identify controlled direct effects. Condition (11) is very similar to condition (1) above for the identification of total effects. Condition (12) requires that mediator-outcome relationship is unconfounded given T_k , X_{ik} and V_k . Note that if there are unmeasured confounding variables of the mediator-outcome relationship so that condition (12) is violated, then estimates of the controlled direct effect will be biased (Judd and Kenny, 1981; Robins and Greenland, 1992; Pearl, 2001; Cole and Hernán, 2002).

In order to ensure that condition (12) holds, additional covariates, beyond those necessary to identify total effects, may need to be included in X_{ik} and V_k . Suppose, for example, that the area in which a hospital was to be constructed was just east of a particular town and that the west side of the town had more industrial factories and hence more air pollution. For simplicity suppose further that those living on the east side of the town would be within 10 miles of the new hospital and those on the west side would not. In this case, a variable indicating whether an individual lived on the east side or the west side of the town would need to be included in the individual level covariate vector, X_{ik} , because this variable would then affect both the mediator, M_{ik} (whether or not the individual had access to clinical care within 10 miles), and the health status outcome, Y_{ik} , because of air pollution. Note that in this example if the total effect, rather than the controlled direct effect, were of interest then the side of town that individuals lived in might not need to be included in the individual level covariate vector, X_{ik} , if information on which individuals lived on which side of the town was not used in making treatment decisions about the construction of the hospital.

In some cases the variables that confound the relationship between the mediator and outcome might be consequences of treatment (Robins and Greenland, 1992; Pearl, 2001). Denote the individual and cluster level covariates that confound the mediator-outcome relationship and are consequences of treatment by Z_{ik} and W_k respectively. In such cases, to identify controlled direct effects condition (12) can be modified to $Y_{ik}(t, m) \perp\!\!\!\perp M_{ik}/T_k, X_{ik}, V_k, Z_{ik}, W_k$; provided this condition holds, along with condition (11), controlled direct effects are still identified and given by $\mathbb{E}[Y_{ik}(t, m)] - \mathbb{E}[Y_{ik}(t, m')] = \sum_{x,v,z,w} \{ \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t, M_{ik}=m, Z_{ik}=z, W_k=w] - \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t, M_{ik}=m', Z_{ik}=z, W_k=w] \} P(Z_{ik}=z, W_k=w | X_{ik}=x, V_k=v, T_k=t) P(X_{ik}=x, V_k=v)$ (cf. Robins 1986, 1987; Pearl, 2001 for proofs of the result for non-clustered data; as with Theorem 1, the extension of the proof to the multilevel setting is straightforward). One would have to use the expression just given rather than that given in (13) if, for example, the building of the hospital led to

employment and higher incomes for some individuals which caused them to move from the west side to the east side of the town. In this case, the variable indicating the side of the town is both a consequence of treatment and confounds the relationship between the mediator and the outcome.

4. Natural Direct and Indirect Effects for a Cluster-level Intervention

In the previous section, we considered controlled direct effects of the form (6) and. These controlled direct effects represented the effect of neighborhood level treatment, T_k , on individual level outcomes, Y_{ik} , intervening to fix the individual level mediator, M_{ik} , to some level m . In contrast with controlled direct effects, natural direct effects fix the intermediate variable to the level it would have been under the presence or absence of treatment. Although perhaps of less direct policy relevance, these natural direct effects may be of interest in understanding the mechanisms through which particular outcomes arise and in assessing the extent to which the effect of treatment is mediated (Robins, 2003; Hafeman and Schwartz, 2009). Let $M_{ik}(t)$ denote the counterfactual outcome for the individual level mediator, M_{ik} , if, possibly contrary to fact, T_k had been set to t . The neighborhood level SUTVA condition for $M_{ik}(t)$ to be well-defined is that the individual level mediator, $M_{ik}(t)$, for individual i in cluster k does not depend on the value of the cluster-level intervention, $T_{k'}$, of any other neighborhood $k' \neq k$. The consistency condition for $M_{ik}(t)$ is $T_k = t \Rightarrow M_{ik}(t) = M_{ik}$. Consider the following expression:

$$\mathbb{E}[Y_{ik}(1, M_{ik}(0))] - \mathbb{E}[Y_{ik}(0, M_{ik}(0))]. \quad (15)$$

The expression given in (15) represents the effect of neighborhood level treatment, T_k , intervening to set the individual level mediator, M_{ik} , to what it would have been if T_k had in fact been set to 0. More generally define the natural direct effect for individual i in neighborhood k by $NDE_{ik}^{t'}(t, t^*) = Y_{ik}(t, M_{ik}(t')) - Y_{ik}(t^*, M_{ik}(t'))$.

In order to estimate natural direct and indirect effects we need, in addition to the ignorability conditions (11) and (12), two other ignorability conditions:

$$M_{ik}(t) \perp\!\!\!\perp T_k | X_{ik}, V_k \quad (16)$$

and

$$Y_{ik}(t, m) \perp\!\!\!\perp M_{ik}(t') | X_{ik}, V_k \quad (17)$$

Condition (16) states that the effect of neighborhood level treatment, T_k , on the individual level mediator, M_{ik} , is unconfounded given X_{ik} and V_k . Condition (17) essentially requires that there is no effect of neighborhood level treatment T_k which itself confounds the relationship between the individual level mediator, M_{ik} , and the outcome, Y_{ik} (Pearl, 2001). Depending on the specific study, assumptions (16) and (17) may or may not hold. Additional discussion is given below. If these assumptions do hold then we have the following Theorem.

Theorem 2. If conditions (11), (12), (16) and (17) hold then the average natural direct effect $\mathbb{E}[NDE_{ik}^{t'}(t, t^*)] = \mathbb{E}[Y_{ik}(t, M_{ik}(t'))] - \mathbb{E}[Y_{ik}(t^*, M_{ik}(t'))]$ is identified and is given by

$$\begin{aligned} \mathbb{E}[NDE_{ik}^{t'}(t, t^*)] &= \sum_{x,v} \sum_m \{ \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t, M_{ik}=m] \\ &\quad - \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t^*, M_{ik}=m] \} \\ &\quad \times P(M_{ik}=m | X_{ik}=x, V_k=v, T_k=t') P(X_{ik}=x, V_k=v). \end{aligned} \quad (18)$$

The quantity on the right hand side of expression (18), the difference in expectations, $\mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t, M_{ik}=m] - \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t^*, M_{ik}=m]$, could be estimated from a multilevel model such as that given in (14). In order to estimate $P(M_{ik}=m | X_{ik}=x, V_k=v, T_k=t')$ a separate regression model will be needed. Generalized linear multilevel models will again be needed if the outcome or mediator is not continuous. Once the quantity in (18) is obtained, standard errors could be calculated by bootstrapping techniques.

Pearl (2001), in the context of causal directed acyclic graphs (Pearl, 1995; Greenland et al., 1999), gives identifiability conditions for controlled directed effects and natural direct and indirect effects for non-clustered data which are somewhat weaker than (11), (12), (16) and (17); some alternative identification conditions are also given by other authors (Petersen et al., 2006; Imai et al., 2009; Hafeman and VanderWeele, 2009); these conditions could similarly be generalized to multi-level data. However, in general, natural direct and indirect effects will not be identified if there is a consequence of the treatment that confounds the relationship between the mediator and the outcome (Avin et al., 2005) unless there is no interaction between the effects of the treatment and the mediator on the outcome in which case controlled direct effects and natural direct effects coincide.

Recall that condition (17) essentially required that there be no effect of neighborhood level treatment, T_k , which itself confounds the relationship between the individual level mediator, M_{ik} , and the outcome, Y_{ik} . Thus if the building of the hospital led to employment and higher incomes for some individuals which caused them to move from the west side to the east side of the town where there was access to the hospital within 10 miles and less air pollution then condition (17) would be violated. As explained in the previous section, controlled direct effects would still be identified in this case; but natural direct effects in general would not be identified. If the variable indicating side of the town of individuals' residence was not affected by the construction of the hospital then (17) might then still hold; alternatively if the side of the town of an individual's residence did not affect health outcomes except through access to clinical care (e.g. if air pollution were not an issue) then (17) might again still hold. However, whenever a variable is a consequence of treatment and confounds the mediator-outcome relationship, natural direct effects are not in general identified.

Now consider an expression such as

$$\mathbb{E}[Y_{ik}(1, M_{ik}(1))] - \mathbb{E}[Y_{ik}(1, M_{ik}(0))]. \quad (19)$$

The expression given in (19) represents the average causal effect of neighborhood level treatment, T_k , of building a hospital mediated through individual level access to clinical care,

M_{ik} . More precisely, the expression given in (19) assumes an intervention to build a hospital in neighborhood k (i.e. setting $T_k = 1$) and compares what would happen to individual level health outcomes, Y_{ik} , if individual level access, M_{ik} , were set to what it would have been if T_k were 1 in contrast to what it would have been if T_k were 0. More generally define the natural indirect effect for individual i in neighborhood k by

$NIE_{ik}^{t'}(t, t^*) = Y_{ik}(t', M_{ik}(t)) - Y_{ik}(t', M_{ik}(t^*))$. Under assumptions (11), (12), (16) and (17), the same assumptions required for natural direct effects, natural indirect effects are also identified.

Theorem 3. If conditions (11), (12), (16) and (17) hold then the average natural indirect effect $\mathbb{E}[NIE_{ik}^{t'}(t, t^*)] = \mathbb{E}[Y_{ik}(t', M_{ik}(t)) - Y_{ik}(t', M_{ik}(t^*))]$ is identified and is given by:

$$\begin{aligned} \mathbb{E}[NIE_{ik}^{t'}(t, t^*)] &= \sum_{x,v} \sum_m \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t', M_{ik}=m] \\ &\quad \times P(M_{ik}=m | X_{ik}=x, V_k=v, T_k=t) P(X_{ik}=x, V_k=v) \\ &\quad - \sum_{x,v} \sum_m \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t', M_{ik}=m] \\ &\quad \times P(M_{ik}=m | X_{ik}=x, V_k=v, T_k=t^*) P(X_{ik}=x, V_k=v). \end{aligned} \quad (20)$$

As with expression (18) in Theorem 2, in order to estimate the quantity on the right hand side of expression (20), the expectation $\mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, T_k=t', M_{ik}=m]$ could be estimated from a multilevel model such as that given in (14) and a separate regression model will be needed to estimate probabilities of the form $P(M_{ik}=m | X_{ik}=x, V_k=v, T_k=t)$. Standard errors for the quantity in (20) could be calculated by bootstrapping techniques. As noted above, the approach can easily accommodate interactions between the effects of the treatment and the mediator on the outcome. For example, if multilevel models for Y_{ik} and M_{ik} are given by:

$$Y_{ik} = \beta_0 + \beta_1 X_{ik} + \beta_2 V_k + \gamma T_k + \theta M_{ik} + \kappa T_k M_{ik} + u_k + \varepsilon_{ik}$$

$$M_{ik} = \delta_0 + \delta_1 X_{ik} + \delta_2 V_k + \lambda T_k + \xi_k + \eta_{ik}$$

with $u_k, \varepsilon_{ik}, \xi_k, \eta_{ik}$ all mutually independent with mean zero, then it is relatively straightforward to show (VanderWeele and Vansteelandt, 2009) that the natural direct effect,

$\mathbb{E}[NDE_{ik}^{t*}(t, t^*)]$, and natural indirect effect, $\mathbb{E}[NIE_{ik}^t(t, t^*)]$, are given by:

$$\mathbb{E}[NDE_{ik}^{t*}(t, t^*)] = \mathbb{E}[Y_{ik}(t, M_{ik}(t^*)) - Y_{ik}(t^*, M_{ik}(t^*))] = \{\gamma + \kappa(\delta_0 + \delta_1 \mathbb{E}[X_{ik}] + \delta_2 \mathbb{E}[V_k] + \lambda t^*)\}(t - t^*)$$

$$\mathbb{E}[NIE_{ik}^t(t, t^*)] = \mathbb{E}[Y_{ik}(t, M_{ik}(t)) - Y_{ik}(t, M_{ik}(t^*))] = (\theta\lambda + \kappa\lambda t)(t - t^*).$$

Similarly, the natural direct effect, $\mathbb{E}[NDE_{ik}^{t*}(t, t^*)]$, and indirect effect, $\mathbb{E}[NIE_{ik}^t(t, t^*)]$, are given by:

$$\mathbb{E}[NDE_{ik}^t(t, t^*)] = \mathbb{E}[Y_{ik}(t, M_{ik}(t)) - Y_{ik}(t^*, M_{ik}(t))] = \{\gamma + \kappa(\delta_0 + \delta_1 \mathbb{E}[X_{ik}] + \delta_2 \mathbb{E}[V_k] + \lambda t)\}(t - t^*)$$

$$\mathbb{E}[NIE_{ik}^{t*}(t, t^*)] = \mathbb{E}[Y_{ik}(t^*, M_{ik}(t)) - Y_{ik}(t^*, M_{ik}(t^*))] = (\theta\lambda + \kappa\lambda t^*)(t - t^*).$$

Note that if there is no interaction between the effects of the treatment and the mediator on the outcome so that $\kappa = 0$, then the expressions for the natural direct effect reduce to $\gamma(t - t^*)$ and the expressions for the natural indirect effect reduce to $\theta\lambda(t - t^*)$.

It is easy to verify that the total average causal effect (of building a new hospital or increasing the level of neighborhood policing) can be decomposed into a natural direct effect and a natural indirect effect. For example:

$$\begin{aligned} Y_{ik}(1) - Y_{ik}(0) &= Y_{ik}(1, M_{ik}(1)) - Y_{ik}(0, M_{ik}(0)) \\ &= Y_{ik}(1, M_{ik}(1)) - Y_{ik}(1, M_{ik}(0)) + Y_{ik}(1, M_{ik}(0)) - Y_{ik}(0, M_{ik}(0)) \quad (21) \\ &= NIE_{ik}^1(1, 0) + NDE_{ik}^0(1, 0) \end{aligned}$$

and similarly,

$$\begin{aligned} Y_{ik}(1) - Y_{ik}(0) &= Y_{ik}(1, M_{ik}(1)) - Y_{ik}(0, M_{ik}(0)) \\ &= Y_{ik}(1, M_{ik}(1)) - Y_{ik}(0, M_{ik}(1)) + Y_{ik}(0, M_{ik}(1)) - Y_{ik}(0, M_{ik}(0)) \quad (22) \\ &= NDE_{ik}^1(1, 0) + NIE_{ik}^0(1, 0). \end{aligned}$$

The expression in (21), for example, decomposes the total effect of building a hospital on individual health outcomes into (i) a natural indirect effect,

$NIE_{ik}^1(1, 0) = Y_{ik}(1, M_{ik}(1)) - Y_{ik}(1, M_{ik}(0))$, the portion of the effect of building a hospital mediated through the increase in access to clinical care that the new hospital provides and (ii) a natural direct effect, $NDE_{ik}^0(1, 0) = Y_{ik}(1, M_{ik}(0)) - Y_{ik}(0, M_{ik}(0))$, the portion of the effect of building a hospital on individual health outcomes if individual access to clinical care within 10 miles had been set to what it would have been had no hospital been built. In this setting the natural direct effect represents the portion of the effect of building a hospital on individual health outcomes mediated by pathways other than improved access to clinical care e.g. by changing the neighborhood itself. More generally,

$$Y_{ik}(t) - Y_{ik}(t^*) = NIE_{ik}^t(t, t^*) + NDE_{ik}^{t*}(t, t^*) \text{ and also}$$

$$Y_{ik}(t) - Y_{ik}(t^*) = NDE_{ik}^t(t, t^*) + NIE_{ik}^{t*}(t, t^*). \text{ Note that expressions (21) and (22), unlike say expression (10), is a decomposition of the total effect of building a new hospital } Y_{ik}(1) - Y_{ik}(0), \text{ rather than of the average joint effect of building a hospital and ensuring access } Y_{ik}(1, 1) - Y_{ik}(0, 0).$$

The decomposition of a total effect into a natural direct and indirect effects may also be of interest in the neighborhood policing example. Let t denote a high level of neighborhood policing and t^* a low level of neighborhood policing. The average total effect $\mathbb{E}[Y_{ik}(t) - Y_{ik}(t^*)]$ can be decomposed into the sum of an average natural direct effect, $\mathbb{E}[Y_{ik}(t, M_{ik}(t)) - Y_{ik}(t^*, M_{ik}(t))]$, and an average natural indirect effect, $\mathbb{E}[Y_{ik}(t^*, M_{ik}(t)) - Y_{ik}(t^*, M_{ik}(t^*))]$. If

the neighborhood policing is effective we might expect the average natural direct effect to be negative: the increased neighborhood policing directly decreases the incidence of crime. However, it is possible that for the average natural indirect effect, the increased neighborhood policing may lead to an increase in the mediator, the average number of hours individuals spend per week walking through their neighborhoods, and may consequently lead to potentially increased exposure to criminal activity. Estimation of the average natural indirect effect would allow a researcher to assess whether the data supports a hypothesis like this. If the indirect effect were positive this would indicate that increased neighborhood policing indirectly increases crime by increasing the time individuals spend walking through the neighborhood. We might still expect the total effect to be negative with direct effect of neighborhood policing on crime dominating the indirect effect. In such a case the direct effect would be of greater magnitude than the total effect; the total would not be as substantial because of the, possibly inadvertent, indirect effect of increasing hours spent walking.

5. Relaxing No-Interference Assumptions

The no-interference assumptions which were required so that counterfactuals of the form $Y_{ik}(t, m)$ and $M_{ik}(t)$ were well defined, and thus so that controlled and natural direct and indirect effect could be defined, were quite strong. In order to define controlled direct effects we needed counterfactuals of the form $Y_{ik}(t, m)$ to be well-defined and so we required that the outcome for individual i in cluster k did not depend on the value of the mediator for any other individual i' in the same or any other cluster nor on the cluster-level treatment of any other cluster k' . In addition, in order to define natural direct and indirect effects, we also needed counterfactuals of the form $M_{ik}(t)$ to be well defined and so we required that the individual level mediator for individual i in cluster k did not depend on the value of the cluster-level treatment of any other cluster k' . We thus required a no-interference assumption at the individual level for the mediator and no-interference assumptions at the cluster level for the treatment. The no-interference assumption at the individual level required for the mediator is a particularly strong assumption in the context of the study of neighborhoods.

In many settings these assumptions are unlikely to hold. Consider, for instance, the example of trying to estimate the controlled direct effects of constructing a new hospital in a particular neighborhood on individual health outcomes intervening on access. The no interference conditions in this context would require that the health outcome for individual i in neighborhood k does not depend on whether any other individual i' in the same or any other neighborhood has access to clinical care within 10 miles, nor on whether a hospital is built in any other neighborhood k' . The individual-level no-interference assumption for the mediator may well fail to hold here. In the case of a potential flu epidemic, for example, access to clinical care within 10 miles for individual i' in neighborhood k may allow for a .u vaccine to be administered to individual i' which in turn may prevent individual i in neighborhood k from contracting the flu. Thus the health status of individual i in this case would depend on whether individual i' in the same neighborhood has access to clinical care. When such violations of the no-interference assumptions occur, counterfactuals of the form $Y_{ik}(t, m)$ are not well defined. In the neighborhood policing example, it may be the case that an increase in the mediator, M_{ik} , hours per week an individual spends walking through the

neighborhood, may for a particular individual lead to increased exposure to crime. However, it might also be the case that, as more and more individuals spend time walking through the neighborhood, criminal activity in fact decreases because of increased public monitoring within that neighborhood. In such a setting we would once again have individual-level interference within a cluster for the mediator (time spent walking) because individual i in neighborhood k may not become a victim of crime because some other individual, i' , in the same neighborhood, by walking, is present at the scene at which the crime would have otherwise occurred.

These examples of interference create difficulties for causal inference and for reasoning about direct and indirect effects. Fortunately, there has been recent work on relaxing no-interference assumptions in the setting of clustered treatments (Hong and Raudenbush, 2006; Hudgens and Halloran, 2008; Verbitsky and Raudenbush, 2009). We will briefly indicate how this work can be applied in the context of estimating direct and indirect effects of cluster-level interventions. Note also other authors have considered the relaxation of no-interference assumptions in the context of general randomized experiments (Rosenbaum, 2007) and of randomizing individuals to a change in neighborhood (Sobel, 2006).

In the context of neighborhood effects it is noted in related work (Sobel, 2006; Hudgens and Halloran, 2008; VanderWeele, 2008a) that if neighborhoods or clusters included in a study are sufficiently geographically separated then the assumption that the outcome for individual i in cluster k does not depend on the neighborhood-level intervention of any other neighborhood k' is likely to hold at least approximately. This is sometimes referred to as an assumption of partial interference (Sobel, 2006; Hudgens and Halloran, 2008). We will therefore focus our attention here on violations of the no-interference assumption that require that the outcome for individual i in cluster k not depend on the the value of the mediator for any other individual i' in the same or any other cluster i.e. on violations of the individual-level no-interference assumption for the mediator. In particular, we will assume that the outcome for individual i in cluster k does depend on the the values of the mediator for other individuals in the same cluster but not in other clusters. We assume that the outcome and mediator for individual i in cluster k does not depend on the intervention in any other cluster k' . The remarks made here, however, can be adapted to settings in which other types of no-interference assumptions are violated. For example, Verbitsky and Raudenbush (2009) consider violations of no-interference assumptions which involve spillover effects from the intervention for one neighborhood to another. Indeed, in some studies the spillover effects may be amongst the primary effects of interest (Sobel, 2006; Hudgen and Halloran, 2008; Verbitsky and Raudenbush, 2009). Here we have essentially avoided between-cluster interference for the treatment and mediator by assuming neighborhoods are geographically separated; within-cluster interference for the treatment is not an issue because treatment is administered at the cluster level; and thus our focus is on individual-level within-cluster interference for the mediator.

Let M_{-ik} denote the values of the mediator for all individuals $i' \neq i$ in cluster k . Let $Y_{ik}(t, m_i, f(m_{-i}))$ be the counterfactual value of the outcome for individual i in cluster k if T_k were set to t , if M_{ik} were set to m and if M_{-ik} were set to m_{-i} where f is some function of m_{-i} . In the context of the hospital construction example, $f(m_{-i})$ might indicate the proportion of

individuals $i' = i$ in neighborhood k which have access to clinical care within 10 miles. In the context of neighborhood policing $f(m_{-i})$ might indicate the average number of hours per week spent walking in neighborhood k for all individuals $i' = i$ in neighborhood k . In this neighborhood policing example we might then expect $Y_{ik}(t, m_i, f(m_{-i}))$ to increase with m_i (because of greater exposure to potential criminal activity) and to decrease with $f(m_{-i})$ (because of greater public monitoring). Using these modified counterfactuals accounting for interference, controlled direct effects and natural direct and indirect effects can still be defined. Let $\mathbf{M}_k$ denote $\{M_{ik}\}_{i=1}^{n_k}$ where n_k denotes the number of individual in cluster k . Let $M_{-ik}(t)$ denote the counterfactual values of the mediator for all individuals $i' = i$ in cluster k if T_k were set to t . We define the controlled direct effect for individual i in neighborhood k , intervening to set $\mathbf{M}_k$ to $\mathbf{m}$ by $CDE_{ik}^m(t, t') = Y_{ik}(t, m_i, f(m_{-i})) - Y_{ik}(t', m_i, f(m_{-i}))$. We define the natural direct effect for individual i in neighborhood k by

$NDE_{ik}^{t'}(t, t*) = Y_{ik}(t, M_{ik}(t'), f(M_{-ik}(t'))) - Y_{ik}(t*, M_{ik}(t'), f(M_{-ik}(t')))$. We define the natural indirect effect for individual i in neighborhood k by

$$NIE_{ik}^{t'}(t, t*) = Y_{ik}(t', M_{ik}(t), f(M_{-ik}(t))) - Y_{ik}(t', M_{ik}(t*), f(M_{-ik}(t*))).$$

In order to identify direct and indirect effects under such interference, modified ignorability conditions are also needed. Let X_{-ik} denote the values of the covariates for all individuals $i' = i$ in cluster k and let h be some function of X_{-ik} . It can be verified, by a modification of the argument given in the proof of Theorem 1, that if

$$Y_{ik}(t, m_i, f(m_{-i})) \prod T_k | X_{ik}, V_k, h(X_{-ik}) \quad (23)$$

and

$$Y_{ik}(t, m_i, f(m_{-i})) \prod \{M_{ik}, f(M_{-ik})\} | T_k, X_{ik}, V_k, h(X_{-ik}) \quad (24)$$

then controlled direct effects are identified and given by $\mathbb{E}[CDE_{ik}^m(t, t')]$

$$\begin{aligned} &= \sum_{x,v,h} \{ \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, h(X_{-ik})=h, T_k=t, M_{ik}=m_i, f(M_{-ik})=f(m_{-i})] \\ &- \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, h(X_{-ik})=h, T_k=t', M_{ik}=m_i, f(M_{-ik})=f(m_{-i})] \} P(X_{ik}=x, V_k=v, h(X_{-ik})=h). \end{aligned} \quad (25)$$

The expression on the right hand side of (25) is given in terms of the observed data and the conditional expectations could be estimated from a multilevel model which extends (14), so as to include terms for $f(M_{-ik})$ and $h(X_{-ik})$, such as:

$$Y_{ik} = \beta_0 + \beta_1 X_{ik} + \beta_2 V_k + \gamma T_k + \theta M_{ik} + \zeta f(M_{-ik}) + \varrho h(X_{-ik}) + u_k + \varepsilon_{ik} \quad (26)$$

where once again u_k and ε_{ik} are assumed to be independent and identically distributed with mean zero. Interaction terms could be included in the multilevel model given in (26) if appropriate. Condition (23) is similar to condition (11) above but requires that it holds for all m_i and $f(m_{-i})$ rather than just for m_i . Condition (24) is similar to condition (12) but it requires that within strata of $T_k, X_{ik}, V_k, h(X_{-ik})$, the effects of M_{ik} and $f(M_{-ik})$ on Y_{ik} are unconfounded rather than merely the effect of M_{ik} on Y_{ik} . If in addition to (23) and (24), if we also have the following two conditions,

$$\{M_{ik}(t), f(M_{-ik}(t))\} \prod T_k | X_{ik}, V_k, h(X_{-ik}) \quad (27)$$

and

$$Y_{ik}(t, m, f(m_{-i})) \prod \{M_{ik}(t'), f(M_{-ik}(t'))\} | X_{ik}, V_k, h(X_{-ik}) \quad (28)$$

then it can be shown by a modification of the argument given in the proof of Theorem 2 that expressions such as $\mathbb{E}Y_{ik}(t, M_{ik}(t'), f(M_{-ik}(t')))$ are identified and given by $\mathbb{E}Y_{ik}(t, M_{ik}(t'), f(M_{-ik}(t')))$

$$= \sum_{x,v,h} \sum_m \mathbb{E}[Y_{ik} | X_{ik}=x, V_k=v, h(X_{-ik})=h, T_k=t', M_{ik}=m_i, f(M_{-ik})=f(m_{-i})] \times P(M_{ik}=m_i, f(M_{-ik})=f(m_{-i}) | X_{ik}=x, V_k=v, h(X_{-ik})=h, T_k=t') P(X_{ik}=x, V_k=v, h(X_{-ik})=h). \quad (29)$$

The expression on the right hand side of (29) is given in terms of the observed data. The conditional expectation can once again be estimated using a multilevel model such as that given in (26). A separate bivariate multilevel regression model will be needed to estimate the probabilities $P(M_{ik} = m_i, f(M_{-ik}) = f(m_{-i}) | X_{ik} = x, V_k = v, h(X_{-ik}) = h, T_k = t')$. Note that (29) can be used to estimate average natural direct effects since

$$\mathbb{E}[NDE_{ik}^{t'}(t, t^*)] = \mathbb{E}[Y_{ik}(t, M_{ik}(t'), f(M_{-ik}(t')))] - \mathbb{E}[Y_{ik}(t^*, M_{ik}(t'), f(M_{-ik}(t')))]$$

Equation (29) can also be used to estimate average natural indirect effects since

$$\mathbb{E}[NIE_{ik}^{t'}(t, t^*)] = \mathbb{E}[Y_{ik}(t', M_{ik}(t), f(M_{-ik}(t)))] - \mathbb{E}[Y_{ik}(t', M_{ik}(t^*), f(M_{-ik}(t^*)))]$$

Condition (27) is similar to condition (16) except that it requires that it hold jointly for $\{M_{ik}(t), f(M_{-ik}(t))\}$ rather than for $M_{ik}(t)$ only. Condition (28) is likewise similar to (17) except that it again requires that the condition hold jointly for $\{M_{ik}(t'), f(M_{-ik}(t'))\}$ rather than for $M_{ik}(t')$ only.

6. Extensions to Clustered Longitudinal Data

Our discussion up until this point has assumed that the cluster-level intervention takes place at a single point in time. In the context of non-clustered interventions, van der Laan and Petersen (2004, 2008) provide discussion of how the definitions of direct and indirect effects can be extended to a longitudinal setting in which the treatment, mediator and confounding variables may change over time. Here we adapt their discussion for the case of direct and indirect effects for clustered longitudinal data.

We index time by $s = 0, 1, \dots, S$. Let Y_{ik} denote the individual level outcome for individual i in cluster or neighborhood k at the end of the study. Let X_{iks} denote individual level characteristics for individual i in cluster k at time s . Let V_{ks} denote cluster-level characteristics for cluster k at time s . Let T_{ks} denote the cluster-level treatment or intervention at time s . Let M_{iks} be some individual level mediator for individual i in cluster k at time s . Let $\bar{X}_{iks}$ denote the vector $(X_{ik0}, X_{ik1}, \dots, X_{iks})$ and similarly let $\bar{V}_{ks}, \bar{T}_{ks}$ and $\bar{M}_{iks}$ denote V_{ks}, T_{ks} and M_{iks} respectively from time periods 0 through s . We will use $\bar{X}_{ik}, \bar{V}_k, \bar{T}_k$ and $\bar{M}_{ik}$ to denote $\bar{X}_{iks}, \bar{V}_{ks}, \bar{T}_{ks}$ and $\bar{M}_{iks}$. Let $Y_{ik}(\bar{t}, \bar{m})$ denote the counterfactual outcome for individual i in cluster k if $\bar{T}_k$ were set to $\bar{t}$ and if

$\bar{M}_{ik}$ were set to $\bar{m}$. Let $M_{iks}(\bar{t}_s)$ denote the counterfactual value of the individual level intermediate, M_{iks} , if $\bar{T}_{ks}$ were set to $\bar{t}_s$. We will assume that the individual and neighborhood level SUTVA conditions apply for every time-point so that $Y_{ik}(\bar{t}, \bar{m})$ and $M_{iks}(\bar{t}_s)$ are well-defined. Let $\bar{M}_{iks}(\bar{t}_s)$ denote $M_{iks}(\bar{t}_s)$ from time periods 0 through s and let $\bar{M}_{ik}(\bar{t})$ denote $\bar{M}_{iks}(\bar{t}_s)$.

This longitudinal setting may be of interest, for example, in the context of the neighborhood policing setting. It may take time, a number of months or years, of increased neighborhood policing before substantial changes in walking behavior or in the incidence of crime takes place; it may be the cumulative effects of neighborhood policing that eventually lead to changes in walking behavior or incidence of crime. Consideration of this clustered and longitudinal setting would thus allow for assessing how the effects of neighborhood policing on crime outcomes change or accumulate over time.

For clustered longitudinal data we define the controlled direct effect by

$CDE_{ik}^{\bar{m}}(\bar{t}, \bar{t}') = Y_{ik}(\bar{t}, \bar{m}) - Y_{ik}(\bar{t}', \bar{m})$. To estimate average controlled direct effects we can use the following theorem.

Theorem 4. If

$$Y_{ik}(\bar{t}, \bar{m}) \prod T_{ks} | \bar{X}_{iks}, \bar{V}_{ks}, \bar{T}_{k(s-1)}, \bar{M}_{ik(s-1)} \text{ for all } s \quad (30)$$

and

$$Y_{ik}(\bar{t}, \bar{m}) \prod M_{iks} | \bar{X}_{iks}, \bar{V}_{ks}, \bar{T}_{ks}, \bar{M}_{ik(s-1)} \text{ for all } s \quad (31)$$

then $\mathbb{E}[Y_{ik}(\bar{t}, \bar{m})]$

$$= \sum_{\bar{x}, \bar{v}} \mathbb{E}[Y_{ik} | \bar{X}_{ik} = \bar{x}, \bar{V}_k = \bar{v}, \bar{T}_k = \bar{t}, \bar{M}_{ik} = \bar{m}] \times \prod_{s=0}^S P(X_{iks} = x, V_{ks} = v | \bar{X}_{ik(s-1)} = \bar{x}_{s-1}, \bar{V}_{k(s-1)} = \bar{v}_{s-1}, \bar{T}_{k(s-1)} = \bar{t}_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}). \quad (32)$$

The right hand side of (32) is given in terms of observed data. Average controlled direct effects for clustered longitudinal data can be estimated from (32) since

$\mathbb{E}[CDE_{ik}^{\bar{m}}(\bar{t}, \bar{t}')] = \mathbb{E}[Y_{ik}(\bar{t}, \bar{m})] - \mathbb{E}[Y_{ik}(\bar{t}', \bar{m})]$. In practice, in order to estimate controlled direct effects from longitudinal data, one could employ multi-level marginal structural models recently developed by Hong and Raudenbush (2008). These models are a generalization for clustered data of the individual level marginal structural models developed by Robins (1999, cf. Robins et al. 2000); these individual level marginal structural models have also been employed in the estimation of direct and indirect effects for non-clustered data (van der Laan and Petersen, 2004; VanderWeele, 2009c). The implementation of multilevel marginal structural models is beyond the scope of this paper; see the work of Hong and Raudenbush (2008) for details.

Condition (30) requires that for each time-period s the effect of treatment, T_{ks} , on outcome, Y_{ik} , is unconfounded given the covariate history, $\bar{X}_{iks}$, $\bar{V}_{ks}$, up until time s , the treatment history, $\bar{T}_{k(s-1)}$, up until time $s - 1$, and the mediator history, $\bar{M}_{ik(s-1)}$, up until time $s - 1$. Condition (31) requires that for every time-period s the effect of the mediator, M_{iks} , on outcome, Y_{ik} , is unconfounded given the covariate history, $\bar{X}_{iks}$, $\bar{V}_{ks}$, up until time s , the treatment history, $\bar{T}_{ks}$, up until time s and the mediator history, $\bar{M}_{ik(s-1)}$, up until time $s - 1$.

We can also define natural direct and indirect effects for clustered longitudinal data. The natural direct effect can be defined as

$NDE_{ik}^{\bar{t}'}(\bar{t}, \bar{t}^*) = Y_{ik}(\bar{t}, \bar{M}_{ik}(\bar{t}')) - Y_{ik}(\bar{t}^*, \bar{M}_{ik}(\bar{t}'))$ for any $\bar{t}, \bar{t}^*$ and $\bar{t}'$. The natural indirect effect can be defined as $NIE_{ik}^{\bar{t}'}(\bar{t}, \bar{t}^*) = Y_{ik}(\bar{t}', \bar{M}_{ik}(\bar{t})) - Y_{ik}(\bar{t}', \bar{M}_{ik}(\bar{t}^*))$ for any $\bar{t}, \bar{t}^*$ and $\bar{t}'$.

We again can decompose the total effect of treatment, $\bar{T}_k$, into a natural direct effect and a natural indirect effect:

$Y_{ik}(\bar{t}, \bar{M}_{ik}(\bar{t})) - Y_{ik}(\bar{t}^*, \bar{M}_{ik}(\bar{t}^*)) = NIE_{ik}^{\bar{t}}(\bar{t}, \bar{t}^*) + NDE_{ik}^{\bar{t}^*}(\bar{t}, \bar{t}^*) = NDE_{ik}^{\bar{t}}(\bar{t}, \bar{t}^*) + NIE_{ik}^{\bar{t}^*}(\bar{t}, \bar{t}^*)$. We can use the following theorem to estimate average natural direct and indirect effects.

Theorem 5. If (30) and (31) hold and if in addition we also have that

$$\bar{M}_{iks}(\bar{t}_s) \prod T_{ks} | \bar{X}_{iks}, \bar{V}_{ks}, \bar{T}_{k(s-1)}, \bar{M}_{ik(s-1)} \text{ for all } s \quad (33)$$

and

$$Y_{ik}(\bar{t}, \bar{m}) \prod \bar{M}_{ik}(\bar{t}') | X_{ik0}, V_{k0} \quad (34)$$

then $\mathbb{E}[Y_{ik}(\bar{t}, \bar{M}_{ik}(\bar{t}^*))]$

$$\begin{aligned} &= \sum_{x_0, v_0} \sum_m \left\{ \sum_{(x_1, \dots, x_S), (v_1, \dots, v_S)} \mathbb{E}[Y_{ik} | \bar{X}_{ik} = \bar{x}, \bar{V}_k = \bar{v}, \bar{T}_k = \bar{t}, \bar{M}_{ik} = \bar{m}] \right. \\ &\quad \times \prod_{s=0}^S P(X_{iks} = x, V_{ks} = v | \bar{X}_{ik(s-1)} = \bar{x}_{s-1}, \bar{V}_{k(s-1)} = \bar{v}_{s-1}, \bar{T}_{k(s-1)} = \bar{t}_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}) \} \\ &\quad \left\{ \sum_{(x_1, \dots, x_S), (v_1, \dots, v_S)} \left\{ \prod_{s=0}^S P(M_{iks} = m_s | \bar{X}_{iks} = \bar{x}_s, \bar{V}_{ks} = \bar{v}_s, \bar{T}_{ks} = \bar{t}_s^*, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}) \right\} \right. \\ &\quad \left. \left\{ \prod_{s=0}^S P(X_{iks} = x_s, V_{ks} = v_s | \bar{X}_{ik(s-1)} = \bar{x}_{s-1}, \bar{V}_{k(s-1)} = \bar{v}_{s-1}, \bar{T}_{k(s-1)} = \bar{t}_{s-1}^*, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}) \right\} \right\}. \end{aligned} \quad (35)$$

The right hand side of (35) is given in terms of observed data. Average natural direct effects for clustered longitudinal data can be estimated from (35) since

$\mathbb{E}[NDE_{ik}^{\bar{t}'}(\bar{t}, \bar{t}^*)] = \mathbb{E}[Y_{ik}(\bar{t}, \bar{M}_{ik}(\bar{t}'))] - \mathbb{E}[Y_{ik}(\bar{t}^*, \bar{M}_{ik}(\bar{t}'))]$. Average natural indirect effects

can be estimated from (35) since $\mathbb{E}[NIE_{ik}^{\bar{t}'}(\bar{t}, \bar{t}^*)] = \mathbb{E}[Y_{ik}(\bar{t}', \bar{M}_{ik}(\bar{t}))] - \mathbb{E}[Y_{ik}(\bar{t}', \bar{M}_{ik}(\bar{t}^*))]$.

Condition (33) requires that for each time-period s the effect of treatment, T_{ks} , on the mediator, M_{iks} , is unconfounded given the covariate history, $\bar{X}_{iks}$, $\bar{V}_{ks}$, up until time s , the treatment history, $\bar{T}_{k(s-1)}$, up until time $s - 1$ and the mediator history, $\bar{M}_{ik(s-1)}$, up until time $s - 1$. Condition (34) is a longitudinal generalization of condition (17) above. It will be violated if for any s there is an effect of treatment, T_{ks} , which itself confounds the

relationship between the mediator, M_{iks} , and the outcome, Y_{ik} . Such an assumption may be reasonable when treatment, T_{ks} , is randomized at each time period s and when the mediator, M_{iks} , is measured shortly after treatment. However, assumption (34) like assumption (17) is a strong assumption and will often be violated.

7. Discussion

We have given conditions under which controlled direct effects and natural direct and indirect effects can be estimated from observational data for a cluster-level intervention both for interventions occurring at a single point in time and for time-varying interventions. The conditions under which such estimation can take place are relatively strong. For the estimation of controlled direct effects, we required that the effect of treatment on the outcome was unconfounded given the covariates and that the effect of the mediator on the outcome was unconfounded given the covariates. For the estimation of natural direct and indirect effects we also required that the effect of treatment on the mediator was unconfounded given the covariates and that there was no consequence of treatment that confounded the mediator-outcome relationship.

In addition to these ignorability (or no unmeasured confounding) conditions we also required certain stability or no-interference conditions. In particular, in order to define controlled direct effects we initially required that the outcome for individual i in cluster k did not depend on the value of the mediator for any other individual i' in the same or any other cluster nor on the cluster-level intervention of any other neighborhood k' . In addition, in order to define natural direct and indirect effects, we also initially required that the individual level mediator for individual i in cluster k did not depend on the value of the cluster-level intervention of any other neighborhood k' . Essentially we made an individual-level no-interference assumption for the mediator as well as neighborhood-level no-interference assumptions for the treatment. The individual-level no-interference assumption for the mediator is particularly stringent and in many contexts will be unlikely to hold. Fortunately, as discussed in Section 5, these assumptions can be relaxed and we have provided definitions and identification conditions for direct and indirect effects under individual-level interference. Nevertheless, the assumptions required for the identification of the direct and indirect effects of a cluster-level intervention are considerable. Whether these assumptions hold approximately in any given study will have to be judged by subject matter experts and determined on a case-by-case basis.

This work has generalized the counterfactual approach to direct and indirect effects (Robins and Greenland, 1992; Pearl, 2001; Robins, 2003) to the multilevel setting. The advantage of this approach over the approach typically advocated in the structural equation modeling literature (Bollen, 1987; Holland, 1988; Raudenbush and Sampson, 1999) is that it allows for the definition, estimation and effect decomposition of direct and indirect effects even in settings in which there are interactions and nonlinearities; as discussed above, the approach does not assume that the direct effects, for example, are homogenous for all levels of the mediator. A further advantage of considering direct and indirect effects in a counterfactual setting is that the no unmeasured confounding assumptions required to interpret estimates causally are made clear. As has been noted above, the identification assumptions required

for natural direct and indirect effects are even stronger than those required for controlled direct effects. The counterfactual approach also makes clear the no-interference assumptions required in defining total, direct and indirect effects and provides a framework within which such no-inference assumptions can be relaxed. In the multilevel setting, the definition and identification is complicated by issues concerning correlated data and by potential interference; but, these complications aside, the counterfactual approach to direct and indirect effects in the individual level setting otherwise carries over in a relatively straightforward way to the multilevel setting.

In this paper we have focused on cases in which the treatment or intervention of interest was at the neighborhood or cluster level. A similar approach to direct and indirect effects could be applied with clustered data in which the treatment was administered at the individual level. Some of the issues discussed in this paper carry over in a straightforward way even when individual level treatment is being considered. The neighborhood treatment variable, T_k , would be replaced by individual level treatment, T_{ik} . The identification assumptions described in this paper would still apply with T_k replaced by T_{ik} throughout; though one would have to consider whether these assumptions were satisfied in any particular study. The multilevel model in (14) could be replaced by

$$Y_{ik} = \beta_0 + \beta_1 X_{ik} + \beta_2 V_k + \gamma T_{ik} + \theta M_{ik} + u_k + \varepsilon_{ik}$$

again with interactions or generalized linear multilevel models being used as appropriate. However, a significant additional complication that arises when treatment is administered at the individual level is that within-cluster interference could then arise both for the treatment variable and for the mediator variable. An approach similar to that described in section 5 for relaxing no-interference assumptions for the mediator would also have to be employed for the treatment variable and such approaches to handle within-cluster inference for treatment variables have been discussed in the literature (Hong and Raudenbush, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008). Future work could develop a fully general approach which allowed for both within-cluster interference and between-cluster interference, at both the treatment stage and the mediator stage, and which considered the identification of not just total, direct and indirect effects but also within-cluster spillover effects and between-cluster spillover effects.

Acknowledgments

The author thanks Steve Raudenbush and Guanglei Hong for several helpful comments on an earlier draft of this paper.

References

- Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. *Proceedings of the International Joint Conferences on Artificial Intelligence*. 2005:357–363.
- Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*. 1986; 51:1173–1182. [PubMed: 3806354]
- Bollen, KA. Total, direct and indirect effects in structural equation models. In: Clogg, CC., editor. *Sociological Methodology*. Washington, DC: American Sociological Association; 1987. p. 37-69.

- Browning CR, Wallace D, Feinberg S, Cagney KA. Neighborhood social processes and disaster-related mortality: The case of the 1995 Chicago heat wave. *American Sociological Review*. 2006; 71:665–682.
- Cole SR, Frangakis CE. The consistency assumption in causal inference: a definition or an assumption? *Epidemiology*. 2009; 20:3–5. [PubMed: 19234395]
- Cole SR, Hernán MA. Fallibility in estimating direct effects. *International Journal of Epidemiology*. 2002; 31:163–165. [PubMed: 11914314]
- Cox, DR. *The Planning of Experiments*. New York: Wiley; 1958.
- Del Conte A, Kling J. A synthesis of MTO research on self-sufficiency, safety and health, and behavior and delinquency. *Poverty Research News*. 2001; 5:3–6.
- Diez-Roux AV, Nieto FJ, Muntaner C. Neighborhood environments and coronary heart disease: A multilevel analysis. *American Journal of Epidemiology*. 1997; 146:48–63. [PubMed: 9215223]
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58:21–29. [PubMed: 11890317]
- Gallop R, Small DS, Lin JY, Elliott MR, Joffe M, Ten Have TR. Mediation analysis with principal stratification. *Statistics in Medicine*. 2009; 28:1108–1130. [PubMed: 19184975]
- Gitelman AI. Estimating causal effects from multilevel group-allocation data. *Journal of Educational and Behavioral Statistics*. 2005; 30:397–412.
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999; 10:37–48. [PubMed: 9888278]
- Hafeman DM, Schwartz S. Opening the black box: a motivation for the assessment of mediation. *International Journal of Epidemiology*. 2009; 38:838–845. [PubMed: 19261660]
- Hafeman DM, VanderWeele TJ. Alternative assumptions for the identification of direct and indirect effects. *Epidemiology*. 2009 in press.
- Holland, PW. Causal inference, path analysis, and recursive structural equations models. In: Clogg, CC., editor. *Sociological Methodology*. Washington, DC: American Sociological Association; 1988. p. 449–484.
- Hong G, Raudenbush SW. Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*. 2006; 101:901–910.
- Hong G, Raudenbush SW. Causal inference for time-varying instructional treatments. *Journal of Educational and Behavioral Statistics*. 2008; 33:333–362.
- Hubbard AE, Ahern J, Fleischer NL, van der Laan M, Lippman SA, Bruckner T, Satariano WA. To GEE or not to GEE: comparing estimating function and likelihood-based methods for estimating the associations between neighborhoods and health. *Epidemiology*. in press.
- Hudgens MG, Halloran ME. Towards causal inference with interference. *Journal of the American Statistical Association*. 2008; 103:832–842. [PubMed: 19081744]
- Imai K, Keele L, Yamamoto T. Identification, inference, and sensitivity analysis for causal mediation effects. Working paper.
- Joffe M, Small D, Hsu C-Y. Defining and estimating intervention effects for groups that will develop an auxiliary outcome. *Statistical Science*. 2007; 22:74–97.
- Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluations. *Evaluation Review*. 1981; 5:602–619.
- Kaufman JS. Re: "Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression". *American Journal of Epidemiology*. 2005; 162:602–603. [PubMed: 16093285]
- Kaufman JS, MacLehose RF, Kaufman S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiologic Perspectives and Innovations*. 2004; 1:4. [PubMed: 15507130]
- Kling JR, Liebman JB, Katz LF. Experimental analysis of neighborhood effects. *Econometrica*. 2007; 75:83–119.
- Lee Y, Nelder JA. Hierarchical generalized linear models. *Journal of the Royal Statistical Society, Series B*. 1996; 58:619–678.

- Pearl J. Casual diagrams for empirical research (with discussion). *Biometrika*. 1995; 82:669–710.
- Pearl, J. Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence. San Francisco: Morgan Kaufmann; 2001. Direct and indirect effects; p. 411-420.
- Peterson ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. *Epidemiology*. 2006; 17:276–284. [PubMed: 16617276]
- Raudenbush SW, Sampson R. Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods and Research*. 1999; 28:123–153.
- Robins JM. A new approach to causal inference in mortality studies with sustained exposure period - application to control of the healthy worker survivor effect. *Mathematical Modelling*. 1986; 7:1393–1512.
- Robins JM. Addendum to a new approach to causal inference in mortality studies with sustained exposure period - application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications*. 1987; 14:923–945.
- Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992; 3:143–155. [PubMed: 1576220]
- Robins, JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran, ME.; Berry, D., editors. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. NY: Springer-Verlag; 1999. p. 95-134.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
- Robins, JM. Semantics of causal DAG models and the identification of direct and indirect effects. In: Green, P.; Hjort, NL.; Richardson, S., editors. *Highly structured stochastic systems*. New York: Oxford University Press; 2003. p. 70-81.
- Rosenbaum PR. Interference between units in randomized experiments. *Journal of the American Statistical Association*. 2007; 102:191–200.
- Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal Educational Psychology*. 1974; 66:688–701.
- Rubin DB. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*. 1978; 6:34–58.
- Rubin DB. Direct and indirect effects via potential outcomes. *Scandinavian Journal of Statistics*. 2004; 31:161–170.
- Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science*. 1997; 227:918–923. [PubMed: 9252316]
- Sobel ME. Effect analysis and causation in linear structural equation models. *Psychometrika*. 1990; 55:495–515.
- Sobel ME. What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *Journal of the American Statistical Association*. 2006; 101:1398–1407.
- Subramanian SV, Kim DJ, Kawachi I. Social trust and self-rated health in US communities: a multilevel analysis. *Journal of Urban Health*. 2002; 79:S21–S34. [PubMed: 12473696]
- van der Laan, MJ.; Petersen, ML. Estimation of direct and indirect causal effects in longitudinal studies. U.C. Berkeley Division of Biostatistics Working Paper Series. 2004. Working Paper 155; <http://www.bepress.com/ucbbiostat/paper155>
- van der Laan MJ, Petersen ML. Direct effect models. *International Journal of Biostatistics*. 2008; 4 Article 23.
- VanderWeele TJ. Ignorability and stability assumptions in neighborhood effects research. *Statistics in Medicine*. 2008a; 27:1934–1943. [PubMed: 18050151]
- VanderWeele TJ. Simple relations between principal stratification and direct and indirect effects. *Statistics and Probability Letters*. 2008b; 78:2957–2962.
- VanderWeele TJ. Mediation and mechanism. *European Journal of Epidemiology*. 2009a; 24:217–224. [PubMed: 19330454]
- VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009b; 20:880–883. [PubMed: 19829187]

- VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*. 2009c; 20:18–26. [PubMed: 19234398]
- VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface - Special Issue on Mental Health and Social Behavioral Science*. 2009; 2:457–468.
- Verbitsky N, Raudenbush SW. Causal inference in spatial settings: A case study of community policing program in Chicago. Working paper. 2009
- Winship C, Morgan SL. The estimation of causal effects from observational data. *Annual Review of Sociology*. 1999; 25:659–707.

Appendix: Proofs

Proof of Theorem 1

$$\begin{aligned}
 & \text{We have that } \mathbb{E}Y_{ik}(t, m) - \mathbb{E}Y_{ik}(t', m) = \sum_{x,v} \{ \mathbb{E}Y_{ik}(t, m) | X_{ik} = x, V_k = v \} - \sum_{x,v} \{ \mathbb{E}Y_{ik}(t', m) | X_{ik} = x, V_k = v \} \} P(X_{ik} = x, V_k = v) \text{ by iterated expectations} \\
 & = \sum_{x,v} \{ \mathbb{E}Y_{ik}(t, m) | X_{ik} = x, V_k = v, T_k = t \} - \mathbb{E}Y_{ik}(t', m) | X_{ik} = x, V_k = v, T_k = t' \} \} P(X_{ik} = x, V_k = v) \text{ by (11)} \\
 & = \sum_{x,v} \{ \mathbb{E}Y_{ik}(t, m) | X_{ik} = x, V_k = v, T_k = t, M_{ik} = m \} - \mathbb{E}Y_{ik}(t', m) | X_{ik} = x, V_k = v, T_k = t', M_{ik} = m \} \} P(X_{ik} = x, V_k = v) \text{ by (12)} \\
 & = \sum_{x,v} \{ \mathbb{E}Y_{ik} | X_{ik} = x, V_k = v, T_k = t, M_{ik} = m \} - \mathbb{E}Y_{ik} | X_{ik} = x, V_k = v, T_k = t', M_{ik} = m \} \} P(X_{ik} = x, V_k = v) \text{ by consistency.}
 \end{aligned}$$

Proof of Theorem 2

$$\begin{aligned}
 & \mathbb{E}Y_{ik}(t, M_{ik}(t')) - \mathbb{E}Y_{ik}(t^*, M_{ik}(t')) \\
 & = \sum_{x,v} \{ \mathbb{E}Y_{ik}(t, M_{ik}(t')) | X_{ik} = x, V_k = v \} - \mathbb{E}Y_{ik}(t^*, M_{ik}(t')) | X_{ik} = x, V_k = v \} \} P(X_{ik} = x, V_k = v) \text{ by iterated expectations} \\
 & = \sum_{x,v} \sum_m \{ \mathbb{E}Y_{ik}(t, m) | X_{ik} = x, V_k = v, M_{ik}(t') = m \} - \mathbb{E}Y_{ik}(t^*, m) | X_{ik} = x, V_k = v, M_{ik}(t') = m \} \} P(M_{ik}(t') = m | X_{ik} = x, V_k = v) P(X_{ik} = x, V_k = v) \text{ by iterated expectations} \\
 & = \sum_{x,v} \sum_m \{ \mathbb{E}Y_{ik}(t, m) | X_{ik} = x, V_k = v \} - \mathbb{E}Y_{ik}(t^*, m) | X_{ik} = x, V_k = v \} \} P(M_{ik}(t') = m | X_{ik} = x, V_k = v) P(X_{ik} = x, V_k = v) \text{ by (17)} \\
 & = \sum_{x,v} \sum_m \{ \mathbb{E}Y_{ik}(t, m) | X_{ik} = x, V_k = v \} - \mathbb{E}Y_{ik}(t^*, m) | X_{ik} = x, V_k = v \} \} P(M_{ik}(t') = m | X_{ik} = x, V_k = v, T_k = t') P(X_{ik} = x, V_k = v) \text{ by (16)} \\
 & = \sum_{x,v} \sum_m \{ \mathbb{E}Y_{ik}(t, m) | X_{ik} = x, V_k = v, T_k = t \} - \mathbb{E}Y_{ik}(t^*, m) | X_{ik} = x, V_k = v, T_k = t^* \} \} P(M_{ik}(t') = m | X_{ik} = x, V_k = v, T_k = t') P(X_{ik} = x, V_k = v) \text{ by (11) and consistency} \\
 & = \sum_{x,v} \sum_m \{ \mathbb{E}Y_{ik}(t, m) | X_{ik} = x, V_k = v, T_k = t, M_{ik} = m \} - \mathbb{E}Y_{ik}(t^*, m) | X_{ik} = x, V_k = v, T_k = t^*, M_{ik} = m \} \} P(M_{ik} = m | X_{ik} = x, V_k = v, T_k = t') P(X_{ik} = x, V_k = v) \text{ by (12)} \\
 & = \sum_{x,v} \sum_m \{ \mathbb{E}Y_{ik} | X_{ik} = x, V_k = v, T_k = t, M_{ik} = m \} - \mathbb{E}Y_{ik} | X_{ik} = x, V_k = v, T_k = t^*, M_{ik} = m \} \} P(M_{ik} = m | X_{ik} = x, V_k = v, T_k = t') P(X_{ik} = x, V_k = v) \text{ by consistency.}
 \end{aligned}$$

Proof of Theorem 3

$$\mathbb{E}Y_{ik}(t', M_{ik}(t)) - \mathbb{E}Y_{ik}(t', M_{ik}(t^*))$$

$$\begin{aligned}
&= {}_{x,v} \mathbb{E} Y_{ik}(t', M_{ik}(t)) | X_{ik} = x, V_k = v] P(X_{ik} = x, V_k = v) - {}_{x,v} \mathbb{E} Y_{ik}(t', M_{ik}(t^*)) | X_{ik} = x, V_k \\
&= v] P(X_{ik} = x, V_k = v) \text{ by iterated expectations} \\
&= {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v, M_t = m] P(M_{ik}(t) = m | X_{ik} = x, V_k = v) P(X_{ik} = x, V_k = \\
&= v) - {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v, M_{ik}(t^*) = m] P(M_{ik}(t^*) = m | X_{ik} = x, V_k = v) P(X_{ik} = \\
&= x, V_k = v) \text{ by iterated expectations} \\
&= {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v] P(M_{ik}(t) = m | X_{ik} = x, V_k = v, T_k = t) P(X_{ik} = x, V_k = v) \\
&- {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v] P(M_{ik}(t^*) = m | X_{ik} = x, V_k = v, T_k = t^*) P(X_{ik} = x, V_k = \\
&= v) \text{ by (16)} \\
&= {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v, T_k = t \uparrow P(M_{ik} = m | X_{ik} = x, V_k = v, T_k = t) P(X_{ik} = x, \\
&= V_k = v) - {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v, T_k = t \uparrow P(M_{ik} = m | X_{ik} = x, V_k = v, T_k = \\
&= t^*) P(X_{ik} = x, V_k = v) \text{ by (17)} \\
&= {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v, T_k = t \uparrow P(M_{ik} = m | X_{ik} = x, V_k = v, T_k = t) P(X_{ik} = x, V_k \\
&= v) - {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v, T_k = t \uparrow P(M_{ik} = m | X_{ik} = x, V_k = v, T_k = t^*) P(X_{ik} \\
&= x, V_k = v) \text{ by (11)} \\
&= {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v, T_k = t', M_{ik} = m] P(M_{ik} = m | X_{ik} = x, V_k = v, T_k = \\
&= t) P(X_{ik} = x, V_k = v) - {}_{x,v} {}_m \mathbb{E} Y_{ik}(t', m) | X_{ik} = x, V_k = v, T_k = t', M_{ik} = m] P(M_{ik} = m | X_{ik} = \\
&= x, V_k = v, T_k = t^*) P(X_{ik} = x, V_k = v) \text{ by (12)} \\
&= {}_{x,v} {}_m \mathbb{E} Y_{ik} | X_{ik} = x, V_k = v, T_k = t', M_{ik} = m] P(M_{ik} = m | X_{ik} = x, V_k = v, T_k = t) P(X_{ik} = x, \\
&= V_k = v) - {}_{x,v} {}_m \mathbb{E} Y_{ik} | X_{ik} = x, V_k = v, T_k = t', M_{ik} = m] P(M_{ik} = m | X_{ik} = x, V_k = v, T_k = \\
&= t^*) P(X_{ik} = x, V_k = v) \text{ by consistency.}
\end{aligned}$$

Proof of Theorem 4

The conditions (30) and (31) together give

$Y_{ik}(\bar{t}, \bar{m}) \prod \{T_{ks}, M_{iks}\} | \bar{X}_{iks}, \bar{V}_{ks}, \bar{T}_{k(s-1)}, \bar{M}_{ik(s-1)}$. We show by induction that for all $r =$

$$\mathbb{E}[Y_{ik}(\bar{t}, \bar{m})] = \sum_{\bar{x}_r, \bar{v}_r} \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) | \bar{X}_{ikr}]$$

$$= \bar{x}_r, \bar{V}_{kr}$$

$$= \bar{v}_r, \bar{T}_{kr}$$

$$= \bar{t}_r, \bar{M}_{ikr}$$

$$= \bar{m}_r] \prod_{s=0}^r P(X_{iks})$$

$$= x, V_{ks}$$

$$= v \bar{X}_{ik(s-1)}$$

$$= \bar{x}_{s-1}, \bar{V}_{k(s-1)}$$

$$= \bar{v}_{s-1}, \bar{T}_{k(s-1)}$$

$0, 1, \dots, S$. $= \bar{t}_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}$. For $r = 0$, we have by the law of iterated

expectations that $\mathbb{E}[Y_{ik}(\bar{t}, \bar{m})] = \sum_{x_0, v_0} \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) | X_{ik0} = x_0, V_{k0} = v_0] P(X_{ik0} = x_0, V_{k0} = v_0)$

and by (30) and (31) that

$$\mathbb{E}[Y_{ik}(\bar{t}, \bar{m})] = \sum_{x_0, v_0} \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) | X_{ik0} = x_0, V_{k0} = v_0, T_{k0} = t_0, M_{ik0} = m_0] P(X_{ik0} = x_0, V_{k0} = v_0)$$

. Thus the result holds for $r = 0$. We show that if the result holds for $r = q$ then it holds also

$$\begin{aligned}
\mathbb{E}[Y_{ik}(\bar{t}, \bar{m})] &= \sum_{\bar{x}_q, \bar{v}_q} \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) \mid \bar{X}_{ikq}] \\
&= \bar{x}_q, \bar{V}_{kq} \\
&= \bar{v}_q, \bar{T}_{kq} \\
&= \bar{t}_q, \bar{M}_{ikq} \\
&= \bar{m}_q \prod_{s=0}^q P(X_{iks} \\
&= x, V_{ks} \\
&= v \bar{X}_{ik(s-1)} \\
&= \bar{x}_{s-1}, \bar{V}_{k(s-1)} \\
&= \bar{v}_{s-1}, \bar{T}_{k(s-1)} \\
&= \bar{t}_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}
\end{aligned}$$

for $r = q + 1$. If the result holds for $r = q$ then the law of iterated expectations we have

$$\begin{aligned}
\mathbb{E}[Y_{ik}(\bar{t}, \bar{m})] &= \sum_{\bar{x}_q, \bar{v}_q} \sum_{x_{q+1}, v_{q+1}} \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) \mid \bar{X}_{ikq} = \bar{x}_q, \bar{V}_{kq} = \bar{v}_q, \bar{T}_{kq} = \bar{t}_q, \bar{M}_{ikq} = \bar{m}_q, X_{ik(q+1)} = x_{q+1}, V_{k(q+1)} = v_{q+1}] \\
&= x_{q+1}, V_{k(q+1)} \\
&= v_{q+1} \bar{X}_{ikq} \\
&= \bar{x}_q, \bar{V}_{kq} \\
&= \bar{v}_q, \bar{T}_{kq} \\
&= \bar{t}_q, \bar{M}_{ikq} \\
&= \bar{m}_q, X_{ik(q+1)} = x_{q+1}, V_{k(q+1)} \\
&= v_{q+1} \prod_{s=0}^q P(X_{iks} \\
&= x, V_{ks} \\
&= v \bar{X}_{ik(s-1)} \\
&= \bar{x}_{s-1}, \bar{V}_{k(s-1)} \\
&= \bar{v}_{s-1}, \bar{T}_{k(s-1)} \\
&= \bar{t}_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}
\end{aligned}$$

. By (30) and (31) this expression is equal to

$$\begin{aligned}
&\sum_{\bar{x}_{q+1}, \bar{v}_{q+1}} \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) \mid \bar{X}_{ik(q+1)} = \bar{x}_{q+1}, \bar{V}_{k(q+1)} = \bar{v}_{q+1}, \bar{T}_{k(q+1)} = \bar{t}_{q+1}, \bar{M}_{ik(q+1)} = \bar{m}_{q+1}] \prod_{s=0}^{q+1} P(X_{iks} \\
&= x, V_{ks} \\
&= v \bar{X}_{ik(s-1)} \\
&= \bar{x}_{s-1}, \bar{V}_{k(s-1)} \\
&= \bar{v}_{s-1}, \bar{T}_{k(s-1)} \\
&= \bar{t}_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}
\end{aligned}$$

. Thus the result holds for $r = q + 1$. Applying the result for $r = S$ we have

$$\begin{aligned}
 \mathbb{E}[Y_{ik}(\bar{t}, \bar{m})] &= \sum_{\bar{x}, \bar{v}} \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) \mid \bar{X}_{ik}] \\
 &= \bar{x}, \bar{V}_k \\
 &= \bar{v}, \bar{T}_k \\
 &= \bar{t}, \bar{M}_{ik} \\
 &= \bar{m} \prod_{s=0}^S P(X_{iks}) \\
 &= x, V_{ks} \\
 &= v \bar{X}_{ik(s-1)} \\
 &= \bar{x}_{s-1}, \bar{V}_{k(s-1)} \\
 &= \bar{v}_{s-1}, \bar{T}_{k(s-1)} \\
 &= \bar{t}_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}
 \end{aligned}$$

and by consistency we have that (32) holds.

Proof of Theorem 5

$$\begin{aligned}
 \mathbb{E}[Y_{ik}(\bar{t}, \bar{M}_{ik}(\bar{t}^*))] &= \sum_{x_0, v_0} \mathbb{E}[Y_{ik}(\bar{t}, \bar{M}_{ik}(\bar{t}^*)) \mid X_{ik0}=x_0, V_{k0}=v_0] P(X_{ik0}=x_0, V_{k0}=v_0) \text{ by iterated} \\
 &\text{expectations} \\
 &= \sum_{x_0, v_0} \sum_m \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) \mid X_{ik0}=x_0, V_{k0}=v_0, \bar{M}_{ik}(\bar{t}^*)=\bar{m}] P(\bar{M}_{ik}(\bar{t}^*)=\bar{m} \mid X_{ik0}=x_0, V_{k0}=v_0) P(X_{ik0}=x_0, V_{k0}=v_0) \\
 &\text{by iterated expectations} \\
 &= \sum_{x_0, v_0} \sum_m \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) \mid X_{ik0}=x_0, V_{k0}=v_0] P(\bar{M}_{ik}(\bar{t}^*)=\bar{m} \mid X_{ik0}=x_0, V_{k0}=v_0) P(X_{ik0}=x_0, V_{k0}=v_0) \\
 &\text{by (34)}
 \end{aligned}$$

Using (30) and (31) we have by the inductive argument of Theorem 4, conditional on $X_{ik0} = x_0, V_{k0} = v_0$, that

$$\begin{aligned}
 \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) \mid X_{ik0}=x_0, V_{k0}=v_0] &= \sum_{(x_1, \dots, x_S), (v_1, \dots, v_S)} \mathbb{E}[Y_{ik} \mid \bar{X}_{ik}=\bar{x}, \bar{V}_k=\bar{v}, \bar{T}_k=\bar{t}, \bar{M}_{ik}=\bar{m}] \prod_{s=0}^S P(X_{iks}) \\
 &= x, V_{ks} \\
 &= v \bar{X}_{ik(s-1)} \\
 &= \bar{x}_{s-1}, \bar{V}_{k(s-1)} \\
 &= \bar{v}_{s-1}, \bar{T}_{k(s-1)} \\
 &= \bar{t}_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}
 \end{aligned}$$

We show by induction that

$$\begin{aligned}
 & P(\overline{M}_{ik}(\overline{t}^*)) \\
 &= \overline{m} \mid X_{ik0} \\
 &= x_0, V_{k0} \\
 &= v_0 = \sum_{(x_1, \dots, x_r), (v_1, \dots, v_r)} P(M_{ikr}(\overline{t}^*_{r+1}) = m_{r+1}, \dots, M_{iks}(\overline{t}^*)) \\
 &= m_s \mid \overline{X}_{ikr} \\
 &= \overline{x}_r, \overline{V}_{kr} \\
 &= \overline{v}_r, \overline{T}_{kr} \\
 &= \overline{t}^*_r, \overline{M}_{ikr} \\
 &= \overline{m}_r \{ \prod_{s=1}^r P(X_{iks} \\
 &= x, V_{ks} \\
 &= v_s \mid \overline{X}_{ik(s-1)} \\
 &= \overline{x}_{s-1}, \overline{V}_{k(s-1)} \\
 &= \overline{v}_{s-1}, \overline{T}_{k(s-1)} \\
 &= \overline{t}^*_{s-1}, \overline{M}_{ik(s-1)} \\
 &= \overline{m}_{s-1} \} \{ \prod_{s=0}^r P(M_{iks} \\
 &= m_s \mid \overline{X}_{iks} \\
 &= \overline{x}_s, \overline{V}_{ks} \\
 &= \overline{v}_s, \overline{T}_{ks} \\
 &= \overline{t}^*_s, \overline{M}_{ik(s-1)} \\
 &= \overline{m}_{s-1} \}
 \end{aligned}$$

By (33) for $s = 0$ and consistency we have,

$$\begin{aligned}
 & P(\overline{M}_{ik}(\overline{t}^*) = \overline{m} \mid X_{ik0} = x_0, V_{k0} = v_0) = P(\overline{M}_{ik}(\overline{t}^*) = \overline{m} \mid X_{ik0} = x_0, V_{k0} = v_0, T_{k0} = t^*_0) \\
 &= P(M_{ik1}(\overline{t}^*_1) \\
 &= m_1, \dots, M_{iks}(\overline{t}^*)) \\
 &= m_s \mid X_{ik0} \\
 &= x_0, V_{k0} \\
 &= v_0, T_{k0} \\
 &= t^*_0, M_{ik0}(t^*_0) \\
 &= m_0) P(M_{ik0}(t^*_0) \\
 &= m_0 \mid X_{ik0} \\
 &= x_0, V_{k0} \\
 &= v_0, T_{k0} = t^*_0)
 \end{aligned}$$

$$= P(M_{ik1}(\overline{t}^*_1) = m_1, \dots, M_{iks}(\overline{t}^*) = m_s \mid X_{ik0} = x_0, V_{k0} = v_0, T_{k0} = t^*_0, M_{ik0} = m_0) P(M_{ik0} = m_0 \mid X_{ik0} = x_0, V_{k0} = v_0, T_{k0} = t^*_0)$$

and thus the result holds for $r = 0$. Suppose the result holds for $r = q$ then

$$\begin{aligned}
 & P(\overline{M}_{ik}(\overline{t}^*) = \overline{m} \mid X_{ik0} = x_0, V_{k0} = v_0) \\
 &= \sum_{(x_1, \dots, x_q), (v_1, \dots, v_q)} P(M_{ik(q+1)}(\overline{t}^*_{q+1}) = m_{q+1}, \dots, M_{iks}(\overline{t}^*) = m_s \mid \overline{X}_{ikq} = \overline{x}_q, \overline{V}_{kq} = \overline{v}_q, \overline{T}_{kq} = \overline{t}^*_q, \overline{M}_{ikq} = \overline{m}_q)
 \end{aligned}$$

$$\begin{aligned}
& \left\{ \prod_{s=1}^q P(X_{iks}=x_s, V_{ks}=v_s \mid \bar{X}_{ik(s-1)}=\bar{x}_{s-1}, \bar{V}_{k(s-1)}=\bar{v}_{s-1}, \bar{T}_{k(s-1)}=\bar{t}_{s-1}^*, \bar{M}_{ik(s-1)}=\bar{m}_{s-1}) \right\} \\
& \left\{ \prod_{s=0}^q P(M_{iks}=m_s \mid \bar{X}_{iks}=\bar{x}_s, \bar{V}_{ks}=\bar{v}_s, \bar{T}_{ks}=\bar{t}_{s-1}^*, \bar{M}_{ik(s-1)}=\bar{m}_{s-1}) \right\} \\
& = \sum_{(x_1, \dots, x_{q+1}), (v_1, \dots, v_{q+1})} P(M_{ik(q+1)}(\bar{t}_{q+1}^*)=m_{q+1}, \dots, M_{iks}(\bar{t}^*) \\
& \quad =m_s \mid \bar{X}_{ik(q+1)} \\
& \quad =\bar{x}_{q+1}, \bar{V}_{k(q+1)} \\
& \quad =\bar{v}_{q+1}, \bar{T}_{kq} \\
& \quad =\bar{t}_{q+1}^*, \bar{M}_{ikq}=\bar{m}_q) \\
& P(X_{ik(q+1)}=x_{q+1}, V_{k(q+1)}=v_{q+1} \mid \bar{X}_{ikq}=\bar{x}_q, \bar{V}_{kq}=\bar{v}_q, \bar{T}_{kq}=\bar{t}_{q+1}^*, \bar{M}_{ikq}=\bar{m}_q) \\
& \left\{ \prod_{s=1}^q P(X_{iks}=x_s, V_{ks}=v_s \mid \bar{X}_{ik(s-1)}=\bar{x}_{s-1}, \bar{V}_{k(s-1)}=\bar{v}_{s-1}, \bar{T}_{k(s-1)}=\bar{t}_{s-1}^*, \bar{M}_{ik(s-1)}=\bar{m}_{s-1}) \right\} \\
& \left\{ \prod_{s=0}^q P(M_{iks}=m_s \mid \bar{X}_{iks}=\bar{x}_s, \bar{V}_{ks}=\bar{v}_s, \bar{T}_{ks}=\bar{t}_{s-1}^*, \bar{M}_{ik(s-1)}=\bar{m}_{s-1}) \right\} \\
& = \sum_{(x_1, \dots, x_{q+1}), (v_1, \dots, v_{q+1})} P(M_{ik(q+1)}(\bar{t}_{q+1}^*)=m_{q+1}, \dots, M_{iks}(\bar{t}^*) \\
& \quad =m_s \mid \bar{X}_{ik(q+1)} \\
& \quad =\bar{x}_{q+1}, \bar{V}_{k(q+1)} \\
& \quad =\bar{v}_{q+1}, \bar{T}_{k(q+1)} \\
& \quad =\bar{t}_{q+1}^*, \bar{M}_{ikq}=\bar{m}_q) \\
& \left\{ \prod_{s=1}^{q+1} P(X_{iks}=x_s, V_{ks}=v_s \mid \bar{X}_{ik(s-1)}=\bar{x}_{s-1}, \bar{V}_{k(s-1)}=\bar{v}_{s-1}, \bar{T}_{k(s-1)}=\bar{t}_{s-1}^*, \bar{M}_{ik(s-1)}=\bar{m}_{s-1}) \right\} \\
& \left\{ \prod_{s=0}^q P(M_{iks}=m_s \mid \bar{X}_{iks}=\bar{x}_s, \bar{V}_{ks}=\bar{v}_s, \bar{T}_{ks}=\bar{t}_{s-1}^*, \bar{M}_{ik(s-1)}=\bar{m}_{s-1}) \right\} \\
& = \sum_{(x_1, \dots, x_{q+1}), (v_1, \dots, v_{q+1})} P(M_{ik(q+2)}(\bar{t}_{q+2}^*) \\
& \quad =m_{q+2}, \dots, M_{iks}(\bar{t}^*) \\
& \quad =m_s \mid \bar{X}_{ik(q+1)} \\
& \quad =\bar{x}_{q+1}, \bar{V}_{k(q+1)} \\
& \quad =\bar{v}_{q+1}, \bar{T}_{k(q+1)} \\
& \quad =\bar{t}_{q+1}^*, \bar{M}_{ikq} \\
& \quad =\bar{m}_q, M_{ik(q+1)}(\bar{t}_{q+1}^*)=m_{q+1}) \\
& P(M_{ik(q+1)}(\bar{t}_{q+1}^*)=m_{q+1} \mid \bar{X}_{ik(q+1)}=\bar{x}_{q+1}, \bar{V}_{k(q+1)}=\bar{v}_{q+1}, \bar{T}_{k(q+1)}=\bar{t}_{q+1}^*, \bar{M}_{ikq}=\bar{m}_q) \\
& \left\{ \prod_{s=1}^{q+1} P(X_{iks}=x_s, V_{ks}=v_s \mid \bar{X}_{ik(s-1)}=\bar{x}_{s-1}, \bar{V}_{k(s-1)}=\bar{v}_{s-1}, \bar{T}_{k(s-1)}=\bar{t}_{s-1}^*, \bar{M}_{ik(s-1)}=\bar{m}_{s-1}) \right\} \\
& \left\{ \prod_{s=0}^q P(M_{iks}=m_s \mid \bar{X}_{iks}=\bar{x}_s, \bar{V}_{ks}=\bar{v}_s, \bar{T}_{ks}=\bar{t}_{s-1}^*, \bar{M}_{ik(s-1)}=\bar{m}_{s-1}) \right\}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{(x_1, \dots, x_{q+1}), (v_1, \dots, v_{q+1})} P(M_{ik(q+2)}(\bar{t}^*_{q+2}) = m_{q+2}, \dots, M_{ikS}(\bar{t}^*)) \\
&= m_s | \bar{X}_{ik(q+1)} \\
&= \bar{x}_{q+1}, \bar{V}_{k(q+1)} \\
&= \bar{v}_{q+1}, \bar{T}_{k(q+1)} \\
&= \bar{t}^*_{q+1}, \bar{M}_{ik(q+1)} = \bar{m}_{q+1}) \\
&\{ \prod_{s=1}^{q+1} P(X_{iks} = x_s, V_{ks} = v_s | \bar{X}_{ik(s-1)} = \bar{x}_{s-1}, \bar{V}_{k(s-1)} = \bar{v}_{s-1}, \bar{T}_{k(s-1)} = \bar{t}^*_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}) \} \\
&\{ \prod_{s=0}^{q+1} P(M_{iks} = m_s | \bar{X}_{iks} = \bar{x}_s, \bar{V}_{ks} = \bar{v}_s, \bar{T}_{ks} = \bar{t}^*_s, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}) \}
\end{aligned}$$

Thus the result holds for $r = q + 1$ and by the principle of induction for all $r = 0, 1, \dots, S$.

Applying the result to $r = S$ we have,

$$\begin{aligned}
&P(\bar{M}_{ik}(\bar{t}^*)) \\
&= \bar{m} | X_{ik0} \\
&= x_0, V_{k0} \\
&= v_0) = \sum_{(x_1, \dots, x_S), (v_1, \dots, v_S)} \prod_{s=1}^S P(X_{iks} \\
&= x_s, V_{ks} \\
&= v_s | \bar{X}_{ik(s-1)} \\
&= \bar{x}_{s-1}, \bar{V}_{k(s-1)} \\
&= \bar{v}_{s-1}, \bar{T}_{k(s-1)} \\
&= \bar{t}^*_{s-1}, \bar{M}_{ik(s-1)} \\
&= \bar{m}_{s-1}) \{ \prod_{s=0}^S P(M_{iks} = m_s | \bar{X}_{iks} = \bar{x}_s, \bar{V}_{ks} = \bar{v}_s, \bar{T}_{ks} = \bar{t}^*_s, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}) \} \\
&\mathbb{E}[Y_{ik}(\bar{t}, \bar{M}_{ik}(\bar{t}^*))] = \sum_{x_0, v_0} \sum_m \mathbb{E}[Y_{ik}(\bar{t}, \bar{m}) | X_{ik0} = x_0, V_{k0} = v_0] P(\bar{M}_{ik}(\bar{t}^*)) \\
&= \bar{m} X_{ik0} \\
&= x_0, V_{k0} \\
&= v_0) P(X_{ik0} \\
\text{Thus, } &= x_0, V_{k0} = v_0)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{x_0, v_0} \sum_m \{ \sum_{(x_1, \dots, x_S), (v_1, \dots, v_S)} \mathbb{E}[Y_{ik} | \bar{X}_{ik} = \bar{x}, \bar{V}_k = \bar{v}, \bar{T}_k = \bar{t}, \bar{M}_{ik} = \bar{m}] \\
&\prod_{s=0}^S P(X_{iks} = x, V_{ks} = v | \bar{X}_{ik(s-1)} = \bar{x}_{s-1}, \bar{V}_{k(s-1)} = \bar{v}_{s-1}, \bar{T}_{k(s-1)} = \bar{t}_{s-1}, \bar{M}_{ik(s-1)} = \bar{m}_{s-1}) \}
\end{aligned}$$

$$\begin{aligned}
& \sum_{(x_1, \dots, x_S), (v_1, \dots, v_S)} \{ \prod_{s=0}^S P(M_{iks} \\
& \quad = m_s \mid \bar{X}_{iks} \\
& \quad = \bar{x}_s, \bar{V}_{ks} \\
& \quad = \bar{v}_s, \bar{T}_{ks} \\
& \quad = \bar{t}^*_s, \bar{M}_{ik(s-1)} \\
& \quad = \bar{m}_{s-1}) \} \{ \prod_{s=0}^S P(X_{iks} \\
& \quad = x_s, V_{ks} \\
& \quad = v_s \bar{X}_{ik(s-1)} \\
& \quad = \bar{x}_{s-1}, \bar{V}_{k(s-1)} \\
& \quad = \bar{v}_{s-1}, \bar{T}_{k(s-1)} \\
& \quad = \bar{t}^*_{s-1}, \bar{M}_{ik(s-1)} \\
& \quad = \bar{m}_{s-1}) \}
\end{aligned}$$

This completes the proof.